Blood, sweat, and cannabis: real-world policy evaluation of controversial issues

Céline Mavrot, Susanne Hadorn & Fritz Sager

Published online: 18 Jun 2023.

Submit your article to this journal ⬚

View related articles ⬚

View Crossmark data ⬚

Routledge
Taylor & Francis Group

Blood, sweat, and cannabis: real-world policy evaluation of controversial issues

Céline Mavrot ⬡[a], Susanne Hadorn ⬡[b] and Fritz Sager ⬡[b]

aInstitute of Social Sciences, University of Lausanne, Lausanne, Switzerland; bKPM Center for Public Management, University of Bern, Bern, Switzerland

ABSTRACT
The motivation of this article is to address the ambivalent position of policy analysis when it intervenes in the real-world policy process through policy evaluation. It tackles the underresearched question of the challenges faced by policy analysis in relation to applied research mandates. It argues that policy analysis is constantly at risk of instrumentalisation by politico-administrative players. The article is based on the evaluation of the medical cannabis policy in Switzerland as a case study. The results point out four specific challenges faced by applied policy analysis: political pressure, scientific integrity, access to sensitive data, and epistemic legitimacy. However, applied policy analysis can contribute to de-escalating controversies by presenting a bigger and contextualised picture of the considered political issues. Policy evaluation can identify deficient implementation processes, but also wider mismatches among legislative and societal processes. Hence, although evidence is subordinated to other factors in the decision-making process, evaluations provide an outside perspective, which can help solving controversies around policies. The article contributes to the literature on the politics of policy analysis by showing that confronting policy analysis with practical problems brings both scientific and policy benefits.

Introduction

Although policy processes are its core object of study, policy analysis tends to have an ambivalent stance towards real-world policymaking. On the one hand, distance is considered as a condition of objectivity in the analysis, while on the other hand, some research streams such as policy evaluation

see their role as providing feedback to policymakers (concerning this last point, see for instance Wollmann, 2007). This article argues that policy evaluation as application-oriented policy analysis can contribute to de-escalating controversies by presenting a bigger and contextualised picture of issues at hand. In other words, policy analysis can help overcome existing political obstacles and contribute to effective policymaking. By considering the whole policy process from an overarching perspective, policy scholars can help policy practitioners fight against memory loss and institutional amnesia (i.e., repeating the same errors) that impedes policy learning and service delivery (Stark & Head, 2019; citing Pollitt, 2000). In addition, by discerning programme failure from implementation failure (Linder & Peters, 1987, p. 29), policy analysis can also move a controversial policy issue to the correct political decision layer. Hence, applied research can bring crucial policy benefits and help find practical solutions to policy issues thanks to its analytical tools. In turn, applied research projects aimed at tackling concrete problems also lead to important scientific benefits by providing policy scholars with a privileged access to data on timely policy issues.

This article offers a contribution to the reflection on 'the politics of policy analysis' (Cairney, 2023; see also Dorren & Wolf, 2023) by delving into the real politics of applied policy analysis based on a concrete example that is representative of the challenges and contributions of policy evaluation. It identifies the main challenges faced by policy scholars in their applied research endeavours – regarding independence, integrity, access to data and legitimacy, and proposes to strengthen the dialogue between theory- and praxis-oriented policy analysis communities.

Although policy evaluation is a long-established field of study (Alkin, 2004; Pawson & Tilley, 1997: Rossi *et al.*, 1999; Scriven, 1991) that has accumulated a lot of know-how on putting political science at the service of public action, it remains somewhat disregarded in the wider policy analysis field. Policy analysis is the theoretical research stream in which scholars decide on their own research programmes without any connection to the politico-administrative agenda. In contrast, policy evaluation is defined here as applied research pertaining to the wider family of policy analysis, but in which research is commissioned by organisations and agencies according to the needs of the moment. The different properties root in the distinctive disciplinary backgrounds that have generated policy analysis and programme evaluation. Policy analysis stems from economics that still claims the label while the origins of evaluation lie in education research with the explicit aim for practical learnings (Stufflebeam & Shinkfield, 2007). As a consequence, we posit policy evaluation is applied research assessing the effects of programmes or public policies. The commissioned and applied nature of policy evaluations as well as their proximity to politico-administrative authorities tend to attract suspicion. However, we hold that an enhanced dialogue between applied and

fundamental policy analysis could benefit both our knowledge on policy action and the potential real-world effects of policy analysis.

This paper discusses the role of policy evaluators in disputed (or contested) policies on the basis of an illustrative case study on cannabis prescriptions for medical purposes in Switzerland that the authors conducted themselves. The policy on medical cannabis triggered polarised parliamentary debates amid an acute conflict between two groups of public agents in charge of the policy's implementation: physicians and legal experts from the national public health agency. The legal experts were in favour of a restrictive interpretation of the law and accused the public agency's physicians of liberalising the medical use of cannabis against the political will. This case is well suited to the study of the relationship between policy analysis and policy praxis because it concentrates the typical pitfalls of applied research, while also showing where its contribution might lie. The evaluation team based its investigation on Michael Patton's utilisation-focused theory (1997), which insists on the need to identify solutions that are viable in a real-world setting. The policy evaluation showed that the main issue was not located within the public agency's praxis, but in a mismatch among legislative, medical, and societal processes.

In this regard and without overestimating the role of evidence in the democratic process (Parkhurst, 2017), we hold that policy evaluation as an applied form of policy analysis can contribute to de-escalate policy-making controversies by providing a dedramatised account of policy processes in two ways:

- It provides a balanced and comprehensive analysis of policy design and implementation from different perspectives, while stakeholders are trapped in their own rationality.
- It can reconstruct the history of the policy sub-sector against institutional amnesia and through that provide an explanation of the ambiguities or shortcomings of the temporal developments of a policy, while stakeholders are focused on the present.

The importance of a contribution of policy analysis to de-escalation can not be stressed enough, because 'when controversies and struggles dominate the policy process, knowledge can gain access to the political debate only as argument' (Radaelli, 1995, p. 176). Hence, de-escalation is a key element to realistically aim at the designing and implementing of genuine 'evidence-informed policies' (Topp *et al.*, 2020). In this paper, we look at what applied policy evaluation makes to policy and the other way around. In the next section, we contrast policy analysis as basic research from policy evaluation as an applied version of policy analysis and derive four challenges for policy evaluation that result from its status as commissioned research. These four challenges

form our analytical framework. We then present our research design as well as our test case based on which we discuss the four challenges for policy evaluation. We thereafter substantiate our argument that applied policy analysis can unlock debates on contested issues and lead to policy change. The conclusion addresses the generalisability of the argument and discusses lessons for both policy analysis and policy practice.

Four challenges of policy evaluation as applied policy analysis

Policy evaluation is a distinct form of research with idiosyncratic features that can lead to specific challenges in political practice. In the following, we define policy evaluation as applied policy analysis and derive four interrelated challenges: the challenge of scientific independence and integrity; the challenge of political pressure in cases of contested policy problems; the challenge of accessibility of sensitive data; and the challenge of epistemic legitimacy and political acceptance. These four key challenges are identified deductively; the scientific literature shows that they are typical of policy evaluations. We do not address here challenges that endanger the realisation of evaluations themselves, such as the lack of resources or of data. We employ the four challenges to study our empirical case.

What is a policy evaluation and why bother?

Policy evaluation is the empirical, transparent, and replicable scientific assessment of the effects of a policy intervention usually commissioned by an actor other than the evaluator (Sager *et al*., 2021). Following this definition, evaluation seems pretty much the same as policy analysis that also systematically and empirically studies the effects of a policy. However, there is a core difference in the evaluation 'assessment'. Evaluations have the explicit objective to pass judgement on the policy they study. While policy analysis focuses on the causality between policy and effect, evaluations provide normative information about this effect. Cognitive philosopher David Hume (1978) (1711–1776) famously introduced the distinction between 'Is' and 'Ought'. While the 'Is' refers to cognition and can be rationally achieved by scientifically sound methods of empirical data collection and analysis, it is not possible to rationally deduce an 'Ought' from the 'Is'. This means that only because science empirically establishes causality between human behaviour and climate change ('Is'), we cannot rationally deduce the need for action against this behaviour. Translating Hume's distinction into systems sociology, the 'Is' arguably belongs to the realm of science while the 'Ought' belongs to the realm of politics.

In the case of policy evaluation, the distinction between 'Is' and 'Ought' helps us understand its position at the intersection between science and

politics. While policy analysis is interested in the 'Is' when studying causality between intervention and effect for the causality's sake, i.e., to develop theory, policy evaluation contains a share of 'Ought' when using the found causality to make a normative statement whether the policy is successful or not. While both policy analysis and policy evaluation employ the same methodological toolkit and are subject to the same scientific standards, policy evaluation needs normative criteria to assess its findings. These criteria most commonly stem from the sphere of politics (Sager & Mavrot, 2021). The criteria can be derived from the evaluated policy itself (Does the policy achieve its own objectives?), given by the commissioner (Does the policy meet demands that are not part of the statutory policy decision but deemed relevant by the commissioning party of the evaluation?), or defined by interested third-parties in case of participatory policy evaluations (Does the policy meet criteria of social desirability, e.g., equity, sustainability).

Given that policy evaluation employs normative criteria set by external actors who have direct political interest in the evaluation results and at the same time commission the evaluation, policy evaluators find themselves in a demanding situation. Their analysis not only must meet scientific standards, but also satisfy their commissioner's information needs. However, the case of contested policies is challenging. Evaluation results are relevant to a wide range of actors with different stances towards the policy and the problem it is supposed to solve. Hence, policy evaluation has a distinctive added value because it has been forced to elaborate related strategies: the triangulation of data, interdisciplinarity, a reflection on evaluation criteria, the process of synthetising data to arrive to a sound conclusion, resistance to instrumentalisation on the field. The disagreement about the policy may not only regard actors inside and outside the political-administrative nexus (such as interest groups, parties, NGOs, or political movements) but also the commissioning agency itself. This is the reason why we talk here of contested policies. The term contested policy captures a case in which the policy is disputed within the responsible administrative agency, which results in conflicts between professional roles, norms, and ethos. A polarised policy depicts a wider societal and/or political dispute, that aligns itself on political cleavages (e.g., partisanship, leading to politicisation) or on social cleavages (Lagroye, 2003; Rennes, 2016). Hence, a contested policy can be polarised when the debate spills over to the media or political arenas (as was the case here at some point), but can also remain contested within more confidential arenas such as the administrative ones.

In this article, we focus on a case of ex-post (i.e., conducted some time after the implementation) and external (i.e., conducted by external experts, in this case academics) policy evaluation aimed at learning. There are other types of policy evaluations. For instance, regarding the temporality, ex-ante evaluations assess the possible risks and effects of a future policy before the

implementation. Regarding the aims of evaluations, financial or performance audits focus more on reviewing than learning processes, and adopt a narrower, often budgetary, perspective. Regarding the nature of the evaluator, self-evaluations are conducted internally by the concerned organisations themselves; there are also evaluations conducted by agencies that have a special role in the state architecture, such as parliamentary committees that review the activities of the executive branch (Fischer *et al.*, 2007). However, the challenges highlighted here – political pressure, scientific integrity, access to sensitive data, and epistemic legitimacy – mostly apply to these other types of evaluations, though with some variations regarding the epistemic legitimacy. From a system perspective, policy evaluation has acquired a consequent importance in policy-making processes. This is for instance the case in Switzerland where the administration is weak and the political system is a non-professional one (no full-time politicians) (Ladner & Sager, 2022, p. 323). Policy evaluation provides policy-relevant information and therefore is part of the policy politics. For policy evaluators as scientists, this situation leads to challenges that we discuss in the following paragraphs. We will use these challenges as an analytical grid for our case study that we present in the next section.

The challenge of political pressure in cases of contested policy problems

The question of independence stems not only from the fact that evaluations are commissioned. Evaluations provide policy-relevant information and hence are of political relevance for all actors involved, which makes pressure from various stakeholders a key ethical challenge in compiling such reports (Morris, 1999, 2007). Pleger *et al.* (2017) found in a meta-analysis that between 42 and 83 per cent of the respondents of respective surveys in the USA, UK, Germany and Switzerland have been subject to pressure from the outside to change their evaluation results. The pressure mainly stemmed from the commissioning agency and took on various degrees. There are different forms of pressure, ranging from the demand to change certain formulations over requests to change interpretations of results to demands to skip whole sections of the reports or threats not to publish the report altogether (Morris & Clark, 2013; Schmidli *et al.*, 2023). Requests to distort findings, ignore data, use invalid data, exclude sources or substantially change conclusions are highly problematic. The main motivation of such attempts is political. Even if the pressure stems from the commissioning agency, they are not necessarily the source of the demands that may come from their superiors and the agency only is the messenger.

It has been argued that in the case of sensitive or in other words disputed policy problems, political pressure is more likely to occur than in ordinary

daily-business evaluations (The LSE GV314 Group, 2013). For evaluators, being confronted with political pressure is an important challenge due to the limited options they have to counter it. Politics takes place (at least partly) in public, evaluation does not. Evaluators therefore cannot address the pressure during the evaluation process by making it public. Within their contract, evaluators are in a defined relationship with their commissioner that they must not violate. Evaluators face the challenge to cope with potential attempts to influence their work when a policy is controversial. Contested policy issues can thus build a strained working context for evaluators.

The challenge of scientific independence and integrity

It is impossible to imagine policymaking today without the use of experts. Nevertheless, some scholars demand that the influence of experts on policymaking should be viewed more critically from a democratic perspective: 'Too often still, experts are seen as individuals possessing special skills or superior knowledge applicable to predetermined domains of decisionmaking; the experts' political power to define the issues and select the very terms of deliberation has received too little notice' (Jasanoff, 2003, p. 162). This quote shows some of the challenges also policy evaluation is confronted with. Evaluation is commissioned research. This means that there is a tension between two ideals that are not always easy to reconcile: the ideal of basic science (including objectivity and independence) on the one hand and the ideal of customer or user orientation on the other (Klerman, 2010; Sager & Mavrot, 2021).

In the evaluation literature, the debate is virulent with two corresponding strands. Carol Weiss (e.g., 1977) was the mascot of the independence coalition, while Michael Quinn Patton (1997) is the inventor and most famous proponent of 'Utilization-focused evaluation'. The independence argument states that utilisation of evaluations is not the responsibility of the evaluators. The evaluators need to produce the best possible answer to the commissioner's questions, i.e., scientifically sound empirical studies. Evaluators thus seek to produce their report as autonomous as possible after clarification of initial questions. Once the evaluators have submitted their report, they no longer have control over what happens with it. The utilisation-focus argument disagrees. The sole reason evaluations exist in the first place is that they inform practice. Unlike basic research that finds its own research questions and has the purpose to contribute to theory, evaluation is not self-sufficient. Utilisation is the essential feature of evaluation in this strand of literature (Giel, 2013). Consequently, it must be the main goal of evaluators that the commissioners deem their findings useful and ultimately use them.

The two ideals can collide in cases of disputed policy problems. Customer orientation may lead to an overweight of the expectations of the

commissioner in the analysis of the data. Evaluators must decide whether they seek the confrontation with their clients or whether and how far they are willing to compromise on certain points. In any case, the tension potentially looms whether evaluators opt for either independence or utilisation-focus. Reliance on strong evaluation norms and scientific standards help resolve this dilemma.

The challenge of accessibility of sensitive data

A minor but also important challenge is data availability. Evaluation is empirical. Besides original data collection like surveys and interviews, the data sources for evaluations often are internal data sets from the agencies that implement the evaluated policy and underlie strict data protection rules (Trivellato, 2019). These sources can be documents that evaluators need for content analysis or they can consist of datasets the agency collects for management purposes (Nielsen & Ejler, 2008). While policy documents are not problematic as they are mere empirical material for content analysis, datasets collected by the agency are not tailored for scientific-analytical purposes and often need re-coding. This includes additional interpretation by the evaluators with which the agency may disagree. In addition, the preparation and provision of data can be a burden for the involved agencies, which in some cases requires an adaptation of data collection strategies to maintain commitment of the different stakeholders (Dekker *et al.*, 2021). Thus, it has been argued that such administrative data is often not optimally exploited for the purpose of evaluations (Lyon *et al.*, 2015). However, such disagreements and challenges are evaluators' daily business and can be resolved.

In the case of contested policies, the negotiation about the interpretation of policy management data can become more difficult when the data owner fears negative consequences from a re-interpretation of the data. Additionally, policy data tend to be sensitive in many policies. When personality rights are affected, data availability can become a major issue. Policy agencies may refuse to grant access to sensitive data if they have reasons to believe that the evaluation may lead to violation of data protection rights. Evaluators have to cope with the tension between the need for a sound empirical basis in order to fulfil the contract and the challenge to convince data owners of their responsible handling of the data.

The challenge of epistemic legitimacy and political acceptance

Michael Scriven (1991) labels evaluation as a 'trans-discipline' meaning that it crosses classic disciplinary borders and is not limited to a given disciplinary silo, neither methodologically nor theoretically. This characteristic is a blessing and a curse for evaluators at the same time. It is a blessing in that it

allows evaluations to be pragmatic in the design and implementation of their research. It is a curse in two ways: first, in the scientific community, evaluators as pluridisciplinary and applied researchers often struggle for epistemic legitimacy as scientists. Academic basic research at times has difficulties to acknowledge applied research as equivalent, which undermines the scientific legitimacy of evaluators (Ritz, 2003). It has been argued that researchers in such situations face an inherent tension between 'a desire to work critically whilst maintaining a level of credibility with policy and funding audiences' (Smith, 2010, p. 189). Second, the transdisciplinary status of evaluation leads to tedious discussions with both policy practitioners and policy experts whether evaluators are capable of evaluating a given policy if they are evaluation experts but not field experts in the evaluated policy (e.g., Levin-Rozalis, 2010). While this reservation can be countered with the analogy of the veterinarian who needs not be a cow to diagnose a cow's disease, it still is relevant for evaluation practice. Doubts against evaluators' competence undermine this acceptance that is a necessary condition for credibility and utilisation of results. Research shows that clients' reservations about skills of the evaluation teams are a key cause of conflicts in policy evaluations (Schmidli *et al.*, 2023). Evaluators therefore face a double challenge. One, they must convince academia that their research meets the same quality standards as basic research does to draw from the symbolic capital of scientific credibility. Two, they also have to prove that they command the competences necessary to evaluate this policy, even though they are evaluation experts rather than policy sector specialists. In the case of contested policy issues, both challenges can intensify when policy stakeholders activate these reservations to question the validity of evaluation findings they may disagree with. In the remainder of this paper, we use these four tensions for evaluators of contested policies as an analytical grid. We present the research design and the empirical case in the next section.

Research design, methods, and data: a policy evaluation before the storm

We will analyse the tensions outline above by means of a single case study: the evaluation of the legislation on cannabis prescriptions for medical purposes in Switzerland conducted between 2017 and 2018.[1] The policy (as evaluated) allows patients with specific diagnoses such as multiple sclerosis to consume cannabis for medical reasons to relieve their suffering. This policy exists in the context of the otherwise prohibited use of this substance in Switzerland. To be able to legally consume medical cannabis, the physicians of the respective patients have to apply for an exemption permit at the Swiss Federal Office of Public Health (FOPH) – the national public health agency, where the application is reviewed individually before a decision is made.

Importantly, the term 'exemption permit' is a legal concept implying that the total number of permits given in the scope of the policy has to be limited to specific cases and the granting of authorisation is not systematic. The evaluation of the policy was mandated by the FOPH in the context of significant tensions between two groups of professionals – physicians and jurists – who are the public agents responsible for policy implementation within the FOPH. We argue that this specific evaluation is an ideal case to analyse the relevance of applied policy analyses, i.e., policy evaluations, in the context of contested policy. Notably, because the policy controversy has not only taken place in the political sphere, but has spilled over into the implementation arena, making the evaluation an emblematic case of contested policies both at the design and implementation levels. This single case study helps us to unravel causal mechanisms (Gerring, 2007), and allows us to illustrate the relevance of the four above-mentioned challenges and how applied policy analysis can be an instrument to solve otherwise deadlocked controversies around policies.

Additionally, this evaluation is particularly suitable to shed light on the potential that applied policy analysis can have in the context of contested policies, since it has been built on an extremely robust set of data. A solid data set is even more important in the analysis of contested policies because it reduces the likelihood of an attack on the credibility of the evaluation. In such situations, mixed-method approaches help us to improve our understanding of the subject under study (Creswell, 2003) and allow us to 'find solutions to practical, real-world problems' (Riccucci, 2010, p. 109). The cannabis evaluation exploited these advantages by basing its assessments on various methods including interviews, a survey, a qualitative and quantitative document analysis as well as a context analysis as displayed in Table 1.

First, we conducted 21 interviews with different stakeholders including both the physicians and lawyers involved in the policy implementation (mid-level bureaucrats of the national public agency) as well as their managers (top-level bureaucrats in a supervising function). To also be able to trace the development over time and the changing context in which the policy had been implemented, former employees of FOPH were also interviewed. Additionally, stakeholders external to the national administration were interviewed, such as the producers of cannabis-based medicines, the cultivator of the cannabis grown in Switzerland for medical purposes, and representatives of the subnational administration (holding implementation tasks in this policy area). Second, we invited 1015 referring physicians who had at least once applied for an exemption permit for one of their patients to participate in an online-survey (response rate: 34.8 per cent; number of answers registered: 353). This survey provided us with an external perspective on the implementation process of the FOPH and gave us insights on the reasons for the steadily increasing numbers of applications.

Table 1. Methods and material – policy evaluation on medical cannabis regulation.

Method	Scope	Source	Information received
Interviews and qualitative document analysis	21 interviews Internal administrative documentation	Current and former implementing agents (physicians and lawyers); high-level bureaucrats in a supervising function; external stakeholders (cultivators of cannabis; producers of cannabis extract; pharmacies allowed to deliver the product; representatives of sub-national units); medical experts	Insights on the implementation processes Motives behind decisions within implementation process Developments regarding production of medical products Division of work, workflows, administrative processes
Survey	353 responses	Physicians applying for exemption permits for their patients	Effects of medicine on patients Reasons behind increasing number of applications External view on FOPH-processes
Quantitative analysis	8400 applications	Quantitative analysis: all applications for exemption permits since the policy came into force All internal documents relevant to application-granting	Information on patterns in the diagnoses allowed for an exemption permit
Context analysis (media analysis and analysis of parliamentary debates)	Time period analysed: 2000–2017	Media reports Parliamentary debates Minutes of parliamentary commission meetings	Information on the initial considerations of policy makers when adopting the policy (e.g., regarding the expected development in this area) Information on the development of the societal debate on the use of medical cannabis
Legal opinion	Initial formulation of the law and development of the implementation praxis	All data Legislation Doctrine on medication law and on psychotropic products	External opinion on the legality of the implementation praxis

Third, the evaluation encompassed a large document analysis, including documents related to the internal administrative processes to examine aspects such as the collaboration between physicians and lawyers within the FOPH. Moreover, a key element of the evaluation was the quantitative analysis of all applications reviewed by the FOPH in the years since the policy came into force. In total, 8400 applications were analysed, in particular to identify patterns in the diagnoses for which exemption permits had been

granted over time. The main interest here was whether there had been an expansion of the diagnoses accepted for exemption permits over the years. Finally, the evaluation was complemented with a context analysis, including a media analysis and an analysis of the parliamentary debates related to the topic during the time period 2000 to 2017. This helped to trace back the initial motives and considerations of policy makers when adopting the policies. The resulting rich data set allowed us to examine the implementation of the policy not only in view of the different perceptions of the actors involved, but also in consideration of the changing context such as the increased societal demand for medical cannabis as well as an objective presentation of the implementation practice over the entire period (through the quantitative examination of the applications).[2] The evaluation team made its study in accordance with evaluation ethics and best practice, based on the literature (Morris, 2007) and the standards of the Swiss Evaluation Society[3] and had no preference for any policy option. Its members' objective was to identify the reasons of the growing misfit between the legislation and the implementation praxis, and to highlight solutions to resolve these contradictions. However, the team deployed various strategies to avoid the instrumentalisation of the evaluation (see case study below).

The data gathered and analysed as described above however forms the basis of the policy evaluation process itself (i.e., how the evaluation team assessed the policy in the frame of its mandate). In this article however, we take a step back and propose an analysis of the challenges faced with applied policy analysis at a meta level. In this analysis, we include the power relationships, strategic plays, and negotiations in which evaluators are involved in every step of the assessment process. In this sense, the present analysis is based on the qualitative methods of participative observation in this administrative setting. The data includes personal notes and other written traces of the evaluation process, in particular emails and minutes of meetings with the evaluation commissioners (high-level administrative agents) and the policy implementers (mid-level administrative agents) and with whom we had continuous contacts during the evaluation (kick-off meetings; interviews; discussions about several aspects such as data transmission, the selection of interviewees, and the evaluation design; intermediate and final discussions of the results). In the following, we will analyse the four tensions identified in the theoretical section of this paper by means of the here outlined data.

Case study: policy evaluation at grips with political and administrative controversies

In this section, we analyse if and how the four challenges presented earlier have manifested in practice in the evaluation of the legislation on cannabis

prescriptions for medical purposes in Switzerland. We also examine how the evaluation team addressed these challenges to ensure the scientific validity of the evaluation.

The challenge of political pressure in cases of contested policy problems

As is often the case for policy evaluations (Bovens *et al.*, 2008), the background of the study was heavily contested since the beginning. The 'prevention of non-communicable diseases' division of the Swiss Federal Office of Public Health (FOPH) decided to commission external academic evaluators for this policy evaluation in the wake of the strong dispute between civil servants about the implementation of legislation on medical cannabis. Since 2012, the law in Switzerland has authorised the exceptional use of medical cannabis for patients on an individual basis, which is subject to special authorisation by the FOPH. The referring physicians of the patients (mainly oncologists, neurologists, and general physicians) must submit an application for such authorisations. Once accepted, the patients receive ad hoc produced *magistral formula* from one of the two pharmacies authorised to deliver cannabis-based medication in the country. The main medical bases for cannabis use include neurological diseases (e.g., Parkinson, multiple sclerosis), pain related to cancers or chemotherapy side effects, and chronic pains. The applications are reviewed by a team of physicians from the FOPH in collaboration with a team of jurists from the same office (all civil servants). However, the agency's physicians and jurists are inserted in different sections and hierarchical lines within the office. The exact division of tasks between them is not clearly defined, especially in cases of disagreement. Theoretically, the physicians examine the validity of the requests, and the jurists overview the legality of the whole process. However, according to one of the jurists, his team is not authorised to give instructions to the physicians, which is frustrating because repeated warnings from the legal unit are: 'when disaster then strikes [such as a legal dispute with applicants], the legal service has to pay the price'.[4]

A strong dissent quickly rose between the physicians and the jurists as the number of granted requests sharply increased from the beginning of the enforcement of the law in 2012 (291 accepted requests) and the time of the study (2309 accepted requests in the first nine months of 2017). While the physicians declared they had kept their authorisation praxis constant (applying the same criteria for each request), the jurists accused them of (increasingly) loosely granting the requests without doing an effective selection. In the jurists' eyes, this meant that the exceptional system foreseen in the law[5] was becoming an automatic authorisation praxis, in which the individual examination of the applications was only conducted pro forma.

Moreover, they accused the teams of physicians of 'complete ignorance and aversion to any legal concerns'.[6] The hierarchy for each of the two professional groups of civil servants were also involved in the dispute, each backing up its own team. Besides the explosive situation within the public health agency, several motions had been submitted in the National Parliament regarding the issue of medical cannabis in the past few years. Some had accused the FOPH of dilettantism, while others had requested that the legal system for access to medical cannabis be simplified. Related parliamentary debates pushed the FOPH to commission the policy evaluation to put the house in order and to be prepared for the spotlight that was going to be shone on this topic.

Between 2004 and 2017, 38 parliamentary objects were submitted and discussed on the topic of cannabis at the national level. In 2008, 63.3 per cent of the Swiss population voted against cannabis depenalisation. Although these political debates mainly focused on the recreational use of cannabis, they also had a strong impact on the debate on medical cannabis. Those who oppose the medical use of cannabis argue that this represents a first step toward a generalised cannabis legalisation. The jurists insisted that the public health agency had to respect the initial will of the parliament to restrict the use of cannabis to exceptional situations when the legislator introduced the possibility of a medical use in the Federal Act on Narcotics in 2012. For them, granting too many requests would go against this will and be a breach in the rule of law, as civil servants cannot substitute politicians as rule-makers. For the FOPH's physicians, the core criteria was to grant authorisations to any patient who met the legal and medical requirements, regardless of the number of applications. In theory, the law required FOPH physicians to exhaustively review the medical history of the patients for whom a request had been submitted, and to verify that all existing medical alternatives had already been tried. De facto, the agency's physicians claimed to not have the resources to proceed to such background verification and trusted the referring physicians in their assessments.

Hence, the policy evaluation happened in a highly polarising situation. Within the public health agency, it was an open war, and even for the experienced policy evaluators, the tensions were unprecedented. The interviews we led with FOPH players were particularly emotional, with cross-accusations from the two teams; each claimed that the other lied and concealed elements from the investigation. In the course of the inquiry, interview partners regularly revealed new documents to the evaluation team and wanted to show private emails as proof for their accusations. We had to refuse to take such data into account, as an agreement had been passed that the inquiry would transparently rely on a set of approved data known by both teams. However, this context also led to rich interviews that lasted up to six hours and allowed us to get to the bottom of the process. Moreover, the

importance of the final assessment of the situation by the evaluation led all players to be highly willing to participate and share their side of the story. This is linked with both advantages and drawbacks; on the one hand, this provided privileged access to highly actual and confidential data, but on the other hand, this meant all players constantly pressured the evaluators. The jurists hinted that if we did not denounce the breaches to the law, we would be covering up an illegal authorisation praxis. The physicians suggested that if we assessed the situation too severely, we would contribute to putting needy patients at risk of not receiving their medication. In addition to that was an important time pressure; the head of the FOPH wanted the results of the study to be ready in a short timeframe for the upcoming parliamentary debates (discussion of several motions). To protect the evaluation team from political and organisational pressions in this conflictual context, we implemented some strategies such as organising several kick-off and follow-up meetings with representatives of both teams of public servants to ensure plenary discussions, a validation of the evaluation protocol by all parties, and engaged ourselves to make the evaluation report publicly available in the end of the process to ensure a transparent display of the procedure through which our final conclusions would be derived. Investigator triangulation (Archibald, 2016) was also crucial to ensure a critical distance with the accusations and claims of each involved parties. As explored in the next section, the pressures on evaluators had to be further countered through tight negotiations and the implementation of a sound research design.

The challenge of scientific independence and integrity

To take distance with the controversial context, we invested in negotiations with the public health agency (commissioner of the evaluation) to consolidate three dimensions of the study. Firstly, we devoted great attention to the triangulation of data collected from the key informants with whom we conducted interviews. We made sure that all stakeholders were involved in the interview module of the inquiry, while all involved parties (the physicians and jurists teams), with whom we were in contact during the preparatory meetings, tried to restrict the involved interviewees to their advantages. To ensure a solid corroboration of the data, the comprehensive inclusion of key informants included, for instance, the secretaries working for the two teams (who were also involved in the management of the requests), former employees of the two team (civil servants who had been involved in the process before switching positions), and other players of the system (e.g., the two pharmacies that produce the *magistral formula*, laboratories that extract cannabis oil, a specialised physician, the direction of the FOPH).

Secondly, despite the time pressure, we extensively negotiated the research design of the study. Again, the objective was to ensure comprehensive source- and method-triangulation to enhance the quality of the data and to avoid any accusations of partiality that would have undermined the legitimacy of the study. As described, a five-module analysis was designed that included a study on the media coverage and of parliamentary debates, an organisational analysis of the workflows within the FOPH, an online survey among prescribing physicians, a quantitative analysis of the applications, and a legal expertise. Two of the modules especially raised debate within the agency: the online survey among prescribing physicians and the quantitative analysis of the applications (see below). However, we held that these modules were essential to obtain a 360 vision of the problem because they thoroughly documented the rationale of prescribing medical cannabis from the physicians' point of view.

Thirdly, we opted for extensive coding of the symptoms and indications for which the FOPH had approved medical cannabis since the beginning of the system. We did so by using the international ICD-10 code system (International Classification of Diseases of the World Health Organization) to ensure a systematic assessment. We coded all requests ($N = 8400$, 2012–2017). The results of this coding module showed that the main reasons for granting requests remained stable over the years, and that the increase in authorisations was related to the increase in requests. The combined examination of the findings from the various modules allowed us to go beyond the declarations of the key players in interviews and reconstruct evolving trends, such as the growing mediatisation of medical cannabis and its increased popularity among patients. Interestingly, the survey among prescribing physicians showed counter-intuitive results. While most stakeholders assumed that the physicians would be against the oversight of their prescriptions by the state, we found that a short majority actually supported the double gatekeeper system in place. The rationale was to protect them against patients who requested cannabis but to whom they did not want to prescribe the product due to insufficient medical indications. The questionnaire among referring physicians also documented their opinion and praxis around the prescription of cannabis (e.g., do they refuse patients' queries to receive cannabis and in which cases; what is their opinion on cannabis legalisation). This picture was important to draw, since the jurists were considering that prescribing physicians were all advocates of cannabis consumption, which was not supported by the survey. Altogether, the comprehensiveness of the research design helped obtain a nuanced understanding of the issue, from a multiple-party perspective. Finally, the evaluation team also had to preserve its scientific independence at the end of the process when the results were presented to the conflicting parties. One of the involved groups of public servants requested changes that

would have altered our interpretation of the results, which we refused. That led to the proposition that the team which did not agree with the final report write a counter-opinion to our study (which, in the end, did not happen).

The challenge of accessibility of sensitive data

For non-state agents, some of the most crucial data for conducting such studies are highly sensitive and difficult to access. Regarding the survey among prescribing physicians and the coding of the granted requests, the FOPH agents were concerned about anonymity and sceptical about transmitting the data to external policy evaluators. They also argued that these two modules would take too much time to implement. The public health agency had to unblock supplementary resources in a short timeframe – over Christmas – to prepare the launch of the online survey among referring physicians. The FOPH also had to work extensively on extracting and anonymising the application database to deliver the data in an appropriate form to the evaluation team. The extra-work generated within the agency intervened with their already limited resources, which did not please the involved civil servants. In addition, the research strategy aimed at objectivity such as the coding raised opposition from both teams. The jurists held that even if the evaluation team thoroughly coded all applications with regard to the medical conditions for which the authorisations had been granted over the years, the physicians could have not been fully transparent when filling the database. As to the agency's physicians, they feared the results of the coding, among other things because the application database had been filled incompletely over the years and lacked information, which could lead to political criticism. The danger of such criticism was high because the state has a special duty of documentation when authorising the medical use of prohibited substances. The evaluation team insisted on the crucial importance of the coding of the request as a mean to gather an objective picture of the evolution of authorisation-granting than what could be achieved through interviews and could help settling the dispute.

The challenge of epistemic legitimacy and political acceptance

A fourth challenge concerns the legitimacy of the research team itself in assessing a complex policy situation with high political and organisational stakes. For this policy evaluation, it was crucial to put together an interdisciplinary research team, especially since the hierarchical head of the FOPH's medical team was a social scientist, along with the majority of the evaluation team. The team of social scientists enjoyed legitimacy for the mandate as its members were specialised in public health policy evaluations, and because policy evaluations are strongly institutionalised within the FOPH. However,

having only social scientists aboard would have been perceived as biased. The subject of the study had a strong juridical component that required an external legal opinion in addition to the policy analysis. The evaluation team's legal expert was a university professor who specialised in pharmaceutical law and enjoyed legitimacy from the agency's jurists. In her final expertise, she prioritised patients' rights over the letter of the law, judging the latter to be no longer appropriate in light of the social developments that had taken place and therefore validated the authorisation praxis of the FOPH in spite of the growing numbers.

A key component of the whole study had been the access to the minutes of the parliamentary commission that had prepared the 2012 revision of the Federal Act on Narcotics. The minutes of the commission are normally strictly confidential, but we were given access to them for this study. Switzerland is a power-sharing semi-direct democracy that builds upon political compromise acceptable for the different political positions to prevent the launch of popular referendums against policy decisions (Vatter, 2008). Therefore, in the Swiss consociationalist system, the debates in parliamentary commissions are kept confidential to ensure consensus-building among political parties. Importantly, the minutes analysed revealed that the legislator thought that the system of exceptional authorisations would be provisory and limited to a few years. This was based on the belief that in the future, a new cannabis-based medicine would quickly enter the market. In this scenario, it was believed that the pharmaceutical companies would develop and industrially produce a fully tested cannabis-based pharmaceutical product that would replace the ad hoc *magistral formula*. This however did not happen, and the shortcomings of the initial formulation of the law began to show over the years. The provisory solution (system of exceptional authorisations, in which only a small number of authorisations may be issued) did not fit the societal trend moving towards a stronger acceptance and popularity of the use of cannabis medicine. This important result was discovered through the legal expertise and the parliamentary analysis, which helped reconstitute the legislators' will that the FOPH's jurists constantly referred to. This result helped pacify the situation within the office; instead of focusing on each team's shortcomings, it pointed out a problem of law obsolescence and the existence of a legislative assumption that had proven wrong. It showed the policy's wider picture, including political processes that happened outside the agency, and from a historical perspective, it explained the ambiguity of the system. By insisting on this unknown historical root of the legislation, the evaluation team offered an acceptable way out for each of the conflicting parties: sending the issue back to the legislator by pointing out the false assumption, and claiming for an update of the legislation at the light of this new information and the recent developments. Both legal security and patients' rights would be respected.

As a consequence, the issue de-escalated both at the administrative and on the political level. First, since the evaluation, the enforcement situation at the FOPH has shown positive development according to a study conducted by an evaluation specialist within the FOPH (Bonassi, 2020). Based on the evaluation, internal guidelines now specify how the physicians and legal unit work together and clarify the responsibilities of each team. Also, the exchange between the units has been institutionalised in regular meetings to improve cooperation, allowing for the resolution of misunderstandings in a structured manner. Second, one year after the evaluation, the policymakers at the national parliament agreed to modify the law. As a result, the prescription system was simplified to a single gatekeeper system[7] (cannabis remains a prohibited product but can be prescribed by referring physicians alone).

Discussion: scientific benefits and policy benefits

In this section we discuss our results, highlighting on the one hand the scientific benefits and on the other hand the policy benefits that can be created through policy evaluations. We also address existing reservations about applied policy studies and the conditions that need to be met for such benefits to occur.

As an applied streams of policy analysis, policy evaluation has evolved as a subdiscipline partly disconnected from more fundamental research streams. Policy evaluation could be viewed as a technocratic endeavour that drifted apart from more critical concerns of policy analysis. However, because of its distinctive position at the crossroads of fundamental research and real-world problems, policy evaluation has had to develop reflections, tools, and procedures to find a balance amid these tensions. However, applied research is far from being theory-free. As Alkin and Christie show in their evaluation theory tree (2004), evaluation is an encompassing academic undertaking ranging from branches dedicated to social accountability to purely epistemological research. Providing accounts of policy dynamics that are both theoretically robust and useful to solve problems is an ambitious endeavour. However, the proximity to the turmoil of political realities does not mean a disengagement from a reflective stance, quite the contrary. Because its activity is located at the heart of power relationships, applied policy research is in the position to provide not only informed and critical accounts of policies in the making, but also essential reflections on the position of policy analysts regarding their object of study.

Scientific benefits

Policy analysis scientific communities can show a certain reluctance towards applied policy studies, such as policy evaluation. A number of

issues are usually raised, including the independence of science and the implication in commissioned studies that topics, timing, and objectives are set by politics. These are important concerns. Government sciences, initially located in the state apparatus and subordinated to the political will, have historically had to separate themselves from politics and to conquer a scientific legitimacy (Ihl *et al.*, 2003; Sager *et al.*, 2018). The divorce was set by developing political and administrative sciences into proper academic spheres and adopting the rules of the scientific game. In a more recent past around the 1960s, policy analysis also distanced itself from public law and its doctrine, which was closely entangled with the action of the state (Payre & Pollet, 2005). This past proximity with the political power resulted in a strong distrust of applied studies, which is still felt today vis-à-vis policy evaluation (Delahais & Devaux-Spatarakis, 2018).

However, policy evaluation provides numerous advantages for policy analysis that shouldn't be dismissed. Applied policy evaluation studies give inestimable access to observation fields and data that would otherwise be difficult to access. It opens the door to highly topical issues, particularly in a context where involved parties are interested in participating in the research because they have a stance to defend. This paves the way for research teams to access confidential data, as well as valuable observations of the daily politico-administrative routine, such as participation in meetings and access to minutes. The interest of such procedures and data is reflected in the growing interest for ethnographic methods and direct observations in public management (Cappellaro, 2017) or policy studies (Dubois, 2015). It goes without saying that instrumentalisation risks are present, but so is the case in fundamental research when scientists negotiate their entries and presence in the field. As an established discipline, policy evaluation has techniques and procedures to protect its independence towards external pressures from politicians, stakeholders, and commissioners (e.g., Perrin, 2019). A stronger dialogue on these issues between policy evaluation and more theoretical streams of policy analysis could be beneficial. As instrumentalisation is a particularly obvious risk for policy evaluations, this discipline has been forced to pay close attention to questions of objectivation, distancing, risks of manipulation, and reflectivity. This know-how is useful for fundamental research as well, which also face similar challenges when doing qualitative research and field inquiries. In addition, policy evaluation has a long tradition of crafting its storytelling to communicate its results to society (in order to maximise the use of recommendations), which might also be of interest to the wider policy analysis community. In return, academic policy analysis has a lot to bring to evaluation research from a conceptual and theoretical perspective.

Policy benefits

Policy evaluations give policy analysts the opportunity (in best-case scenarios) to have some real-life impact with their studies. However, being able to prove policy evaluation's impartiality is an absolute prerequisite to generate policy benefits. Various techniques exist to preserve the independence of a study. First, just like in any fundamental research, crossing data, sources, and key informants – in other words, triangulation (Creswell, 2005) – is crucial in achieving independence. A report based on a sound triangulation of sources not only produces more reliable results, but also enjoys stronger legitimacy among potential users (Patton, 2001). This aspect is crucial given the particularly vivid framing contests in which controversial policies are embedded in (Bovens & 't Hart, 2016). Because policy evaluations consist of contracts between commissioners and research teams, they might be under higher scrutiny to prove their independence. Second, another key point is the importance of interdisciplinarity. The majority of the time, public policies bring together different fields that can be best understood with specialised knowledge (e.g., law, public health, natural sciences). Setting up interdisciplinary evaluation teams can be a requirement to seriously tackle complex transversal political problems by increasing the 'validity of knowledge generated' (Jacob, 2008, p. 182). This addresses one of the most recurrent criticisms against policy evaluators by specialised public agents who often believe that policy analysts do not have the required knowledge to understand and assess their praxis (Levin-Rozalis, 2010).

Third, because policy evaluators talk with many stakeholders about highly disputed issues, the question of how they weigh and synthesise gathered data to come to their final conclusions is particularly sensitive. Measurements must be made based on clear indicators (Wollmann, 2007) and scales, and assessments conducted based on declared criteria (Bellamy *et al.*, 2001) and transparently displayed synthesis processes (Scriven, 1994). Evaluators have the duty to map and reflect on the power and interest games at play among the stakeholders in the considered policy field (Bryson *et al.*, 2011). Establishing solid rules between policy-makers and policy analysts is also a prerequisite to strike a balance between proximity and independence in applied research situations (de Graaf & Hertogh, 2022). Finally, the formulated recommendations must fit the requirements of the praxis: political window of opportunity, organisational constraints, precision and realism, and stakeholder language (Patton, 1997). Once these basic conditions have been met, the stage is set for a possible policy benefit.

This article is based on a single case study, but the four identified challenges of applied policy analysis and the described contributions are generalisable. The case analysed here is highly typical of applied policy analysis: an

external policy evaluation, commissioned by authorities to face a political and/or an organisational crisis, timely situated in anticipation of political debates. It is representative of the contribution of policy evaluation to policy analysis on the one hand, and to the betterment of public policies on the other hand.

Conclusion: when policy analysis meets real-world issues

The final political decision will of course always be dependent on overweighting factors, such as the balance of political powers, external events, or the perceived state of public opinion. In other words, we should detach ourselves from 'romantic stories of 'evidence-based policymaking' in which we expect policymakers to produce 'rational' decisions' (Cairney, 2018, p. 200). In fact, if it is a challenge to 'induce policy (…) learning from highly public, politically charged forms of feedback such as that produced by evaluators', scholars can nevertheless have a bearing in trying to inject the results 'into the right places at the right time' (Bovens & 't Hart, 2016, p. 662). While policy evaluations inevitably come with the risks associated with the reopening of disputed legislative dossiers (Mastenbroek et al., 2016), they can be used as a chance for policy betterment. The role policy analysis and academic policy advice have to play in concrete policymaking is particularly timely and valuable in a post-truth era (Pattyn et al., 2022).

   The highlighted elements help open the door for a learning process in the policy subsystem. Policy analysis can offer a transversal enlargement of the perspective by reconstructing the whole system of interests, positions, and motives and rationalities around a policy. It can also provide a longitudinal enlargement of the perspective by retracing the historical developments of the policy, that might result in incoherent layers. By unfolding this picture, policy analysis can point out stumbling blocks and possible solutions. Because policy evaluation retraces whole policy pathways, sometimes with a very privileged access to data, it can be an important element to offset institutional amnesia – understood as the process of memory loss across the development of policies (Stark & Head, 2019). In the studied case, it did so on the regulatory side of amnesia (Hall et al., 2000; cited in Stark & Head, 2019), as the implementation praxis increasingly departed from the initial formulation of the law due to evolving social and medical trends. Hence, policy analysis can aim at reconciling the requirements of high-quality research by retracing complex policy trajectories, and of applied research by contributing to solve pressing political issues and 'to influence, rather than simply bemoan, the pathologies of the policy process' (Cairney, 2016). If successfully achieved, policy evaluations constitute an opportunity; they offer comprehensive analyses that provide a bigger picture and can potentially contribute to de-escalating polarisation.

Notes

1. The final policy evaluation report has been published (Mavrot et al., 2019).
2. A comprehensive analysis of the results of this enquiry are presented in Mavrot (2023).
3. https://www.seval.ch/app/uploads/2018/08/SEVAL-Standards-2016_e.pdf.
4. Interview statement of an FOPH-jurist. For the interview excerpts, translation from German and French.
5. Art. 8 Federal Act on Narcotics and Psychotropic Substances: https://www.fedlex.admin.ch/eli/cc/1952/241_241_245/en.
6. Interview statement of one of the responsible FOPH-jurists.
7. https://www.bag.admin.ch/bag/fr/home/medizin-und-forschung/heilmittel/med-anwend-cannabis/gesetzesaenderung-cannabisarzneimittel.html.

Acknowledgements

Disclosure statement

Funding

Notes on contributors

Céline Mavrot is an Assistant Professor for health policy at the Institute of Social Sciences, University of Lausanne, Switzerland. She specialises in comparative policy analysis, public health, and policy evaluation. She has conducted numerous policy evaluation mandates for public agencies and non-governmental organisations and teaches policy evaluation in various curricula. In 2020, she won the Award of the Swiss Evaluation Society together with Susanne Hadorn and Fritz Sager. In 2020–2021, she was a visiting postdoctoral scholar at the University of California, Los Angeles (UCLA) and at the Yale University School of Public Health. Her work has been published in journals such as the *American Review of Public Administration*, *Public Administration*, *Research Policy*, the *Journal of Comparative Policy Analysis*, *Climate Policy*, *Public Policy and Administration*, and the *European Journal of Political Research*.

Susanne Hadorn is a postdoc and policy evaluator at the KPM Center for Public Management, University of Bern. In her PhD thesis she analysed the management of

collaborative networks and her broader research interests include policy implementation, policy evaluation and health policy. She has published articles on topics such as street-level compliance, policy mixes, the use of policy advice during the Covid-19 pandemic and the independence of evaluations in journals such as *Public Policy and Administration*, *Evaluation* and *Research Policy*. Together Céline Mavrot and Fritz Sager, she won the Award of the Swiss Evaluation Society.

Fritz Sager is a Professor of Political Science at the KPM Center for Public Management at the University of Bern, Switzerland. His research focuses broadly on public policy and public administration. Specific fields of interest are the politics of expertise, policy implementation, evaluation and the history of administrative ideas. He has published in numerous international Political Science, Public Administration and Public Policy journals. Recent research regards executive politics, blame avoidance and evidence-based policy-making during the Covid-19 pandemic in Switzerland and abroad. His research has won several awards.

ORCID

Céline Mavrot 🅙 http://orcid.org/0000-0001-9603-5790
Susanne Hadorn 🅙 http://orcid.org/0000-0002-4470-7708
Fritz Sager 🅙 http://orcid.org/0000-0001-5099-6676

References

Alkin, M. (Ed.). (2004). *Evaluation roots. Tracing theorists' views and influences*. Sage.

Alkin, M., & Christie, C. A. (2004). An evaluation theory tree. In M. Alkin (Ed.), *Evaluation roots. Tracing theorists' views and influences* (pp. 12–66). Sage.

Archibald, M. M. (2016). Investigator triangulation: A collaborative strategy with potential for mixed methods research. *Journal of Mixed Methods Research*, *10*(3), 228–250. https://doi.org/10.1177/1558689815570092

Bellamy, J. A., Walker, D. H., McDonald, G. T., & Syme, G. J. (2001). A systems approach to the evaluation of natural resource management initiatives. *Journal of Environmental Management*, *63*(4), 407–423. https://doi.org/10.1006/jema.2001.0493

Bonassi, T. (2020). *Medizinalcannabis: Nutzung einer Evaluation im Bundesamt für Gesundheit in einem interdisziplinären und politischen Umfeld* [Unpublished manuscript]. Administrative source.

Bovens, M., & 't Hart, P. (2016). Revisiting the study of policy failures. *Journal of European Public Policy*, *23*(5), 653–666. https://doi.org/10.1080/13501763.2015.1127273

Bovens, M., 't Hart, P., & Kuipers, S. (2008). The politics of policy evaluation. In R. E. Goodin, M. Moran, & M. Rein (Eds.), *The Oxford handbook of public policy* (pp. 319–335). Oxford University Press.

Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, *34*(1), 1–12. https://doi.org/10.1016/j.evalprogplan.2010.07.001

Cairney, P. (2016). *The politics of evidence-based policy making*. Palgrave Macmillan.

Cairney, P. (2018). Three habits of successful policy entrepreneurs. *Policy & Politics*, *46*(2), 199–215. https://doi.org/10.1332/030557318X15230056771696

Cairney, P. (2023). The politics of policy analysis: Theoretical insights on real world problems. *Journal of European Public Policy* (introduction to the special issue, reference to be completed).

Cappellaro, G. (2017). Ethnography in public management research: A systematic review and future directions. *International Public Management Journal*, *20*(1), 14–48. https://doi.org/10.1080/10967494.2016.1143423

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Sage.

Creswell, J. W. (2005). *Education research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Pearson Education.

de Graaf, G., & Hertogh, M. (2022). Policy research under pressure: The case of the ministry of justice in the Netherlands. *Journal of European Public Policy*, *29*(3), 468–489. https://doi.org/10.1080/13501763.2020.1839535

Dekker, R., Geuijen, K., & Oliver, C. (2021). Tensions of evaluating innovation in a living lab: Moving beyond actionable knowledge production. *Evaluation*, *27*(3), 347–363. https://doi.org/10.1177/1356389021997848

Delahais, T., & Devaux-Spatarakis, A. (2018). Évaluation des politiques publiques et sociologie: état des lieux d'une relation distanciée. *Sociologies Pratiques*, *36*(1), 47–56. https://doi.org/10.3917/sopr.036.0047

Dorren, L., & Wolf, E. E. A. (2023). How evidence-based policymaking helps and hinders policy conflict. *Policy & Politics*, *11*, 1–22. https://doi.org/10.1332/030557321X16836 237135216

Dubois, V. (2015). Critical policy ethnography. In F. Fischer, D. Torgerson, A. Durnová, & M. Orsini (Eds.), *Handbook of critical policy studies* (pp. 462–480). Edward Elgar Publishing.

Fischer, F., Miller, G. J., & Sidney, M. S. (Eds.). (2007). *Handbook of public policy analysis. Theory, politics and methods*. CRC Press, Taylor & Francis.

Gerring, J. (2007). The case study: What it is and what it does. In C. Boix & S. C. Stokes (Eds.), *The Oxford handbook of comparative politics* (pp. 90–122). Oxford Academic.

Giel, S. (2013). *Theoriebasierte Evaluation: Konzepte und methodische Umsetzung*. Waxmann.

Hall, C., Scott, C., & Hood, C. (2000). *Telecommunications regulation: Culture, chaos and interdependence inside the regulatory process*. Routledge.

Hume, D. (1978). *A treatise of human nature III*. (L. A. Selby Bigge & P. H. Nidditch, Eds.). Clarendon Press.

Ihl, O., Kaluszynski, M., & Pollet, G. (Eds.). (2003). *Les sciences de gouvernement*. Economica.

Jacob, S. (2008). Cross-disciplinarization: A new talisman for evaluation? *American Journal of Evaluation*, *29*(2), 175–194. https://doi.org/10.1177/1098214008316655

Jasanoff, S. (2003). (No?) Accounting for expertise. *Science and Public Policy*, *30*(3), 157–162. https://doi.org/10.3152/147154303781780542

Klerman, J. A. (2010). Contracting for independent evaluation: Approaches to an inherent tension. *Evaluation Review*, *34*(4), 299–333. https://doi.org/10.1177/0193841X10370088

Ladner, A., & Sager, F. (2022). Switzerland: The politics of PA in a multi-party semi-direct consensus democracy. In A. Ladner & F. Sager (Eds.), *Handbook on the politics of public administration* (pp. 321–330). Edward Elgar Publishing.

Lagroye, J. (2003). *La politisation*. Belin.

Levin-Rozalis, M. (2010). Cybernetics: A possible solution for the 'knowledge gap' between 'external' and 'internal' in evaluation processes. *Evaluation and Program Planning*, *33*(4), 333–342. https://doi.org/10.1016/j.evalprogplan.2009.12.007

Linder, S. H., & Peters, B. G. (1987). A design perspective on policy implementation: The fallacies of misplaced prescription. *Review of Policy Research*, 6(3), 459–475. https://doi.org/10.1111/j.1541-1338.1987.tb00761.x

The LSE GV314 Group. (2013). Evaluation under contract: Government pressure and the production of policy research. *Public Administration*, 92(1), 224–239. https://doi.org/10.1111/padm.12055

Lyon, F., Gyateng, T., Pritchard, D., Vaze, P., Vickers, I., & Webb, N. (2015). Opening access to administrative data for evaluating public services: The case of the justice data lab. *Evaluation*, 21(2), 232–247. https://doi.org/10.1177/1356389015577507

Mastenbroek, E., van Voorst, S., & Meuwese, A. (2016). Closing the regulatory cycle? A meta evaluation of *ex-post* legislative evaluations by the European Commission. *Journal of European Public Policy*, 23(9), 1329–1348. https://doi.org/10.1080/13501763.2015.1076874

Mavrot, C. (2023). Bureaucratic entrepreneurship and morality politics: Dividing lines within the state. *European Journal of Political Research*, 62(2), 486–509. https://doi.org/10.1111/1475-6765.12559

Mavrot, C., Hadorn, S., Sprecher, F., & Sager, F. (2019). *Evaluation spezifischer Vollzugsaufgaben des BAG im Rahmen des Betäubungsmittelgesetzes (BetmG)*. Study Commissioned by the Swiss Federal Office of Public Health. Center for Public Management and Institute for Public Law, University of Bern. 92 p.

Morris, M. (1999). Research on evaluation ethics: What have we learned and why is it important? *New Directions for Evaluation*, 1999(82), 15–24. https://doi.org/10.1002/ev.1133

Morris, M. (2007). *Evaluation ethics for best practice: Cases and commentaries*. Guilford Press.

Morris, M., & Clark, B. (2013). You want me to do WHAT? Evaluators and the pressure to misrepresent findings. *American Journal of Evaluation*, 34(1), 57–70. https://doi.org/10.1177/1098214012457237

Nielsen, S. B., & Ejler, N. (2008). Improving performance? Exploring the complementarities between evaluation and performance management. *Evaluation*, 14(2), 171–192. https://doi.org/10.1177/1356389007087538

Parkhurst, J. (2017). *The politics of evidence: From evidence-based policy to the good governance of evidence*. Taylor & Francis.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Sage.

Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22(3), 329–336. https://doi.org/10.1177/109821400102200307

Pattyn, V., Blum, S., Fobé, E., Pekar-Milicevic, M., & Brans, M. (2022). Academic policy advice in consensus-seeking countries: The cases of Belgium and Germany. *International Review of Administrative Sciences*, 88(1), 26–42. https://doi.org/10.1177/0020852319878780

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage Publications.

Payre, R., & Pollet, G. (2005). Analyse des politiques publiques et sciences historiques: quel(s) tournant(s) socio-historique(s)? *Revue Française de Science Politique*, 55(1), 133–154. https://doi.org/10.3917/rfsp.551.0133

Perrin, B. (2019). How to manage pressure to change reports: Should evaluators be above criticism? *American Journal of Evaluation*, 40(3), 354–375. https://doi.org/10.1177/1098214018792622

Pleger, L., Sager, F., Morris, M., Meyer, W., & Stockmann, R. (2017). Are some countries more prone to pressure evaluators than others? Comparing findings from the

United States, United Kingdom, Germany, and Switzerland. *American Journal of Evaluation*, *38*(3), 315–328. https://doi.org/10.1177/1098214016662907

Pollitt, C. (2000). Institutional amnesia: A paradox of the 'information age'. *Prometheus*, *19*(1), 5–16. https://doi.org/10.1080/08109020050000627

Radaelli, C. M. (1995). The role of knowledge in the policy process. *Journal of European Public Policy*, *2*(2), 159–183. https://doi.org/10.1080/13501769508406981

Rennes, J. (2016). Les controverses politiques et leurs frontières. *Études de Communication*, *2*(47), 21–48. https://doi.org/10.4000/edc.6614

Riccucci, N. M. (2010). *Public administration: Traditions of inquiry and philosophies of knowledge*. Georgetown University Press.

Ritz, A. (2003). *Evaluation von New Public Management: Grundlagen und empirische Ergebnisse einer Reform-Evaluation aus der schweizerischen Bundesverwaltung*. Haupt.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (1999). *Evaluation: A systematic approach*. Sage.

Sager, F., Hadorn, S., Balthasar, A., & Mavrot, C. (2021). *Politikevaluation. Eine einführung*. Springer.

Sager, F., & Mavrot, C. (2021). Participatory vs expert evaluation styles. In M. Howlett & J. Tosun (Eds.), *The Routledge handbook of policy styles* (pp. 395–407). Routledge.

Sager, F., Rosser, C., Mavrot, C., & Hurni, P. (2018). *A transatlantic history of public administration: Analyzing the USA, Germany and France*. Edward Elgar.

Schmidli, F. H., Pleger, L. E., & Hadorn, S. (2023). Don't you forget about me: Independence of evaluations from the perspective of US evaluation clients – an exploratory study. *Evaluation*, *29*(1), 110–132. https://doi.org/10.1177/13563890221138876

Scriven, M. S. (1991). *Evaluation thesaurus* (4th ed.). Sage Publications.

Scriven, M. (1994). The final synthesis. *Evaluation Practice*, *15*(3), 367–382. https://doi.org/10.1177/109821409401500317

Smith, K. (2010). Research, policy and funding – academic treadmills and the squeeze on intellectual spaces. *The British Journal of Sociology*, *61*(1), 176–195. https://doi.org/10.1111/j.1468-4446.2009.01307.x

Stark, A., & Head, B. (2019). Institutional amnesia and public policy. *Journal of European Public Policy*, *26*(10), 1521–1539. https://doi.org/10.1080/13501763.2018.1535612

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. Jossey-Bass.

Topp, L., Mair, D., Smillie, L., & Cairney, P. (2020). Skills for co-creation. In V. Šucha & M. Sienkiewicz (Eds.), *Science for policy handbook* (pp. 32–42). Elsevier.

Trivellato, U. (2019). Microdata for social sciences and policy evaluation as a public good. In N. Crato & P. Paruolo (Eds.), *Data-driven policy impact evaluation* (pp. 27–45). Springer.

Vatter, A. (2008). Swiss consensus democracy in transition: A re-analysis of Lijphart's concept of democracy for Switzerland from 1997 to 2007. *World Political Science*, *4*(2), Article 1. https://doi.org/10.2202/1935-6226.1042

Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis*, *3*(4), 531–545. http://www.jstor.org/stable/42783234.

Wollmann, H. (2007). Policy evaluation and evaluation research. In F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), *Handbook of public policy analysis. Theory, politics and methods* (pp. 393–402). Taylor & Francis.