



Evaluation of Legislation: Skating on Thin Ice

Evaluation

16(3) 279–293

© The Author(s) 2010

Reprints and permission: sagepub.

co.uk/journalsPermissions.nav

DOI: 10.1177/1356389010370252

<http://evi.sagepub.com>



Werner Bussmann

Federal Office of Justice, Switzerland

Abstract

While parliaments and politicians increasingly call for the evaluation of legislation, the methodological implications of statutes as an evaluand have not received much interest. This article explores the challenges and methods of evaluating statutes. It describes the methodology and the costs of eight all-encompassing evaluations of statutes at Swiss federal level. It then presents approaches to evaluating statutes. A distinction is made between statutes with the aim of problem-solving and those relating to social ordering. While the former are concerned with a 'policy' (objectives combined with intervention instruments), the latter are concerned with the 'basic order', which regulates the rights and indirectly the interactions of various actors, but which does not seek a specific outcome. A good part of the legislation aimed at problem-solving can be broken down into its main components; the evaluation of these components does not necessarily differ from the evaluation of programmes. Quite often, however, this approach is not feasible for various reasons. Strategies are presented to evaluate statutes in a more comprehensive manner.

Keywords

causal attribution, evaluation methodology, legislation, Switzerland

Evaluation Methodology and the Evaluation of Legislation

Government policies in many European countries are primarily regulated by statutes (*lois, Gesetze*). In most countries, they are adopted by parliament and thus have high legitimacy. They set out – in an abstract and generalized form – the rules that apply to citizens and/or to government agencies. They regulate entire policy fields or treat problems in a coherent manner. In general, they apply to the whole population or to the whole territory concerned. In many cases, the general rules laid down in statutes need to be put in more specific terms in the form of ordinances (by-laws, statutory instruments, *ordonnances, Verordnungen*). Implementation often is carried out by decentralized or federate units. This is certainly the case with EU legislation, which is implemented at national or subnational level.

Corresponding author:

Werner Bussmann, Federal Office of Justice, Bern, Switzerland.

Email: werner.bussmann@bj.admin.ch

The types of evaluands and their legal status – projects, programmes, statutes or whole policy domains – have not received much interest in evaluation methodology. Most bibliographical entries on the subject of ‘evaluation of legislation’ stem from the legal sciences (e.g. Karpen, 2002; Mader, 2001, 2003a, 2003b; Schäffer, 2005, 2007).

Evaluation methodology is somehow at odds with statutes. Its main object has been and still is the project or programme, which has clear boundaries in time and has specific beneficiaries. Statutes, however, are intended for an unlimited period of time and normally apply to an entire territory and population. Attributing effects to statutes poses problems. Some of these problems were tackled quite early on under the heading of ‘broad-aim programs’ in the much cited article by Weiss and Rein (1970). The authors regard the experimental approach as unsuitable for broad-aim programmes and recommend instead a qualitative study of development and change. In recent years, this topic has not generated much interest. For instance, only two articles in the journal *Evaluation* (Klein Haarhuis and Niemeijer, 2009; Sverdrup, 2003) have been devoted to the subject of the evaluation of legislation.

Evaluation of legislation, however, is becoming an issue. The media and the public, parliaments and the executive, increasingly want to know how statutes are implemented and what results they bring. Parliaments ask for information either from the executive branch or from their own evaluation or audit units. The British parliament has ‘begun taking steps towards greater post-legislative scrutiny’ (HM Government, 2008; Law Commission, 2006). Other parliaments are equally interested in the evaluation of legislation. In Germany, evaluation is closely related to law-making and the implementation of statutes (see the term *Gesetzesfolgenabschätzung*, Böhret and Konzendorf, 2001). The European Union is increasingly trying not only to evaluate financial programmes, but legislation as well (European Commission, 2008). Therefore, a clearer understanding of the issues involved in the evaluation of legislation is needed.

This is important because expectations regarding the evaluation of legislation are high. Politicians and public authorities would like to know whether statutes are achieving the expected results. Will evaluations be able to create a solid basis for policy decisions? Or will policy decisions based on evaluation of statutes rest on a foundation as thin as ice?

This article approaches the subject of the evaluation of legislation by asking two questions:

1. Are there specific methodological challenges in evaluating statutes?
2. If yes, how can these challenges be met?

We will answer these two questions 1) by – inductively – taking an inventory of a sample of Swiss evaluations of legislation and 2) by – analytically – deriving methodological conclusions.

Evaluation of Statutes: Some Preliminary Thoughts

Public policy evaluands differ widely with regard to their aggregation level. The following paragraphs take a closer look at programmes, statutes and policies.

Programmes usually consist of several projects that have a common objective. They are usually found at a low aggregation level. Although programmes, especially in multilevel governance, may be fairly complex, they are normally built around a certain logic (‘A leading to B and then to C’). Programmes have specific target groups, whose evolution can be compared with the situation of other sections of the population not affected by the programme. Thus, it is often possible to make empirically backed statements on cause–effect relationships.

Table I. Comparison of evaluation of programs, statutes and policy fields

	Evaluation of programs	Evaluation of statutes	Evaluation of policy fields
Aggregation level	+	++	+++
Complexity	+	++	+++
Targets	Simple, coherent	Multiple, often contradictory	Multiple, often contradictory
Methods	Longitudinal and cross-sectional comparisons, experiments	International comparisons, longitudinal comparisons, expert opinion	International comparisons, longitudinal comparisons, expert opinion, evaluation synthesis
Evaluative Statements	Descriptive, normative (nominal comparisons), often evidence-based statements on cause-effect relationships	Descriptive, normative (nominal comparisons), sometimes evidence-based statements on cause-effect relationships	Descriptive, normative (nominal comparisons), plausible interpretations based on expert opinion

At a higher aggregation level, there are statutes. Typically they regulate a certain subject matter (such as primary education, food quality, banking, development aid) or cross-sectional issues (such as environmental quality). When setting a framework for a subject matter, statutes often try to achieve multiple aims, some of which may be contradictory. As statutes usually apply to the whole territory of a country, it is hard to find counterfactual evidence. (What would the situation be, if the statute were not in force?) International comparisons can in principle serve as a substitute if the conditions in other countries are easily comparable.

At the highest aggregation level, there are policy fields such as educational, agricultural or environmental policy. Policy fields are aggregations of programmes or statutes focused on specific outcomes. Thus they are complex ‘ensembles’. The programmes or statutes that form a policy interact and mutually reinforce or weaken the effects. This is why establishing causal relationships is a difficult task. Actual policy evaluations (e.g. OECD reports) very much rely on expert opinion and on evaluation synthesis.

Unfortunately, the situation with regard to the evaluation of programmes, statutes and policies is not as straightforward as has been depicted. ‘Complexity’ is not primarily a feature of the system under examination (programme, statute or policy), but depends on the objectives of the observer as well (Dery, 1984: 68–81). Thus, a programme can be examined in a rather simple way, if overall achievements (e.g. number of projects completed) are at the centre of the interest. It can also be studied in a most sophisticated way, if knowledge purposes and academic interests prevail (e.g. observing and explaining actor networks in policy formulation and implementation). Similarly, policies can be presented in a simple and straightforward way by treating them as black boxes and by looking at aggregate input (e.g. budget figures) and outcome figures (e.g. number of university graduates). They can also be analysed in a very detailed manner by looking at the mutual effects and counter-effects of the programmes involved. Complexity and non-complexity lie in the ‘eye of the beholder’.

This brings us back to the question of evaluating legislation. Statutes are undoubtedly more complex than individual programmes (of which they are sometimes composed). But does this mean that they are necessarily more demanding to analyse than programmes? What are the strategies for

coping with the diversity and complexity of the issues involved? To answer these questions, we will look at a sample of evaluations of statutes in Switzerland.

The Law as an Evaluand in Switzerland

Evaluations and Evaluation of Statutes

Evaluation in Switzerland has gained considerably in importance in recent decades (Bussmann, 2008). Balthasar (2007: 298) found 308 evaluations at federal level in Switzerland for the 1999–2002 period, i.e. approximately 100 per year. Around half of these are highlighted in the annual reports of the Federal Council (government), the Federal Audit Office and the Parliamentary Control of the Administration. Unfortunately, there are no figures available on whether the evaluations concern projects, programmes, statutes or policies. Balthasar (2007: 320) found that 19.1 percent of the evaluations examined were addressed to the parliament and its members, which is an indicator that they are related to legislation. According to our own estimates, less than 10 percent of evaluations relate to statutes.¹ However, the number of all-encompassing evaluations, i.e. evaluations that cover all aspects of a statute, is much smaller. Based on the ARAMIS database and our own experience, we have identified the following eight ‘all-encompassing evaluations’, which means that they are evaluations of the whole statute.² They can thus serve as ‘pioneers’ for studying the problems and the possible solutions involved in the evaluation of legislation. Two studies (studies 7 and 8) that are yet not complete are also included in order to pinpoint recent trends in the evaluation of statutes.

Findings

In presenting the findings on the eight studies, we will start with the similarities and then consider the differences.

Impetus from parliament. Formally, all eight evaluations were triggered by parliament, which either inserted an evaluation clause into the statute (studies 1, 2, 3, 6, 8) or requested the evaluation by means of a parliamentary procedural request (a ‘postulate’) (studies 4, 5, 7). However, in studies 1, 2, 7, 8, the administration favoured evaluation from the outset. It presented the idea of an evaluation clause (1, 2, 8) or started evaluation activities before the parliament asked for them (study 7).

Combination of commissioned studies and in-house involvement. In all eight studies, there was considerable in-house involvement. It consisted in preparing, commissioning and utilizing the studies (all eight studies), but also in making extensive in-house studies (study 6), in synthesizing reports of cantonal programme managers (study 2), in writing a report to the parliament based on the findings (studies 4 and 5), in supporting the evaluation team by assuming specific tasks (studies 5 and 7) and in preparing the questionnaire for the survey (study 4). On average, in-house involvement was/is just as important as external mandates. Thus the conduct of the evaluation of statutes in the Swiss federal administration is of hybrid nature.

Above average costs. All but one (study 4) of the eight studies incurred far above average costs.³

Table 2. Profile of eight evaluations of statutesⁱ

Subject	Methodology	External / internal costs (in Swiss francs)
1. Energy law (1990–2000)	58 studies on specific energy programs with a synthesis report	6,300,000/no figures av.
2. Aid to victims of crimes (1996, 1998, 2000, 2008)	4 studies at 2-3-year intervals on satisfaction of victims with support provided, on criminal law procedures, on cooperation between police and counseling services and on subsidies for the training of counseling services	270,000/300,000
3. Regional policy (2004)	Analysis of data on regional development and on aid decisions, interviews with regional coordinating agencies in the mountain regions, case studies	200,000/40,000
4. Divorce law (2005)	Survey of 500 lawyers and judges experienced in divorce cases	25,000/40,000
5. Gender equality / equal salaries (2006)	Interviews with employee organizations, with HR officers of large and medium-sized companies and with mediation offices; analysis of court decisions and of mediation results, analysis of salary gaps between men and women	300,000/ 300,000
6. Competition law (2008)	5 external and 12 internal studies: organizational analysis, activity analysis, international comparative law studies, studies of outcomes of competition authorities' decisions	315,000/600,000
7. Justice reform (ongoing, 2008–2012)	Surveys of courts and lawyers, data analysis, analysis of court decisions	380,000/ 380,000
8. Organ transplantation (2005–2011)	Baseline measurement, monitoring system, formative evaluation, final summative evaluation	912,000/ no figures av. yet

ⁱ Information sources for the 8 studies (often with multiple publications) are:

study 1: Balthasar 2000; Balthasar and Rieder 2000

study 2: <http://www.bj.admin.ch/bj/de/home/themen/gesellschaft/opferhilfe/evaluationen.html>

study 3: [http://www.seco.admin.ch/themen/00476/00496/00497/index.html? > Evaluation IHG](http://www.seco.admin.ch/themen/00476/00496/00497/index.html?>Evaluation%20IHG)

study 4: <http://www.bj.admin.ch/bj/de/home/dokumentation/medieninformationen/2005/0.html>

study 5: <http://www.bj.admin.ch/bj/de/home/dokumentation/medieninformationen/2006/2006-02-16.html>

study 6: <http://www.weko.admin.ch/dokumentation/00216/index.html?lang=de>

study 7: <http://www.bj.admin.ch/bj/de/home/dokumentation/medieninformationen/2007/2007-07-03.html>

study 8: <http://www.bag.admin.ch/transplantation/00694/04138/index.html?lang=de>

Single–multiple research tools. Only one study (4) relied on one research tool only. In the other seven evaluations, multiple research tools were used. Two approaches can be observed:

- In studies 1 and 2, the different components of the statutes were examined in individual studies. In study 1, a synthesis report was prepared.
- In studies 3, 5, 6, 7 and 8, various research tools were used to investigate different aspects of the statute (e.g. outputs, reactions of target groups, evidence on final results, etc.). Partly, triangulation took place, e.g. in the study 5 on gender equality: the inquiry among HR officers of large and medium-sized firms, among employees' organizations and the analysis of court decisions all confirmed the hypothesis that the Gender Equality Act was well accepted and had no negative side-effects.

One to ten years. Study 4 was conducted in less than a year, studies 3, 5 and 6 were conducted over two to three years, while the other studies (with multiple evaluations) took/take longer, one of them ten years (study 1).

One–multiple points of time. Studies 3, 4, 5 and 6 were done at a single point of time, while studies 1, 2, 7 and 8 were/are carried out at multiple points of time.

Causal Attribution of Results to Statutes

In the broad and complex settings of legal acts, causal attribution, i.e. obtaining convincing evidence that the legal document is the cause of the outputs, impacts and the final results, is a difficult task. All the statutes examined cover the entire Swiss territory and/or the entire population. Thus, an experimental approach by which causal attribution is achieved by looking for differences between the target and the control group was not feasible. Some of the studies, however, developed strategies to deal with the problem of causal attribution.

Gender Equality with Regard to Salaries. In this evaluation, the implementation–results chain, i.e. the mechanism by which the legal act affects the legal position of its addressees (firms and employees) and brings about changes in their behaviour (e.g. a reduction in gender discrimination by firms) which in turn might have further consequences (e.g. reducing the salary gap between men and women) was examined by individual research modules. The results were somewhat contradictory: on the one hand, the target groups and the main stakeholders (employer’s associations, HR officers, trade unions), in general, were quite satisfied with the Gender Equality Act. On the other, analysis of salaries showed that women’s salaries lagged 25 percent behind men’s salaries; in 40 percent of cases, this difference could not be explained by objective factors (age, education, career path, etc.). Thus there was inconsistency between the opinion of how well the Gender Equality Act was working and its final results. In part, the explanation for this inconsistency is that the Gender Equality Act has little effect on the practices of many small and medium-sized enterprises. An agreement has recently been reached between employers and trade unions in order to bridge the gap between the salaries of men and women. It concentrates primarily on small and medium-sized enterprises.

Competition policy. In this evaluation as well, the implementation–results chain was examined in a thorough manner. The clients’ and experts’ satisfaction with competition policy, as confirmed by the surveys, demonstrated the importance of the policy for the economy. By making an international comparison, the strengths and weaknesses of Swiss competition policy were revealed. When preparing studies on the final results, a surprising discovery was made: not one single empirical study of the effects of competition authorities’ decisions on market development was found anywhere in the world. Two parallel studies on the results of Swiss competition authority decisions in certain markets (including prices and productivity) were commissioned. They could not identify any specific results of those decisions, except in the case of the electricity market. It seems that the possible positive or negative effects of competition authorities’ decisions in the relevant markets cannot easily be detected over the ‘white noise’ of economic activities.

Support for crime victims. Separate studies examined the impact of the Federal Act on the Provision of Support to Victims of Crime (Victim Support Act) in various fields (satisfaction of victims with support provided, criminal law procedures, cooperation between police and counselling services, subsidies for the training of counselling services; see Figure 1). These studies were not trying,

however, to provide empirical evidence that the support offered to victims was improving their social and professional reintegration (when compared with a situation where no support was provided). A feasibility study revealed that such a research project would encounter serious data problems: substantive data on the reintegration of victims who had made use of public help and of those victims who had *not* made use of the available support would have had to have been collected and analysed. This was judged to be too expensive as well as too intrusive on the lives of crime victims.

Difficulty of Producing Counterfactual Evidence

Our modest overview of eight evaluations of statutes – all with a comprehensive approach – shows the difficulties of such an endeavour. None of the studies, with the exception of study 1, which examined the effectiveness and efficiency of multiple subsidies for energy conservation, could produce counterfactual evidence. This is not primarily the fault of the evaluators but a consequence of the comprehensive character of the statutes. In five of the six evaluations already completed, the findings of the different elements of the evaluation were not internally consistent (e.g. satisfaction of stakeholders with the statute, yet no evidence of success regarding impacts/outcomes). Although all of those evaluations but one (study 4 on the divorce law) tried to produce evidence on the implementation chain, there are still missing links, e.g. implementation at regional or local level (study 3) or the enforcement in small and medium-sized enterprises (study 5).

The first question of whether evaluation of legislation faces specific challenges can thus be answered with a ‘yes’. We now look at the second question, of how these challenges can be met. Based on our experience with evaluating statutes we propose strategies for coping with the vastness and complexity of statutes. We first start with the inherent nature of the statutes involved.

Approaches to Evaluating Statutes

Legislation as Social Ordering or Social Problem-Solving

Legislation can serve two purposes. On the one hand, it can have the aim of social ordering, i.e. providing a framework for the behaviour of persons or organizations without specifying the outcome of the ensuing legal, social and economic processes. Civil, criminal and economic legislation belong to this type of legislation. On the other hand, it can have the aim of social problem-solving by correcting social, economic and environmental processes in order to achieve a specified outcome. Policies to alleviate poverty, to improve the environment, to combat violence or to favour technological progress belong to this category. Naturally it is quite often the case that legislation is of hybrid nature, containing provisions with elements of both types.

The distinction between social ordering and social problem-solving and the consequences for evaluation can be shown by looking at a statute of hybrid type, Switzerland’s Victim Support Act.

The Swiss legislation on victim support rests on three pillars:

1. Counselling offices give victims of crimes immediate support and advice (information about medical aid, psychological aid and legal advice).
2. Schemes for providing financial compensation exist for those who are not compensated by the aggressor or through insurance and who belong to the lower income bracket.
3. Criminal procedures have been adapted to the needs of victims, e.g. by avoiding direct confrontation between the aggressor and the victim or by excluding the general public from the trial in certain circumstances (e.g. sexual abuse).

Counseling services for victims Victims' satisfaction ↓ Reintegration into social life and into labor market	Financial compensation Victims' satisfaction ↓ Reintegration into social life and into labor market	Victims' rights in criminal proceedings Avoiding secondary victimization Balancing the rights of defendants and victims, restricted access of the general public to trials
Social problem solving		Social ordering

Figure 1. Swiss support for crime victims and its three pillars

Pillars 1 and 2 are intended to help victims who are in danger of feeling neglected and not obtaining justice. These measures should help victims to reintegrate into social and economic life. This is a form of social problem-solving.

Pillar 3 serves similar purposes. It helps to avoid secondary victimization during the trial. In one sense, this is a form of social problem-solving. At the same time, it belongs to a much larger question, i.e. 'justice'. It involves the question of the rights of the public prosecutor, of the victim and of the defendant. It also involves the right of the general public to attend trials and to be informed about court judgments. A balance has to be found between all these rights. If, for some reason, the rights of the victim are improved, this might have implications for other groups, e.g. the rights of the defendant to be able to conduct a proper defence or the right of the general public to be given information about trials. All of this belongs to social ordering. Legislation triggers complex processes in which the justice authorities (courts, prosecutors, prison/probation authorities) have an important say. The theme is not 'policy' (objectives combined with intervention instruments), but relates to the 'basic order' (i.e. regulating the rights and indirectly the interactions of various actors with an undefined outcome).

Implications for Evaluation

The distinction between social ordering and social problem-solving legislation has implications for evaluation (see Table 3).

- Social ordering: statutes which lay down a basic order, i.e. a framework for the behaviour of persons or organizations, can only be partially approached by the ordinary goal-oriented evaluation approach. An understanding of the complex socioeconomic-legal order is needed. Basic or applied research on the interaction of the law and the economic or social actors can satisfy those needs. As the social order is constantly changing, research is always lagging behind and barely successful in mapping the actual state of things.
- Social problem-solving: statutes that try to correct problems in a certain policy area can often be broken down into their main topical parts. Those parts, in turn, can be evaluated like programmes (component evaluation: Donaldson, 2005: 101–2; Scriven, 1991: 84). In practice, however, the different parts of the statute may interact with each other, causing a multi-treatment problem (isolation of effects of each program not feasible) and they may interact with other statutes (creating problems of unclear boundaries of the evaluand).
- Hybrid nature: statutes quite often have elements of social ordering and of social problem-solving. As a consequence, a combined approach is needed. Those parts of the statute that

Table 3. Types of legislation and evaluative approaches

Types of legislation	Statutes		
	Social ordering	Hybrid	Problem solving
Content of legislation	Framework for behavior of persons or organizations	Elements of both	Taking steps to correct a (social, economic or environmental) problem
Objective of evaluation	Describing and understanding the complex socio-economic-legal order	Elements of both	Linking objectives, government activities, addresses' behavior and final results
Evaluation strategy	Basic and applied research (social, economic, environmental, legal studies)	Elements of both	Break the statute down into its main parts and evaluate the parts in a program-like manner

belong to social ordering need an approach aimed at understanding the processes of social and economic interaction in the relevant ‘basic order’. Other components of the statute can, if they do not interact with each other, be isolated and evaluated like individual programmes.

In our experience, in around half of the statutes the component approach works. This means that it is possible to isolate parts of the statute and evaluate them like programs. A typical case of this approach is study 1 (see Table 2) which consisted of 58 individual studies. In the other half of statutes, an integrated evaluation approach is needed. We will leave statutes aside that aim at social ordering (such as civil, penal and economic legislation) and which necessitate substantive research from the relevant academic (sub)disciplines (civil and private law, business economics). Instead we will concentrate on statutes that aim at social problem-solving. Again, based on the eight studies depicted in Table 2 and based on experience with further studies, we will outline approaches to evaluating statutes in a more comprehensive manner and present their advantages and shortcomings.

Possible Strategies for an All-Encompassing Evaluation of Legislation

The Experimental Strategy

The experimental strategy, as sound and convincing as it is from a methodological point of view, does not work for an all-encompassing evaluation of legislation. Continental European legislation is synonymous with ‘full coverage’ and thus in contradiction to the experimental strategy. Impact attribution in the case of full coverage programmes is extremely difficult in a strict methodological sense (Rossi and Freeman, 1993: 333–62: chapter omitted in Rossi et al., 2004).

This does not exclude an experimental approach altogether. In many cases, modern legislation has in-built ‘experimental zones’. Authorities are allowed to test new treatment and implementation models (such as in Switzerland with regard to penal institutions, drug addiction, health care, traffic regulation, education, electronic voting, communication and information technology, welfare policy, etc.) and those experiments can be evaluated according to the canons of the experimental approach (either quantitatively or qualitatively). Frequently, federalist and decentralized governance arrangements, giving elbowroom for implementation agencies, create differences in implementation, thus allowing for quasi-experimental evaluation designs. This is also the case if there is rich quantitative

data on the target population (as in labour market legislation). We would urge legislators to provide windows of opportunity for experimentation. At the same time, we have to be realistic and acknowledge that experimentation will be the exception and not the rule: it is the inherent nature of Continental European legislation that it encompasses the whole territory and population.

Descriptive Strategy

The descriptive strategy is completely different from the experimental one. It leaves causal questions aside. It simply tries to describe the situation, often combined with a comparison between budgeted and real figures.

This approach dominates in managerial governance models (coming under headings such as new public management, performance management, ‘controlling’ and the like). It rests on the assumption that input, output and outcome figures will give a sufficiently accurate picture of the situation.

This strategy has its merits in the evaluation of legislation. It provides essential data on implementation of legislation and on the development of the objectives which legislation tries to achieve.

At the same time, the descriptive strategy can be misleading. The success of a statute, as demonstrated by the positive development of outcome indicators, may be caused entirely by favourable general conditions (such as economic growth). If those conditions change for the worse, ineffective legislation may become a tremendous handicap. Conversely, a statute and the public interventions it includes may be working sufficiently well, but the effects may be overshadowed by adverse general conditions. In a situation of crisis, these interventions will not be applied vigorously and a chance to improve the situation will be missed. In legislation, making correct causal attributions is of prime interest because policy-makers want to know whether specific interventions should be reduced or reinforced, or be applied in an additional policy field.

Strategy of Plausible Arguments

Given the limited applicability of the experimental approach and at the danger of adopting a misconceived descriptive (or purely managerial) approach, a pragmatic third strategy would be useful. It does not rely on scientific ‘proofs’ as intended by proponents of the experimental approach (yet in practice seldom realized), but on plausible arguments.

Many of the methods developed for the evaluation of *programmes* (single case studies, realist evaluation, tests of mediation: Mark and Henry, 2006: 319–39) are of limited value for the evaluation of legislation, which has to be carried out at a fairly high level of aggregation. Of course, an evaluation of a specific statute can incorporate findings from such studies which will serve as elements in an assessment of the effectiveness of that statute. Case studies (Yin, 2009), for instance, can be helpful in getting first insights into the complex working of legislation at one or different sites.

The current intention in all-encompassing evaluations of statutes in Switzerland is to find as much empirical evidence as possible on the implementation of the statute by government agencies, on target groups’ reactions, on secondary effects and on final results. This approach is in certain respects related to Michael Scriven’s ‘Modus operandi method’ (Scriven, 1974 and 1991: 234; later somewhat extended under the heading of ‘observation’ in Scriven, 2005: 43–7, see also Davidson, 2005: 75–6, and – under the name of ‘modus narrandi’ – Gysen et al., 2006), but goes beyond the analysis of an individual case or treatment. It owes a lot to approaches using programme theory (Chen, 1990; Pawson and Tilley, 1997; Stame 2004; Virtanen and Usikylä, 2004; Weiss, 1998). It is also much in line with Mayne’s ‘contribution analysis’ (Mayne, 2001, 2008). Being able to show that implementation took place as planned, that the target group behaved as predicted and that

outcome indicators moved into the 'right' direction are plausible arguments that these changes happened because of the statute.

Of course, it cannot be excluded that favourable external effects (economic improvements, changes in social values or learning processes in the population, etc.) could account for the positive result. Therefore, efforts have to be made to rule out alternative explanations on the positive (or adversely: negative) effects of the statute under investigation. The task is to find as many pieces of evidence to piece together the mosaic of the chain of effects. Three possible sources can be considered for improving the plausibility of arguments.

Utilizing variation. All possible variation can serve as evidence of a statute 'working' or 'not working'. Longitudinal comparisons from a methodological point of view are not considered a powerful evaluation instrument, but can serve as evidence, especially if the three following preconditions, preferably in combination, are met: 1) the new or modified statute incorporates significant changes in comparison with the previous legislation, 2) no important other changes with effects on implementation and target group's behaviour take place *or* these effects can be isolated (e.g. by adjusting data for those additional changes) and 3) there is sufficient and qualitatively acceptable data allowing for before/after comparisons, preferably data with a lot of observation points (in order to eliminate trends that exist independent of legislation).

Although in evaluation of legislation cross-sectional variation with regard to target and control groups seldom exists due to the encompassing nature of statutes, there is still a lot of variation to be found, e.g. with regard to

- implementation activities (implementation by decentralized or federal units or spontaneous variations due to implementation differences among regional/local government agencies or centrally planned implementation differences);
- observed changes in the addressee's behaviour (e.g. regional, educational, professional subgroups of the target population)
- other countries (e.g. international comparative law studies; more seldom, international policy comparisons).

Combining observation on variation (e.g. combining implementation variation with variation in target subgroup's behaviour, supplemented by international comparisons) can give valuable indications of possible patterns of causation.

Drawing on knowledge development. Sometimes knowledge development has taken place since the preparation of the statute (e.g. theoretical advances, establishment of best practice, well documented success stories). In such cases, the empirical evidence found (e.g. utilizing variation) can be triangulated with the current scientific knowledge.

Combining retrospective evaluation with prospective elements. Evaluations of statutes should not only yield a judgement on their effectiveness but can provide guidance to policy-makers on how to improve the statute or on whether it should be abolished. Combining retrospective evaluation with prospective elements can serve this purpose. It can help to find out where the emphasis of empirical investigation should be: on those findings that are most important for policy decisions. If for instance there is sufficient evidence that a statute is, in general, achieving its purpose at reasonable cost, evaluation efforts will concentrate on exploring cross-sectional variation (e.g. linking differences in implementation with differences in target groups' reaction) in order to improve the implementation process and effectiveness.

The strategy of plausible arguments is intended for decision-making. It should provide the best possible information when a decision is to be made. The strategy lies in the presumption that plausible arguments are better than weak arguments and also better than highly persuasive arguments that arrive too late.

Planning evaluation of statutes. The strategy of plausible arguments may appear to be overly modest. This impression is however deceptive. When taken as a rule for all legislation, it sets a rather high standard. In order to bear fruit, the evaluation of legislation has to be planned in good time. The Swiss manual on legislation⁴ asks for an evaluation concept for all legislation, preferably at the time a bill is presented to parliament and at the latest during the enactment of the statute. In this concept, the information requirements with regard to the evaluation of the statute have to be outlined. If necessary an evaluation has to be planned.

Two recent examples of planned all-encompassing evaluations of statutes. Increasingly, integrated approaches to the evaluation of statutes are being adopted. Evaluation concepts are developed before the statute is enacted. Table 2 lists two examples (studies 7 and 8). The evaluation concepts for the evaluation of the statute on organ transplantations and of the justice reforms are based on similar assumptions:

- A baseline measurement allows for a longitudinal comparison.
- Progress of implementation is monitored. In the case of organ transplantation, a formative evaluation is taking place. In case of the justice reforms, interim reports are submitted two and four years after the enactment of the reforms.
- Summative evaluation takes place six years after the enactment of the statute.

It is expected that the presentation of interim results will lead to fewer demands from members of parliament for an immediate change of legislation (in the case of minor deficiencies) or will allow informed action to be taken (in the case of major deficiencies). The report on the final results should provide a solid assessment of the reforms under scrutiny.

There are several dangers involved in the approach presented. Methodologically, summative evaluation can be distorted by formative evaluation, if the evaluation team takes an active part in the implementation process. Also, the longitudinal comparison on which the evaluation design is based can be distorted by other intervening changes that cannot be foreseen and controlled. Unforeseen political issues may arise and may influence the political process and the utilization of evaluation findings. In the case of the evaluation of the justice reform, additional funding has been reserved for such a case. Being aware of the biggest dangers involved should help to avoid serious errors in this ambitious effort at evaluating legislation.

Conclusion

First, the evaluation of legislation has in the past been ignored in evaluation methodology. One of the main evaluation textbooks (Rossi et al., 2004), for instance, has dropped 'full coverage programs' from the list of contents. Statutes, however, are simply too important and information demands are too high to leave this subject aside. Evaluation of legislation should receive increased methodological attention.

Second, due to their scope and complexity, statutes cannot be evaluated according to the 'gold standard' of evaluation. This is not a free pass for purely impressionist assessments. Having no strict

standards does not mean that any standard should be abandoned. We would argue for a strategy aimed at furnishing plausible arguments for decision-makers at the appropriate moment. Evaluation needs to be interwoven with the legislative process more closely than it has been up to now.

Third, evaluation, according to Scriven (1991: 139), is the process of determining the merit, worth or value of something. In public policy, the merit, worth or value of a project, programme or a statute needs to be assessed with regard to its actual impact on the lives of people, on businesses or on the environment. Many efforts in evaluation methodology are aimed at making the determination of results as convincing as possible. Trying to *prove* effectiveness will, however, not work in the case of many statutes for various reasons (complexity, full coverage and thus no counterfactual evidence, multi-treatment, unclear boundaries, etc.). Members of parliament who increasingly ask for hard evidence on the impact of legislation will not be satisfied. However, there are many information sources on important aspects of statutes, e.g. on implementation activities, on the behaviour of the target population or on counterparts from other countries, that can be used in a combined way. The art and craft of the evaluation of legislation involves making useful comparisons and appropriate analogies (Rose, 1993) between countries, policy instruments, regions, implementation systems and the like. It also involves tracking the 'modus operandi' of the statute and examining contradictory evidence with regard to the various aspects of the statute. Often the evaluation of legislation will not make it possible to make a final judgement on the effectiveness and efficiency of a statute. But it can produce plausible arguments that can help to raise the quality of democratic deliberation.

Fourth, while statutes aim to provide a solid basis for the behaviour of economic or social actors, their rationale empirically often rests on thin ice. Evaluation of legislation cannot fully resolve this paradox. It is hoped it will be able to say where the ice is sufficiently thick to skate on and to advise on alternatives (e.g. limitations in the time or scope of a statute) where the ice appears to be too thin.

Notes

An earlier version of the article was presented at the 2008 EES Lisbon Conference. The author is grateful to two anonymous reviewers for their helpful comments.

1. In December 2009, 1323 evaluations were listed in the ARAMIS database on research at federal level which has been in existence since 1997. Only 107 of them had the word 'statute' in their title as well.
2. Not included in the list are the multiple evaluations of the Swiss Health Insurance Act. The various aspects of the Swiss Health Insurance Act (e.g. complementary medicine, reduction of premiums, planning and lists of approved hospitals, medical tariffs) are covered by whole families of evaluations on each subject, some studies retrospective, other studies prospective or both. A presentation of the evaluation of the Swiss Health Insurance Act would merit an article of its own.
3. Balthasar (2007: 302), based on a survey of all evaluations carried out by the Swiss Federal Administration from 1999-2002, arrived at average costs of 115,000 Swiss francs per evaluation.
4. Last chapter on evaluation, accessible under <http://www.gl.admin.ch> (in German and French)

References

- Balthasar, A. (2000) *Energie 2000: Programmwirkungen und Folgerungen aus der Evaluation*. Chur and Zürich: Rüegger.
- Balthasar, A. (2007) *Institutionelle Verankerung und Verwendung von Evaluationen*. Zürich and Chur: Rüegger.
- Balthasar, A. and S. Rieder (2000) 'Learning from Evaluations: Effects of the Evaluation of the Swiss Energy 2000 Programme', *Evaluation* 6(3): 245–60.

- Böhret, C. and G. Konzendorf (2001) *Handbuch Gesetzesfolgenabschätzung (GFA)*. Baden-Baden: Nomos.
- Bussmann, W. (2008) 'The Emergence of Evaluation in Switzerland', *Evaluation* 14(4): 499–506.
- Chen, H.-T. (1990) *Theory-Driven Evaluations*. Newbury Park, CA: SAGE.
- Davidson, J. E. (2005) *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: SAGE.
- Dery, D. (1984) *Problem Definition in Policy Analysis*. Lawrence, KS: University Press of Kansas.
- European Commission (2008) 'DG MARKT Guide to Evaluating Legislation', unpublished document, Brussels.
- Gysen, J., H. Bruyninckx and K. Bachus (2006) 'Modus Narrandi: A Methodology for Evaluating Effects of Environmental Policy', *Evaluation* 12(1): 95–118.
- HM Government (2008) *Post-Legislative Scrutiny: The Government's Approach Presented to Parliament by the Lord Privy Seal, Leader of the House of Commons and Minister for Women and Equality by Command of Her Majesty*. URL: http://www.lawcom.gov.uk/docs/lc302_government_response.pdf.
- Karpen, U. (2002) *Evaluation of Legislation*. Baden-Baden: Nomos.
- Klein Haarhuis, C. M. and E. Niemeijer (2009) 'Synthesizing Legislative Evaluations: Putting the Pieces Together', *Evaluation* 15(4): 1–23.
- Law Commission (2006) *Post-Legislative Scrutiny: Presented to the Parliament of the United Kingdom by the Secretary of State for Constitutional Affairs and Lord Chancellor by Command of Her Majesty* (LAW COM No. 302). URL: <http://www.official-documents.gov.uk/document/cm69/6945/6945.pdf>.
- Mader, L. (2001) 'Evaluating the Effects: A Contribution to the Quality of Legislation', *Statute Law Review* 22/2: 119–31.
- Mader, L. (2003a) 'Evaluating the Effects of Legislation: – The Current Situation in Switzerland', *Legislação, Cadernos de Ciência de legislação* 33/34: 85–106.
- Mader, L. (2003b) 'Zum aktuellen Stand der Gesetzesfolgenabschätzung in der Schweiz', in U. Karpen and H. Hof (eds) *Wirkungsforschung zum Recht IV*, pp. 96–110. Baden-Baden: Nomos.
- Mark, M. and G. Henry (2006) 'Methods for Policy-Making and Knowledge Development Evaluations', in J. F. Shaw, J. C. Green and M. M. Mark (eds) *Handbook of Evaluation: Policies, Programs and Practices*, pp. 317–39. London: SAGE.
- Mayne, J. (2001) 'Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly', *Canadian Journal of Program Evaluation* 16: 1–24.
- Mayne, J. (2008) 'Contribution Analysis: An Approach to Exploring Cause and Effect', *ILAC Brief* 16 (May): 1–4.
- Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. Newbury Park, CA: SAGE.
- Rose, R. (1993) *Lesson Drawing in Public Policy: A Guide to Learning across Time and Space*. Chatham, NJ: Chatham House.
- Rossi, P. H. and H. E. Freeman (1993) *Evaluation: A Systematic Approach* (5th edn). Newbury Park, CA: SAGE.
- Rossi, P. H., M. Lipsey and H. E. Freeman (2004) *Evaluation: A Systematic Approach* (7th edn). Thousand Oaks, CA: SAGE.
- Schäffer, H. (2005) *Evaluierung der Gesetze/Gesetzesfolgenabschätzung in Österreich und im benachbarten Ausland*. Vienna: Manz.
- Schäffer, H. (2007) *Evaluierung der Gesetze/Gesetzesfolgenabschätzung* (II). Vienna: Manz.
- Scriven, M. (1974) 'Maximizing the Power of Causal Investigations: The Modus Operandi Method', in W. J. Popham (ed.) *Evaluation in Education: Current Applications*, pp. 68–84. Berkeley, CA: McCutchan.
- Scriven, M. (2005) 'Causation', in S. Mathison (ed.) *Encyclopedia of Evaluation*, pp. 43–7. Thousand Oaks, CA: SAGE.
- Scriven, M. (1991) *Evaluation Thesaurus* (4th edn). Newbury Park, CA: SAGE.
- Sverdrup, S. (2003) 'Towards an Evaluation of the Effects of Laws: Utilizing Time-Series Data of Complaints', *Evaluation* 9(3): 325–39.

- Stame, N. (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation* 10(1): 58–76.
- Virtanen, P. and Usikylä P. (2004) 'Exploring the Missing Link between Cause and Effect: A Conceptual Framework for Understanding Micro–Macro Conversions in Programme Evaluation', *Evaluation* 10(1): 77–91.
- Weiss, C. H. (1998) *Evaluation* (2nd edn). Upper Saddle River, NJ: Prentice Hall.
- Weiss, R. S. and M. Rein (1970) 'The Evaluation of Broad-Aim Programs: Experimental Design, its Difficulties, and an Alternative', *Administrative Science Quarterly* 15(1): 97–109.
- Yin, R. K. (2009) *Case Study Research: Design and Methods* (4th edn). Thousand Oaks, CA: SAGE.

Werner Bussmann is senior advisor in the Legislative Projects and Methodology Unit of the Swiss Federal Office of Justice (Bundesrain 20, 3003 Bern, Switzerland). His research interests include the institutionalization of evaluation, prospective evaluation methods and the utilization of evaluations for legislative purposes. [email: werner.bussmann@bj.admin.ch]