

Addressing Multilevel Program Complexity by Evaluation Design: The Comparative Evaluation of Cantonal Tobacco Prevention Programs in Switzerland

Fritz Sager, Susanne Hadorn, Céline Mavrot

Centre of Competence for Public Management, University of Bern

Paper prepared for ICPP 2015, Milano, T08P13 – What’s new in policy evaluation? Exploring theoretical and methodological approaches to assess complex programs

Preliminary draft - work in progress. Comments are welcome!

Abstract

In this paper, we present the evaluation design for a complex multilevel program recently introduced in Switzerland. The evaluation embraces the federal level, the cantonal program level and the project level where target groups are directly addressed. We employ Pawson and Tilley’s realist evaluation approach in order to do justice to the varying context factors that impact the cantonal programs leading to varying effectiveness of the implemented activities. The application of the model to the canton of Uri shows that the numerous vertical and horizontal relations play a crucial role for the program’s effectiveness. As a general learning for the evaluation of complex programs, we state that there is a need to consider all affected levels of a program and that no monocausal effects can be singled out in programs where multiple intervention address the same problem even if target groups may differ.

Introduction

The present paper presents the approach of the comparative evaluation of the cantonal tobacco prevention programs in Switzerland and links this Swiss case study with current trends in evaluation practices. The complex frame of the Swiss tobacco prevention program offers an interesting case for various dimensions of policy evaluation. Within a set of funding criteria defined at the federal level, sub-national administrations are incentivized to frame their own, customised prevention programs. The main expectation of the policy is that coordinated programs lead to synergies and hence better outcomes than uncoordinated projects. As a result of this allocation of responsibilities, each cantonal program consists of an individually composed set of projects with a distinct emphasis on structural prevention, behaviour oriented prevention and information activities. An important question of the evaluation is whether the cantons are the appropriate level of action to adopt effective policies.

The evaluation concept of the comparative evaluation of cantonal programs is presented in this paper. The aim is to contribute to the advancement of complex program evaluation designs. We claim to achieve this, as the comparative evaluation has to focus on different levels in the case at hand. Firstly, the individual projects have to be assessed, secondly, each cantonal program will be subject to evaluation. Lastly, the eleven program evaluations are used as a basis for an inter-cantonal comparison, where the composition of the programs as well as the impact of the specific cantonal context are studied more closely. The comparative analysis will thus strive at presenting explanations for cantonal differences in efficacy and efficiency, both based on the policy concept as well as the implementation phase.

The paper starts with a short introduction to decentralized policy implementation in Swiss federalism in which the cantonal tobacco prevention programs are situated. We then present our evaluation design with which these cantonal programs are assessed and illustrate its realization with the findings from the evaluation of the canton of Uri. Ultimately, the paper endeavors to provide some conclusions about the effects of policy co-formulation as well as the challenges of multi-level governance with regard to the selected case study.

Swiss federalism and implementation by federal delegation

In Switzerland, federalism is the most distinctive feature of the political system after direct democracy (Sager/Zollinger 2011). Only a few West European countries have systems in which not only the federal government but also the constituent state enjoys legislative, executive and judicative power (Ismayr 2009). In addition to Switzerland, traditionally this can be observed in Germany and Austria and more recently in Belgium and Spain.

The implementation of federal policies in Switzerland is strongly shaped by this institutional setting. The federal programs are implemented by the cantons and communes. 'While the Federation holds the legislative power in many areas, responsibility for implementing federal policies resides to a large extent with the cantons.' (Vatter 2007: 91). Therefore, in a great number of policy domains, the federal level is dependent on the cantons for the implementation of federal legislation. Due to the high degree of legislative autonomy of the Swiss cantons, the delegation applies not only to the actual implementation of federal laws (i.e. the right to act) but also to the adaption of these provisions to the local situation (i.e. the right to decide). The cantons are not only implementing, but also programming authorities.

According to Linder (1987; 2010) the implementation of federal policies by the cantons is advantageous both for the Federation and for the cantons. Whereas for the central government the

advantage lies in reducing its workload, for the cantons, the advantage lies in being able to control their own program priorities and in adapting policy implementation to the local context. Despite this optimistic view, 'implementation by federal delegation' draws mixed reviews (Kissling-Näf/Wälti 2007: 504). While some lament the lack of federal control over implementation, others value the receptivity and adaptability as well as the flexibility and experimental character of decentralized implementation. A further drawback is that not every canton can afford to provide for well-funded and professionalized implementation units (Sager 2003). An additional problem that derives from the 'implementation by federal delegation' is uneven results between the cantons (Sager/Rüefli 2005).

Both advantages and drawbacks also come into play when policy programs launched at the federal level are meant to lead to respective coordinated policy action at the cantonal level. For the case of palliative care, Moser and Sager (2015) show cantonal path dependencies to play a crucial role for whether cantons respond to vertical stimulus or not. In the case of alcohol control policy, Sager and Rielle (2013) identify administrative structures as a core factor for the adoption of cantonal policy programs. Accordingly, Mavrot and Sager (2014) argue so called vertical epistemic communities to play a major role for the establishment of cantonal tobacco prevention programs. The case of these programs is presented in the following section.

The Tobacco prevention fund and the cantonal tobacco prevention programs

Tobacco prevention policy in Switzerland took only off after 2001 with the National Tobacco Prevention Program 2001-2005. However, until the early 2000, due to the central state weakness regarding policy implementation, coordination at the cantonal level mainly derived from voluntary inter-cantonal efforts. Nevertheless, considerable change has taken place since the creation of the Fund for tobacco prevention in 2004 as a leading national actor. Loose and partial inter-cantonal coordination is gradually being replaced by a more vertical coordinating approach including all aspects of smoking prevention.

The Fund was created after the national parliament decided to revise the tobacco taxation legislation in 2003. Taxes increased, whereby an important source of income dedicated to smoking prevention was obtained. The Fund was created to manage this new income by funding any relevant smoking prevention project both those from non-governmental organizations (NGOs) and those from cantonal administrations. This aspect of its mission gives the Fund a significant weight in influencing the policymaking. Its decision-making power in financial allocation allows the Fund to act in a rather voluntaristic way. The strategic orientation of the Fund is provided by the National Tobacco Program, which is designed by a broad panel of actors under the leadership of the Federal Office for Public Health

(FOPH) and approved by the Federal Government every couple of years. It includes the following lines of action: preventing the beginning and promoting the cessation of consumption, protecting against second-hand smoke exposure, raising public awareness, encouraging research activities, developing framework conditions to foster preventive actions, and finally, building a national network of tobacco prevention actors (RS 641.316, art.2). The Fund managed to gain great influence in the relevant policymaking processes despite its proportionally weak resources in terms of personnel. Since its creation, its collaborator numbers have always been very low for an organization deciding the allocation of millions of francs. Hence, the two major strengths of the Fund are: a) its significant budget and its privilege to choose which prevention projects will be funded, and b) the large network of experts it has rapidly built, which largely compensates for its internal weaknesses. The Fund's high financial capacity and authority to make allocation decisions provide it with considerable power of persuasion and of steering.

The resulting new scheme for cantonal tobacco programs is that cantons can apply for considerable funds from the Fund when proposing an overall evidence-based set of prevention projects that will be coordinated by a cantonal program office and implemented by various actors from the tobacco prevention policy network. Both the program coordination and the projects can be either managed by public or private entities. The Fund expects the program format to result in more coordinated action and synergy effects that will lead to better effectiveness in terms of outcomes and impact. At the same time, this scheme leads to very complex horizontal and vertical relations that impact program effectiveness. First, the Fund at the Federal level horizontally needs to coordinate with the FOPH that demands that cantonal programs are in line with the national tobacco prevention policy. Second, the Fund is aiming at coordinating the various cantonal efforts by the selective allocation of funds. However, not only cantons can apply for project funding but also NGOs that act at the national scale. In the case where such NGO projects are implemented in a canton applying for a program, these national projects also need to be coordinated even though they were originally not designed to be part of a cantonal program.

Third, cantons question the Fund's legitimacy to steer their programs as the tax money does not belong to the Federal level but is simply bound to be spent on tobacco prevention. Consequently, the Association of Cantonal Preventions Experts¹ acts as sounding board for the cantonal programs and interferes with the Fund's policy when deemed necessary. Fourth, the cantonal coordinators running the programs are in a difficult double role in which they have to manage the variety of projects in their canton and at the same time are accountable to the Fund. This double role is even more difficult as most of the cantonal coordinators are not managers but prevention specialists, i.e. neither trained in

¹ Vereinigung der kantonalen Beauftragten für Gesundheitsförderung in der Schweiz – VBGF.

managing large projects nor in monitoring and collecting evaluation data they are bound to provide to the Fund in order to get the funds. Fifth, the Swiss cantons are fully fledged political systems that vary not only in social, economic, and cultural aspects but also in terms of politics and administrative organization. Context hence plays a crucial role for program effectiveness.

In the following, we present our overall evaluation design to assess the cantonal programs under these complex conditions.

Capturing vertical complexity: a realistic multilevel evaluation design

The complex frame of the Swiss tobacco prevention program offers an interesting case for various dimensions of policy evaluation. It is, for example, necessary to examine if the sub-national administrations, who are now entitled to frame their own, customized program, are the appropriate level of action to adopt effective policies. As a result of this allocation of responsibilities, each cantonal program consists of an individually composed set of projects with a distinct emphasis on structural prevention, behavior oriented prevention and information. Given the depicted framework, the qualitative evaluation has to focus on different levels. First, the individual projects have to be assessed, second, each cantonal program will be subject to evaluation. Last, the eleven program evaluations are used as a basis for an inter-cantonal comparison, where the composition of the programs as well as the impact of the specific cantonal context are studied more closely. The comparative analysis will thus strive at presenting explanations for cantonal differences in efficacy and efficiency, both based on the policy concept as well as the implementation phase. Consequently, the overall evaluation of the cantonal programs is not mandated in form of an evaluation contract but financed by the Fund as a basic research project.

The main questions with regard to the single cantonal programs are

- Which if any progress can be stated in tobacco prevention?
- Has the program met its goals?
- Which are the drivers and which are the obstacles to progress in tobacco prevention?
- Which, if any, is the value added of the coordinated program as opposed to uncoordinated projects?
- Overall, was the cantonal program a success?

In the following, we first present our general theoretical approach to this evaluation, namely realistic evaluation, before we turn to our analytical model, the different evaluation elements and the respective hypotheses.

General approach: realistic evaluation

Realistic evaluation (Pawson/Tilley 1997) emphasizes the importance of context in analyzing public policies. More specifically, realistic evaluations break political programs down into the three components of program context (C), mechanism (M) and outcome (O). According to realistic evaluation, political programs can trigger a range of mechanisms of change. The specific mechanisms triggered depend on the program context. This notion of complex causation means that the relationship between causal mechanisms and their effects is not fixed, but contingent. Context matters because it "turns (or fails to turn) causal potential into causal outcome" (Pawson/Tilley 1997: 69). As a result, a political program can lead to a variety of outcomes in different contexts. This gives rise to different context-mechanism-outcome configurations, briefly referred to as CMO configurations. Realistic evaluation is about conceiving and testing these CMO configurations in order to provide insight on what works for whom and under what circumstances (Sager 2008).

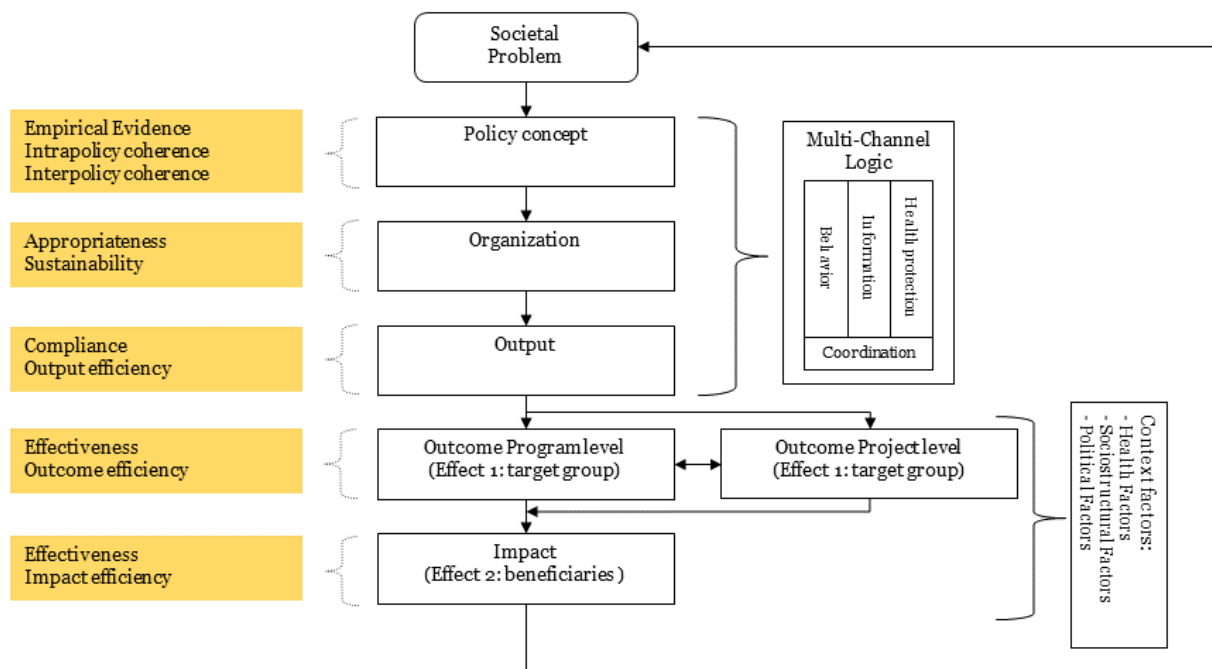
Pawson and Tilley (1997: 159) emphasize the importance of building on theory in determining the focus of an evaluation. Moreover, the recourse to theory and to the results of previous empirical studies in a certain field facilitates the accumulation of knowledge. In his next step, Pawson (2002a; 2002b) transfers this logic of theory-driven and context-sensitive evaluation to the level of research synthesis. He claims that the accumulation of knowledge could be enhanced through what he calls 'realist synthesis', i.e. a summary of existing studies that follows the CMO pattern, enquiry as to the mechanisms they detected, the contexts they found to be relevant and the outcomes they observed. The result would be a collection of potentially similar CMO configurations which could be fused into more abstract ones.

Pawson's understanding of context is closely linked to the notion of contingency. Contingency implies that the meaning of social action depends on the conditions under which it takes place. This means that context refers to both situations and addressees of policy interventions as it is the addressees' perception, under given circumstances, which defines the meaning of social action and thus determines the outcome of a mechanism in a CMO configuration. This notion of context strongly refers to the particular time of implementation. In the following we present the different elements of our evaluation model and formulate the respective hypotheses.

Evaluation model

Our general model for cantonal programs is presented in Figure 1. The focus of this evaluation is on the program level and we therefore refrain from hypotheses at the project level. However, the actual impact of the project only takes place at this latter level where the final target groups are addressed. We therefore use project data to measure outcome and impact.

Figure 1: Evaluation model



Adapted from Sager and Hinterleitner (2014: 444)

In what follows we define the different stages, name the employed evaluation criteria and formulate causal hypotheses.

Policy concept

The policy concept embraces the causal model of a program and the planned implementation structure as put to paper in the cantonal proposals. The concept hence is the program theory before it is put into practice. The criteria we use to evaluate the concept are threefold: empirical evidence, internal coherence and external coherence (Knoepfel et al. 1997; Ledermann/Sager 2009). Empirical evidence refers to the backing of the assumed effects on scientific studies in the fields of public health research

and policy analysis, or on policy experience previously gained by the local actors. Internal coherence means the fit of the different elements of a program, i.e. the avoidance of internal contradictions in the program logic and potential implementation deficits. External coherence refers to the policy context of a program and asks whether the program is in line with other relevant programs in the canton that might interfere with the program logic. This regards both other prevention programs and programs from other policy areas such as economic or education policy. An important issue in this respect is the so called ‚Multi-Channel‘-logic stating that behavior change in target groups is the result of a combination of multiple channels such as regulation, incentives and information that may be limited to tobacco prevention or may go beyond substances and focuses on addictive behavior and its conditions as such (Lund/Aarø 2004; Bala et al. 2013). A program is expected to combine different fields of action operating in a variety of settings. An appropriate policy concept also is expected to account for the distinction between policy failure and implementation failure in that organizational aspects should be planned in accordance to the expected causal effects of the foreseen interventions (Ledermann/Sager 2009). The hypotheses for the policy concept focus on the effectiveness of the program:

- The better the empirical evidence of a program, the better its effectiveness with regard to a) outcome and b) impact.
- The stronger the ‚Multi-Channel‘-logic of a program, the better its effectiveness with regard to a) outcome and b) impact.
- The more precise the objectives of a program are defined, the better its effectiveness with regard to a) outcome and b) impact.
- The higher the internal coherence of a program, the better its effectiveness with regard to a) outcome and b) impact.
- The higher the external coherence of a program, the better its effectiveness with regard to a) outcome and b) impact.

Organization

Organization refers to a cantonal program’s implementation structure as in fact established during implementation realized. The realized implementation structure of a program does not necessarily correspond to the planned implementation structure. We understand organization as an institutional structure. Institutions distribute decision power by rules and routines and coin actor identities and their interpretations of situations. Organization entails the allocation of decision and action competencies, financial and personnel means, and expertise (Sager 2004; Sager/Rielle 2013).

Evaluation criteria are appropriateness and sustainability of an implementation structure. Appropriateness refers to the question of how far the organization is in line with the actual tasks for service provision. Particularly important in this respect is coordination capacity in order to do justice to the ‚Multi-Channel‘-logic of the programs. Sustainability regards long-term security of the programs and projects in order to avoid stop-and-go-problems damaging effectiveness. The hypotheses read as follows:

- The more appropriate the implementation structure of a program, the better its outcomes.
- The more sustainable the implementation structure of a program, the better its outcomes.

We now turn to the output side of the program. Following Patton (1997: 193-194), the success of public policy, thus understood, can be measured on the three levels: outputs, outcome, and impact.

Outputs

Outputs represent the products of a policy. These are public interventions or services which attempt to change actors' behavior. To achieve this, outputs create a direct relationship between the competent implementation actor and the political target group, where the political target group refers to the group of actors whose behavior is seen by public politics as relevant to the problem in question. If the planned outputs are not delivered in the intended manner in terms of time or scope, the policy cannot be expected to achieve its results due to implementation failure. In the case of the cantonal programs, outputs basically regard management services towards the project implementers, i.e. governance and performance control, contract management, and internal and external information and coordination.

Evaluation criteria for output performance are compliance and efficiency. However, as the latter does not have a direct relation to effectiveness but only considers the cost-service-ratio, we do not use the efficiency criterion in this evaluation. Compliance refers to the degree to which planned services actually are delivered. This results in the following hypothesis:

- The more the program implementation complies with the program concept, the better its outcome.

Outcome and Impact

Outcome designates the changes in the behavior of the political target group brought about by the outputs. The addressees of a given policy are not necessary equal to the beneficiaries of the policy.

Rather, they are the ones the policy sees as responsible for causing the the problem at stake. It is thus the addressees' behaviour that needs to be changed in order for the groups suffering from a given problem to benefit. E.g., in order to protect small children in traffic, the mobilists caution has to be increased. The target group's reactions to output may be as expected, take on a non-intended form, or indeed not take place at all.

Finally, impact implies the totality of intended or non-intended effects which the achieved changes in the policy's target group's behavior have on the problem which the relevant policy is aiming to solve. If the addressees change their behaviour in the intended way without an effect on the problem situation, the causal hypotheses of the policy is likely to be wrong (policy failure). Where addressees' behavioural change has a positive effect on the problem and an effect which corresponds to the political objective, one refers to an effective policy.

The criteria for both is effectiveness, i.e. the degree to which expected effects in term of outcome or impact are in fact achieved by the delivered outputs. Again, efficiency shall not be considered. The core hypothesis refers to the causal link between the two effect stages:

- The higher a program's outcome effectiveness, the higher is its impact effectiveness.

Context

Policy programs always take place in a specific context (Pawson/Tilley 1997; Pawson 2002a; 2002b). This context is defined by factors that exist exogenously and cannot be altered (or only with great difficulty), and which, to a certain degree, define the structure within which the program planning and implementation process takes place. The purpose of the inclusion of these contextual factors is so that, at the end, the success factors can be differentiated according to the framework of the project (Pawson/Tilley 1997; Befani et al. 2007; Sager/Andereggen 2012). This means that contextual factors impact the way target groups react to an intervention. Context hence is responsible to the fact that the same intervention once works and another time does not. We therefore integrate context into our model and formulate three distinct general hypotheses regarding problem prevalence, socio-economic context and political context.

- Comparatively large tobacco-related and general health problems in a canton increase the program's impact effectiveness.
- An unfavorable socio-structural context in a canton decreases the program's impact effectiveness.
- A favorable political context in a canton increases the program's impact effectiveness.

Applying the model: Findings for the Canton of Uri

Within the scope of our mandate, Uri was the first canton subject to evaluation. The previously discussed evaluation model was applied to the case and the five elements policy concept, organization, output and lastly outcome and impact were analyzed. Below, the main findings of the evaluation will be discussed with reference to the defined hypotheses.

Method

The evaluation of the cantonal tobacco prevention program in Uri was based on the analysis of primary and secondary data, whereby the data sources differed between the five evaluation elements. More specifically, the evaluation element context was analyzed based on data provided by the Swiss statistical office, national and cantonal health surveys, newspaper articles, economic reports, parliamentary debates as well as diverse cantonal documents. For instance, as a specific part of the context analysis, we conducted a media analysis including all articles published in the period between 01.03.2009-31.12.2014 in the “Neue Urner Zeitung”, using search terms relevant to tobacco prevention (such as “Tabak”, “Passivrauch”, etc.). The thereby collected publications were assessed by means of a qualitative content analysis. The analysis of the second evaluation element, the policy concept, was mainly based on a qualitative content analysis of the canton’s program paper approved by the Fund. Lastly, the elements output, outcome and impact were assessed through a time series of self-evaluation data provided by the cantonal implementing partners on a yearly basis, as well as through additionally conducted semi-structured interviews with key players.

Policy concept

The case study of Uri fully confirmed the crucial importance of the policy concept underlying a program. This is the case regarding all four dimensions of the policy concept: the empirical evidence, the precision of objectives, as well as the internal and external coherence of the program.

The two main learnings regarding the empirical evidence are the following: first, the canton based the concept more on its own past experience than on other available scientific evidence. The use of its own experience proved to be a positive factor. It allowed to build on existing activities, to avoid mistakes made in the past and to better identify the prevention gaps in the canton. Making an initial assessment of the cantonal situation was of particular importance in this context where cantons are

seeking for a better coordination of activities. It proved to be a good opportunity for promoting concerted efforts. However, this process could have been favorably complemented by a better use of scientific evidence, which was widely neglected. The canton had difficulties to link its concrete program with theoretical literature, and to base the program on scientific evidence. Second, it turned out that when evidence was used –whether scientific or stemming from experience– this was more at a macro program level than for specific projects. As a consequence, this lack of consideration of evidence at the project level led to policy failures reducing the effectiveness of the program.

The precise definition of the prevention objectives appeared to be a crucial part of a sound policy concept. Interestingly, the canton showed more difficulties to formulate precise objectives at the macro (i.e. program) than the micro (i.e. project) level. A specific process happened in Uri, which led to a highly imprecise definition of the program’s objectives, and hampered the effectiveness of the program. The tobacco prevention programs are being financed on a matching co-funding basis (canton-Fund). In order to increase the national contribution, Uri included as cantonal self-contribution many activities that bear only little relationship to tobacco prevention. As an effect, this blurred the whole readability of the program and complicated its implementation. Whilst the defined output goals – that have no or insufficient reference to smoking prevention – could be achieved, they did not translate into a change in behavior of the target groups.

The insufficient internal coherence of the program of Uri reveals a lack of understanding of the whole smoking prevention cantonal program logic. Even though these cantonal programs are a new policy instrument, and the cantons need time to adapt to these new processes, our case study highlighted important inconsistencies between output and outcome objectives within projects. The lack of causality between these two categories of objectives hampered the success of many projects. Second, incoherencies regarding the links between the outcome objectives (i.e. projects) and the impact objectives (i.e. program) were identified. This was especially the case in regard to the target group smokers. These policy failures constitute a basic incoherence which directly impacted the efficiency of the program. In contrast, the outcome-impact chain concerning youth was compact and coherent, thus providing the basis for a successful implementation. However, our case study also showed that a dense range of measures aimed at one target group runs the risk of duplications.

Regarding the external program coherence, our case study highlighted two main negative factors. First, the neglect of the medical setting constituted an important lack in matter of smoking prevention. The cantonal program could have benefited from an inclusion of this setting allowing to reach many beneficiaries in a facilitating context. Second, the saturation of schools was highlighted: they are approached by numerous external demands, and have to make a selection between competing themes (violence, social network, etc.). It is therefore crucial to achieve a good external coordination with the

school services in order to obtain access to the setting. We also observed two important positive factors regarding external program coherence: the building of synergies with other addiction services such as for example drug counselling services (coordination of objectives), and the building of synergies with other health promotion programs, for example in order to activate already existing partner-networks (procedural coordination).

Our case study on Uri supports the hypotheses on the policy concept. Evidence, precision, as well as coherence are crucial factors of a program's conception. Interestingly, the canton of Uri had more difficulties with the project than the program level as regards the use of evidence and the precision of objectives. Regarding coherence, the links between outputs, outcomes and impacts were too weakly conceptualized.

Organization

Looking at the canton's implementation structure, the case of Uri confirmed the importance of organizational fit. The main findings can be divided into two groups. Firstly, in accordance with the predefined hypotheses, the structure on the cantonal level proved to be a determining factor with regard to the effectiveness of the tobacco prevention activities. Secondly, the evaluation results showed that there is a need for an extension of the evaluation model which stresses the importance of the collaboration between cantonal and national actors. Although this level of collaboration was not considered in the initial evaluation model, which only considered the cantonal level, its importance became increasingly evident, which is why the analysis was enriched accordingly.

The first group of findings regarding the internal structure of the program concerns the cantonal level, where both factors beneficial and detrimental for the program's effectiveness were identified. Firstly, as beneficial factor, the launch of the cantonal program demonstrated the political interest for tobacco prevention, leading to increased legitimation of the related activities. Secondly, the newly created central program manager enabled the actors to improve their networks and thus to identify and use synergies. Thirdly, the collaboration within the canton was additionally facilitated by the relatively small size of the cantonal administration. Due to the pre-existing closeness of the various administration units, the exchange between the different departments was informal and allowed for simple access to different settings. Fourthly, the program manager's capacity to include central actors such as the police and political municipalities had immediate effects on the program's effectiveness. In cases of strong collaboration, the activities' performance was increased due to better access to settings and increased legitimacy. Where collaboration with such central partners remained weak, the activities could often not be implemented as planned and lacked public acceptability. Lastly, the

relatively inflexible and restrained budget of the canton in conjunction with the partially unpredictable financing of the Fund led to uncertainties in some projects. Where national funding was ceased unexpectedly, the canton failed to be sufficiently flexible and could not raise the necessary funds to continue the concerned activities, leading to losses regarding effectiveness.

The last point leads us to the second group of findings in regard to the program's structure, namely the collaboration of cantonal and national actors, where additional factors relevant to effectiveness were identified. Along with the above mentioned restrictions concerning the partially uncertain national funding of the prevention activities, other changes in regulations as e.g. evaluation procedures led to disruptions of the implementation. Besides the immediate practical implications such as the underfunding of certain activities, the perceived lack of transparency caused growing discontent on the cantonal level. In the case of Uri, these problems of cooperation already started while the program was designed and continued in the subsequent approval stage, processes which were lengthy and unpredictable in the view of the canton. The relatively long preparation phase and the uncertainty concerning the Fund's approval led to decreasing political support and growing demotivation of the implementing partners. This rather unfortunate start of the collaboration had implications on the implementation phase and the dissatisfaction of the main cantonal actors could not be reduced until the program's completion.

In conclusion, the case of Uri demonstrates that whilst studying the program structure, it is not sufficient to exclusively consider the cantonal level. The importance of the organizational structure cannot be assessed without considering the collaboration between the cantonal and the national level. Even if the cantonal structure supports the successful implementation of the program, the overall success of the prevention activities depends on a functioning collaboration between the Fund and the cantonal authorities. In regard to the first hypothesis, it can be concluded that the appropriateness of the program's structure is decisive for its success. As postulated in the second hypothesis, the sustainability of the program structure, especially in regard to financial security, proved to be an additional important factor contributing to the program's overall performance.

Output

The evaluation showed that a vast majority of output objectives were achieved by the implementing partners. However, a closer look at the implemented activities exposed substantial weaknesses in regard to the services' effects on the target groups. The comparison of the level of achievement of both output and outcome goals indicates that there was in some cases a lack of causality between these two levels of analysis. As will be discussed below, a much smaller proportion of outcome targets

than output goals were achieved. One of the underlying reasons were insufficiently ambitious output goals, which often did not go beyond the development of organizational structures which were necessary to provide the actual services. Hence, output goals were achieved without ensuring that the target group was reached. Looking at the causes for non-achievement of output goals, the analysis identified certain management problems. In a number of cases, insufficient commitment and flexibility of the implementing partners caused an absence or delays of responses to arising issues. For instance, the importance of multipliers was recognized in the program's conception, while their advantages were not exploited in the implementation due to inadequate inclusion of these key actors.

In summary, the assessment of the output level requires in a first step a qualitative assessment of the target definitions in order to provide well-informed statements about the level's effectiveness and efficiency. Moreover, in cases of well-defined goals, managerial capacity is decisive for the success or failure of service provision. In general, the appropriateness of the outputs increases the activities' performance in regard to the outcome dimension.

Outcome and Impact

As discussed above, a much smaller proportion of outcome goals were achieved in comparison with the outputs. Some of the underlying reasons for the moderate success in the outcome dimension (such as the incoherency between the two levels as well as the partially problematic definition of output goals) were outlined above, whereas others will be discussed in the following. To that end, outcome achievement will be analyzed by looking first at the composition of the projects and second at the context in terms of tobacco coalitions.

Firstly, the case of Uri provided strong evidence for the importance of the previously discussed 'Multi-Channel'-logic. Target groups which were approached with a well-balanced set of projects in different settings were more likely to be addressed in an effective manner than groups with a weakly conceptualized approach. More specifically, the first main target group, youth, benefitted from an advantageously planned set of projects in the settings school (both compulsory education and post-obligatory education), leisure and sports, as well as market. The subsequent successful implementation of these activities did not only target the adolescence in different surroundings, but also with a balanced selection of behavioral and conditional prevention as well as information activities. These projects were designed for different ages within the superordinate target group. By contrast, the second main target group smokers could not be addressed effectively. Already in the program design, the chosen approach included only a very restricted number of projects for smokers and only a small proportion of the available funds were allocated to this group. Besides this not very

advantageous basis, the respective projects turned out to be ineffective in the implementation phase and reached only a very small proportion of its audience. Taken together, smokers in Uri did not sufficiently benefit from the smoking cessation aids offered by the program's activities due to poor planning and unsuccessful implementation of individual projects. Overall, the 'Multi-Channel'-logic turned out to be a promising approach on the one hand, but also resource intensive on the other hand.

Secondly, the context of the project and, more precisely, the interplay between the settings' internal actors was decisive for success or failure of the prevention projects. In both the settings school and sports as well as within certain other projects, supportive tobacco coalitions were key success factors, embedding the projects at the core of the settings. In other cases, such as the project test purchases, missing commitment and support of relevant partners (in this case the police) hampered the implementation and consequently the effectiveness of the prevention.

In summary, the case of Uri supports the three initial outcome-related hypotheses. The findings show that the better the outcome performance of the individual projects, the stronger the overall preventive effectiveness of the program. Additionally, activities which follow a well-designed and rigorously implemented 'Multi-Channel'-logic (in other words a coherent variety of different prevention approaches) lead to better effects on the target groups than isolated projects. Lastly, projects which are situated in a supportive context with committed partners are more likely to yield the desired success.

Context

Our model aims at investigating the role of the cantonal context along the three following dimensions: the initial extent of the public health problem, the socio-structural context, and the political context. These three dimensions turned out to be crucial factors for the effectiveness of the program, although not always showing the expected effect.

We first hypothesized that a comparatively large tobacco-related problem in the canton would increase the program's effectiveness, given the relative higher potential for improvement. In fact, this was not the case in Uri, where a particularly important problem of snuff tobacco negatively impacted the effectiveness of the program. More specifically, very few improvement could be achieved in this regard, and moreover, the activities directed towards snuff rather lowered the overall acceptance for the program within the population. Snuff is a traditionally established practice in the canton of Uri and is associated with local particularism. In this context, prevention regarding snuff was perceived as an external intrusion drove from the capital city without any regard to the local culture.

Second, the socio-structural context negatively affected the program in two respects. On the one hand, the size of the canton and the population were too small for certain activities to be implemented successfully. For example, the canton lacked the critical mass of interested persons to conduct smoking cessation group-courses. For this reason, many of the planned activities had to be cancelled. On the other hand, the small size of the canton also rapidly led to a strong feeling of saturation in regard to tobacco prevention. The program was large and the program managers reported a certain weariness among the population, who felt overwhelmed by smoking prevention. Finally, as already mentioned, the strong inter-knowledge networks and the short distances between prevention actors allowed for rapid action and were identified as positive socio-structural factors.

Third, our hypothesis regarding the political context was not confirmed, as in the case of Uri the administration managed to launch and to implement an important tobacco prevention program in a little experimented canton with minimal political support. Firstly, the prevention entrepreneurs managed to establish some preventive regulatory standards, which secured long-term activities to a certain extent. Secondly, the program leaders managed to compensate for the minimal political support at the local level through the national support, obtained as a result of the new cantonal smoking prevention programs.

To sum up, two of our context hypotheses were not supported by the case study: the initially widespread snuff problem could not be reduced, and the program was successfully implemented with only little political support. The hypothesis related to the socio-structural context was supported, especially as regards the size of the canton.

Conclusion

In this paper, we presented the evaluation design for a complex multilevel program embracing the federal level, the cantonal program level and the project level where target groups are directly addressed. We employ Pawson and Tilley's realist evaluation approach in order to do justice to the varying context factors that impact the cantonal programs leading to varying effectiveness. The application of the model to the canton of Uri shows that the numerous vertical and horizontal relations play a crucial role for the program. As a general learning for the evaluation of complex programs, we first state that the inclusion of all affected levels of a program should be considered and that no monocausal effects can be singled out in programs where multiple interventions address the same problem even if target groups may differ. Second, the case of Uri strongly implies to go beyond the borders of the actual program organization and include factors that do not directly interfere with the policy delivery as such. In the case of Uri, a crucial organizational factor was the relationship between

the federal level with the Fund as main actor and the cantonal program coordinator. Problems in cooperation of these two actors arose in various phases of the program, decreasing the local political support and hampering the implementation. Hence, this finding implies a stronger inclusion of the federal level and, in particular, the history of the collaboration between the national and cantonal actors for outcome and impact evaluation. Finally, Uri is just the first of at least sixteen cantons implementing a tobacco prevention program along the lines of the Fund. Eleven of those new cantonal programs are currently being evaluated within the frame of our research project. The new tobacco policies resulting from these cantonal programs provide stimulating examples of a complex multi-level governance system, combining national and sub-national steering processes. The comparison of these findings will allow for more robust results as to our hypotheses than could be gathered within this single case application.

References

- Bala Malgorzata M., Strzeszynski Lukasz, Topor-Madry Roman, Cahill Kate (2013). "Mass media interventions for smoking cessation in adults", *Cochrane Database of Systematic Reviews* 6: CD004704.
- Befani Barbara, Ledermann Simone, Sager Fritz (2007). "Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application", *Evaluation: The International Journal of Theory, Research and Practice* 13(2): 171–192.
- Ismayr Wolfgang (2009). "Die politischen Systeme Westeuropas im Vergleich", in Ismayr Wolfgang (ed.), *Die politischen Systeme Westeuropas*, 4th ed, Wiesbaden: VS Verlag für Sozialwissenschaften: 9–64.
- Kissling-Näf Ingrid, Wälti Sonja (2007). "The implementation of public policies", in Klöti Ulrich et al. (eds), *Handbook of Swiss Politics*, 2nd ed, Zürich: NZZ: 501–524.
- Knoepfel Peter, Varone Frédéric, Bussmann Werner, Mader Luzius (1997). "Evaluationsgegenstände und Evaluationskriterien", in Bussmann Werner et al. (eds), *Einführung in die Politikevaluation*, Basel/Frankfurt a.M.: Helbing & Lichtenhahn: 78–118.
- Ledermann Simone, Sager Fritz (2009). "Problem erkannt, aber nicht gebannt. Der Nutzen einer Verknüpfung von Konzept- und Umsetzungsevaluation am Beispiel der schweizerischen Strategie 'Migration und Gesundheit'", *Zeitschrift für Evaluation* 8(1): 7–25.
- Linder Wolf (1987). *Politische Entscheidung und Gesetzesvollzug*, Bern: Haupt.
- Linder Wolf (2010). *Swiss Democracy. Possible Solutions to Conflict in Multicultural Societies*, 3rd ed, New York: Palgrave.
- Lund Johan, Aarø Leif Edvard (2004). "Accident prevention. Presentation of a model placing emphasis on human structural and cultural factors", *Safety Science* 42: 271–342.
- Mavrot Céline, Sager Fritz (2014). "Smoking Prevention Policies in Switzerland: Administrative Coalition and Policy Shifts", Paper presented at the 23rd World Congress of Political Science, Montreal, 19-24 July 2014.

- Moser Ladina, Sager Fritz (2015). "Pfadabhängigkeit in der Mehrebenensteuerung: Das Beispiel Palliative Care", *Swiss Political Science Review*. Early online view: DOI: 10.1111/spsr.12168.
- Patton Michael Q. (1997). *Utilization-Focused Evaluation: The New Century Text*, 3rd ed, Thousand Oaks/London/New Delhi: Sage Publications.
- Pawson Ray (2002a). "Evidence-based Policy: In Search of a Method", *Evaluation* 8(2): 157–181.
- Pawson Ray (2002b). "Evidence-based Policy: The Promise of 'Realist Synthesis'", *Evaluation* 8(3): 340–358.
- Pawson Ray, Tilley Nick (1997). *Realistic Evaluation*, London/Thousand Oaks/New Delhi: Sage Publications.
- RS 641.316 : *Ordonnance du 5 mars 2004 sur le fonds de prévention du tabagisme (OFPT)*. Berne : Chancellerie fédérale.
- Sager Fritz (2003). "Kompensationsmöglichkeiten föderaler Vollzugsdefizite", *Swiss Political Science Review* 9(1): 309–333.
- Sager Fritz (2004). "Verwaltung, Politik und Wissenschaft in der kantonalen Alkoholprävention", *Sozial- und Präventivmedizin. International Journal of Public Health* 49(3): 208–215.
- Sager Fritz (2008). "Securing the Long-Term Bases of the Dual System: A Realistic Evaluation of Apprenticeship Marketing Projects in Switzerland", *Journal of Vocational Education & Training* 60(3): 327–341.
- Sager Fritz, Anderegg Céline (2012). "Dealing with Complex Causality in Realist Synthesis: The Promise of Qualitative Comparative Analysis (QCA)", *American Journal of Evaluation* 33(1): 60–78.
- Sager Fritz, Rielle Yvan (2013). "Sorting Through the Garbage Can. Under What Conditions Do Governments Adopt Policy Programs?", *Policy Sciences* 46(1): 1–21.
- Sager Fritz, Rüefli Christian (2005). "Die Evaluation öffentlicher Politiken mit föderalistischen Vollzugsarrangements", *Swiss Political Science Review* 11(2): 101–129.
- Sager Fritz, Zollinger Christine (2011). "The Swiss Political System in a Comparative Perspective", in Trampusch Christine, Mach André (eds), *Switzerland in Europe*, London/New York: Routledge: 27–42.
- Sager Fritz, Hinterleitner Markus (2014). „Evaluation“, in Schubert Klaus, Bandelow Nils C. (eds), *Lehrbuch der Politikfeldanalyse*, München/Wien: De Gruyter/Oldenbourg: 437–462.
- Vatter Adrian (2007). "Federalism", in Klöti Ulrich et al. (eds), *Handbook of Swiss Politics*, 2nd ed, Zurich: Neue Zürcher Zeitung: 77–99.