

Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Mention : « Sciences de la Vie et de la Santé »

par Françoise JABOT

L'évaluation des politiques publiques

**Cadre conceptuel et étude de son utilisation par les décideurs des
institutions régionales de santé en France**

12 novembre 2014

Membres du jury :

Rapporteurs :	Monsieur Pierre LOMBRAIL	PU-PH, Université de Paris 13, Paris
	Madame Elisabeth MONNET	PU-PH, Université de Franche Comté, Besançon
Examineurs :	Monsieur François ALLA	PU-PH, Université de Lorraine, Nancy Directeur de thèse
	Monsieur Jean TURGEON	Professeur associé, Ecole nationale d'administration publique, Québec Co-directeur de thèse
	Monsieur Steve JACOB	Professeur, Université de Laval, Québec
	Monsieur Denis ZMIROU	PU-PH, Université de Lorraine, Nancy

REMERCIEMENTS

Aux Professeurs François Alla et Jean Turgeon qui m'ont guidée avec tact et bienveillance tout au long de ce travail. Tout en restant attentifs à mes arguments, vous m'avez évité de m'égarer sur des chemins parallèles. Toute ma gratitude pour votre soutien et votre grande disponibilité, particulièrement dans la dernière étape de ce parcours.

Aux Professeurs Pierre Lombrail, Elisabeth Monnet, Steve Jacob et Denis Zmirou de me faire l'honneur d'examiner ces travaux. Votre expertise sur le sujet ne peut que me faire progresser.

A toutes les personnes rencontrées dans le cadre de ce travail pour m'avoir donné de leur temps précieux et livré leur point de vue avec beaucoup de sincérité.

A Michel Legros et Jean-Marie André qui m'ont toujours accordé leur confiance et encouragée dans cette voie.

A Annie-Claude Marchand pour sa contribution à ce travail, la justesse de ses analyses et son immense disponibilité à mes nombreuses sollicitations.

A tous mes collègues et complices d'évaluation qui m'ont accompagnée ces dernières années, tout particulièrement : Françoise pour avoir tracé le chemin de l'évaluation à l'ENSP, René pour m'avoir encouragée à le suivre, Murielle pour les nombreuses aventures partagées, Catherine pour tous les projets menés ensemble, Nadine, Cristina, Léopold, Elise, Jean, Françoise, Elisabeth, Dominique pour votre précieuse et fidèle collaboration pédagogique et la convivialité de nos échanges, Stéphane pour ses qualités inégalables d'animateur de groupe, Nathalie pour nos projets de recherche présents et à venir, Anne pour m'avoir intéressée à d'autres approches, et toutes les personnes qui, sans le savoir, m'ont beaucoup appris.

A Maud, ma traductrice préférée et Hugo, pédagogue prometteur.

A Christian pour son soutien constant pendant toutes ces années.

Sada lamndo, lamndo ganndo.

Saa wuuri lamndo ; saa lanndinna majjata.

Si tu interrogues, interroge celui qui sait.

Tant que tu vis, interroge ; ayant interrogé, tu ne t'égareras pas.

(Proverbe peulh)

RESUME

Contexte : Malgré une volonté affichée traduite dans des textes et procédures, l'évaluation en France n'est que peu intégrée dans le processus de décision. Pourtant, dans le secteur de la santé confronté à de multiples défis, l'évaluation pourrait être une approche utile à la détermination des choix. L'utilisation de l'évaluation dépend d'une pluralité de facteurs, dont les connaissances produites et leur mode de production, les caractéristiques des décideurs et autres utilisateurs et le contexte sociopolitique et institutionnel de l'évaluation.

Objectifs : Les objectifs de ce travail sont d'apprécier l'utilisation de l'évaluation dans les politiques de santé, d'identifier les leviers associés à cette utilisation et d'apprécier la capacité de l'évaluation à satisfaire les attentes et favoriser l'utilisation de ses productions.

Méthode : Une revue de littérature a permis d'enrichir la compréhension du concept d'utilisation, d'identifier les facteurs influents et de faire émerger les problématiques associées. Considérant l'évaluation comme un système d'action complexe, un modèle basé sur une approche systémique a été construit et mis à l'épreuve à travers quatre étapes d'analyse des évaluations des plans régionaux de santé publique (PRSP) : (1) analyse globale de 16 évaluations ; (2) lien processus/utilisation dans une région ; (3) résultats à court/moyen terme dans 9 régions ; (4) études de cas approfondies et analyse multicritères dans 5 régions.

Résultats : Une première analyse a éclairé les enjeux du contexte et la contribution des PRSP à la cohérence des politiques régionales de santé. L'examen du processus d'évaluation dans une région a montré la relation entre finalités, démarche et utilisations de l'évaluation. L'observation des suites de l'évaluation dans neuf régions a identifié différentes formes d'utilisation et les principaux facteurs intervenant. La dynamique de changement a été appréhendée dans la globalité des interactions entre le contexte, les acteurs et l'évaluation dans cinq régions.

Discussion : Les retombées de l'évaluation sont plus à chercher du côté des savoirs accumulés et des évolutions des pratiques que dans des décisions formelles transformant de façon radicale les politiques. Le contexte, la crédibilité de l'évaluation, l'engagement et la motivation des acteurs sont des facteurs déterminants. Des pistes de recherche sont envisagées en vue d'approfondir les conditions du développement des capacités et de la culture d'évaluation, elles-mêmes indispensables à une meilleure exploitation de l'évaluation.

Mots clés : complexité, décision, diffusion des connaissances, évaluation, politiques de santé, utilisation.

ABSTRACT

Title: Evaluation of public policies: a conceptual framework and a study of its use by the decision-makers of the health regional institutions in France

Context: Despite the willingness showed in texts and procedures, evaluation in France is little integrated in the decision making process. However, in the health sector which faces multiple challenges, evaluation should be a useful approach to select choices. The use of evaluation depends on multiple factors such as, knowledge and its production process, characteristics of decision makers and others users, and the political and institutional context of the evaluation

Objective: The objectives were: to assess the use of evaluation on health policies; to identify levers associated with use; to assess the capacity of evaluation to fit with decision-makers needs and to enhance usability.

Method: Literature review allowed to enrich the understanding of the concept of use and to identify the main influent factors as well as the related issues. Regarding evaluation as a complex system, a model based on a systemic approach was built and tested in evaluations of regional public health plans (PRSP) in 4 steps: (1) global analysis of 16 evaluations; (2) relation process/use in on region; (3) use at short/medium term in 9 regions; (4) case studies and multicriteria analysis in 5 regions.

Results: A first analysis enlightened the context and the contribution of the PRSP to the coherence of regional policies. A deeper process analysis carried out in one region pointed out the relation between final aims, management and evaluation use. The examination of evaluation consequences conducted in nine regions identified different forms of use and the major factors associated with them. The dynamic of change has been apprehended as a whole through the interactions between context, users and evaluation in five regions.

Discussion: The effects of evaluation are more obvious in terms of knowledge building and evolution of practice than in radical change of policies. Context, evaluation credibility, actors' commitment and motivation are key factors. Future research should help to better understand how to foster the culture and the capacities of evaluation. These are important prerequisites to a wider use of evaluation.

Key words: complexity, decision-making, evaluation use, health policy, knowledge dissemination

SOMMAIRE

<i>Présentation de la thèse</i>	1
<i>Chapitre 1 - Les fondamentaux de l'évaluation</i>	5
1. L'évaluation, opération de jugement, orientée vers l'action	5
2. Les fondements théoriques de l'évaluation	6
3. L'évolution de l'évaluation	8
3.1. L'évaluation sur quatre générations	8
3.2. La diffusion de l'évaluation en quatre vagues	9
3.3. Les enjeux de l'évaluation aujourd'hui	10
4. L'évaluation, une activité spécifique	11
4.1. La distinction entre évaluation, recherche et expertise	12
4.2. L'évaluation, une pratique composite, contrastée, clivée, pluridisciplinaire..	13
<i>Chapitre 2– L'évaluation en France, un positionnement instable</i>	15
1. Repères historiques sur l'évaluation des politiques publiques	15
1.1. L'institutionnalisation de l'évaluation des politiques publiques	16
1.1.1. La naissance de l'évaluation des politiques publiques : une évaluation de type démocratique	16
1.1.2. L'influence de la commission européenne et des contrats avec les collectivités	19
1.2. L'évaluation à l'aune de la culture du résultat	20
1.2.1. Le virage de la performance avec la LOLF	20
1.2.2. La révision générale des politiques publiques	21
1.2.3. La modernisation de l'action publique	21
1.3. Les acteurs de l'évaluation des politiques publiques	22
1.3.1. Les acteurs institutionnels	23
1.3.2. Le rôle fédérateur de la société française de l'évaluation	24
1.3.3. Le faible investissement du monde universitaire	25
1.3.4. La place du secteur privé	26
1.4. Des pratiques hétérogènes mais un fond culturel commun	27
1.5. Peut-on toujours parler d'une « évaluation à la française ? »	28
2. L'évaluation dans le domaine de la santé	29
2.1. L'émergence progressive de l'évaluation dans le secteur de la santé	29
2.1.1. Le développement de la programmation en santé	30
2.1.1.1. Le rôle d'entraînement de l'échelon régional	30
2.1.1.2. L'évaluation au niveau national	32
2.1.2. La mise en place d'agences spécialisées	33

2.1.2.1. Dans le domaine des soins	33
2.1.2.2. Dans le domaine de la santé publique	34
2.1.2.3. Dans le champ médicosocial	37
2.2. La pratique de l'évaluation dans les institutions régionales de santé	38
2.2.1. L'évaluation dans les services déconcentrés	38
2.2.2. L'évaluation dans les ARS	40
2.2.3. Des pratiques hétérogènes	41
2.3. L'évaluation, un concept hybride en tension entre plusieurs visions	42
<i>Chapitre 3 – La problématique</i>	<i>45</i>
1. Objectifs et questions de recherche	45
2. La revue de littérature	46
3. La problématique de l'utilisation au carrefour de plusieurs enjeux (article soumis)	46
<i>Chapitre 4 : La méthode</i>	<i>77</i>
1. Les hypothèses de recherche	77
2. Démarche de recherche et recueil de données	78
2.1. Etape 1 : analyser les évaluations régionales	78
2.2. Etape 2 : analyser le processus d'évaluation et sa relation avec l'utilisation de ses conclusions	78
2.3. Etape 3 : identifier les résultats des évaluations dans neuf régions	79
2.4. Etape 4 : étudier les facteurs d'influence de l'utilisation dans cinq régions...	79
<i>Chapitre 5 : L'influence de l'évaluation sur les politiques régionales</i>	<i>81</i>
1. Les enseignements des évaluations des PRSP sur l'évolution des politiques régionales de santé (publication acceptée)	83
2. L'influence du contexte institutionnel sur le processus d'évaluation : l'exemple de l'évaluation du PRSP de la région PACA (publication)	107
3. La prise en compte des résultats des évaluations dans un contexte réformateur : analyse dans neuf régions (publication)	121
4. L'évaluation, un système d'action complexe: analyse des répercussions de l'évaluation dans cinq régions (article soumis)	131
<i>Chapitre 6 – Synthèse et discussion</i>	<i>161</i>
1. Les usages délibérés et non délibérés de l'évaluation	161
1.1. Le défi de la traçabilité de l'utilisation	162
1.2. Le périmètre de la cible : les utilisateurs	163
2. Les conditions du changement	165
2.1. Le terrain de l'évaluation : atouts et contraintes	165
2.1.1. Les travaux récents	165
2.1.2. Nos résultats face aux conclusions de ces deux recherches	167

2.2.	Les dispositions et comportements des acteurs envers l'évaluation	168
2.3.	Les interactions entre le contexte et les acteurs	170
3.	Le potentiel de l'évaluation	172
3.1.	La crédibilité, élément fondamental du processus d'évaluation	173
3.1.1.	La crédibilité en tension entre validité et acceptabilité	173
3.1.2.	La crédibilité de la démarche d'évaluation	174
3.2.	Trois acteurs clés du dispositif d'évaluation	177
3.2.1.	Le chargé d'évaluation, agent de changement	178
3.2.2.	L'évaluateur : devoir d'indépendance ou devoir d'ingérence ?	179
3.2.3.	Le président de l'instance d'évaluation	181
3.2.4.	L'alliance ou la répartition des rôles	182
	<i>Conclusion</i>	<i>185</i>
	<i>Références bibliographiques</i>	<i>191</i>

Présentation de la thèse

Le rôle et la fonction de l'évaluation ne cessent d'interroger les chercheurs, la question de son utilité étant l'une des plus explorées depuis des décennies (Christie, 2007). Rendre compte de ce qui a été fait, introduire de la rationalité dans la prise de décision, améliorer la performance des politiques, traduisent les principaux buts recherchés, ensemble ou séparément, conduisant à des pratiques diversifiées, en termes de méthode et de conduite d'évaluation. Cependant, malgré un discours engagé en sa faveur, l'instauration de prescriptions et l'installation de procédures pour promouvoir et encadrer la démarche, l'évaluation n'exerce qu'une influence modeste sur la prise de décision.

La situation de l'évaluation en France illustre cette ambiguïté. Engagée au plus haut niveau, implantée dans les institutions, omniprésente dans les textes et les discours, l'évaluation proclamée comme une exigence associée à toute action publique, reste une pratique marginale aux retombées incertaines et souvent accessoires. Le décalage entre cette volonté affichée et une réalité bien en deçà des ambitions aiguise l'intérêt. L'opacité du concept serait-elle une entrave à sa mise en œuvre ? Le nombre de définitions de l'évaluation n'épuise ni le sens attaché à la question de la *valeur* ni les controverses sur la meilleure manière de l'apprécier. Régulation/délibération, sanction/valorisation, problématisation/remise en question, compréhension/quantification, le concept débattu à travers ces multiples oppositions, se prête mal à une démarche fluide entre l'initiation d'une démarche d'évaluation et un usage maîtrisé de sa production.

Dans le domaine des politiques de santé, l'évaluation cristallise des attentes à plusieurs niveaux. L'accroissement des inégalités de santé en dépit, ou à cause, des interventions conduites dans ce domaine, est devenu la préoccupation prioritaire des décideurs. Ce constat critique pour le système de santé interroge la pertinence des stratégies et les effets des politiques conduites ; la mise à jour des mécanismes à l'œuvre étant un préalable à l'élaboration d'actions correctrices. Par ailleurs, les restrictions budgétaires représentent une contrainte majeure conduisant à des choix orientés sur les actions les plus efficaces et à une répartition équitable des ressources. Enfin, la gouvernance de la santé procède aujourd'hui d'un rééquilibrage entre l'Etat, les collectivités territoriales et les citoyens, appelant à plus de concertation entre ces acteurs, dans le souci de faire émerger collectivement des solutions adaptées aux besoins locaux. Face à ces défis pluriels, l'évaluation serait une voie pour, à la

fois, produire la connaissance, répondre à des questions à dimension sociétale, assurer une meilleure répartition des ressources, créer un espace de débat avec divers groupes d'intérêt en vue de l'élaboration de politiques plus efficaces.

Comme dans les autres secteurs, les aspirations formulées ne se traduisent pas toujours par une exploitation des travaux commandés, amenant à s'interroger sur la reconnaissance, le rôle et l'influence de l'évaluation sur les politiques de santé.

L'objectif de cette thèse est de **comprendre les liens qui unissent l'évaluation et la décision en santé**, ouvrir et décrypter la boîte noire entre la production de connaissances à partir de l'évaluation et l'utilisation de ces connaissances pour nourrir la décision.

Deux questions de départ ont orienté cette recherche.

La première interroge la *réalité du changement impulsé par l'évaluation* : dans quelle mesure et à quelle condition, l'évaluation exerce-t-elle une influence sur la prise de décision et le cadre de l'action en santé, à savoir, sur les acteurs, les organisations et les systèmes ?

La seconde questionne la *capacité de l'évaluation à s'ajuster aux attentes et aux besoins de son commanditaire et à favoriser l'utilisation de ses productions* : dans quelle mesure, l'évaluation peut-elle répondre aux attentes des décideurs, à savoir, produire une connaissance de nature à nourrir une décision, dans un univers de complexité croissante ?

Ce travail est centré sur **les politiques régionales de santé**. L'échelon régional a été retenu pour plusieurs raisons. D'une part, considéré comme pertinent pour la mise en œuvre— et selon les époques, l'élaboration - des politiques, il est aussi terrain d'expérimentation de réformes successives, laboratoire de la territorialisation et lieu d'observation privilégié du potentiel créatif régional dans la conjugaison des logiques descendantes, ascendantes et transversales. Les deux lois relatives à la politique de santé ont, ces dix dernières années, introduit des changements notables affectant le pilotage des interventions de santé et impulsé la mise en œuvre de leur évaluation. L'obligation faite à toutes les régions d'évaluer leur plan régional de santé publique (PRSP) constituait une opportunité unique d'exploitation d'un matériau disponible complétant les travaux que nous avons conduits dans ce domaine.

Le document est découpé en six chapitres.

Les deux premiers chapitres ont vocation à introduire le sujet dans son contexte. Aborder la question de l'utilisation de l'évaluation nécessite une clarification du concept et des enjeux de l'exercice afin de mieux saisir les atouts et les obstacles au développement de la pratique en situation réelle. Le premier chapitre souligne quelques notions clés sur l'évaluation tandis que le deuxième situe la pratique dans le contexte français.

Le troisième chapitre développe la problématique de recherche élaborée à partir d'une analyse de la littérature. Celle-ci a permis d'élargir la compréhension du concept d'utilisation, d'identifier les principaux facteurs en jeu et de faire émerger des enjeux associés, d'une part, les questions en lien avec la diffusion et le partage des connaissances, et d'autre part, les défis de l'évaluation confrontée à la complexité des politiques. Ce travail nous a permis de construire un modèle théorique pour analyser l'utilisation des évaluations des politiques régionales de santé.

Le quatrième chapitre expose les différentes étapes de la démarche de recherche. Cependant, le questionnement de ce travail de thèse a émergé au fil des années d'activité d'enseignant-chercheur et de praticien de santé publique. Il est nourri de savoirs théoriques et d'apprentissages accumulés à partir des évaluations réalisées, des formations délivrées et reçues et des échanges qui ont contribué à l'alimenter.

Le cinquième chapitre présente les résultats des travaux conduits aux différentes étapes. Il comprend quatre sections, chacune d'elles exposant les publications correspondantes.

Le sixième chapitre fait la synthèse des travaux et ouvre sur des perspectives de recherche dont les principaux axes sont proposés dans la partie conclusive.

Ce travail a donné lieu à cinq productions principales, insérées dans le corps de la thèse : **quatre articles** dont deux ont été publiés et **un chapitre d'ouvrage** à paraître :

- Jabot F, Turgeon J, and Alla F. A conceptual framework for a better understanding of evaluation influence on policy decision making process. *Evaluation* (soumis).
- Jabot F, Marchand AC. Regard sur la performance du système de santé au niveau régional en France. In: Suárez-Herrera JC, Contandriopoulos AP, Hartz Z et al. (eds) *L'amélioration continue de la performance des systèmes et services de santé : un enjeu*

incontournable, sous la direction Presses de l'Université de Montréal & Conselho Nacional de Secretários de Saúde do Brazil (à paraître décembre 2014).

- Jabot F, Marchand AC (2014) L'évaluation des plans de santé publique : un levier de changement pour les politiques régionales de santé en France ? *Global Health Promotion*, 21, supp.1, 64-69.
- Jabot F, Turgeon J, and Carbonnel L (2011) The Evaluation of the PACA Regional Public Health Plan: Reconciling the Managerial, Scientific and Democratic Finalities. *Evaluation and Program Planning*, 34, 196-205.
- Jabot F, Turgeon J and Alla F. Contributions and limits of evaluation in reviewing health policies: in-depth analysis in five French regions. *Evaluation* (soumis).

Chapitre 1 - Les fondamentaux de l'évaluation

Le concept d'évaluation est difficile à cerner tant il renvoie à des acceptions et des références théoriques variées selon les époques, les contextes et les individus. Polysémique et généraliste, le mot évaluation est généralement qualifié par un autre terme qui en précise l'objet (évaluation de politique, de programme, de pratiques, de réseaux, de personnes), les finalités poursuivies (évaluation sommative, formative, managériale, démocratique), le type d'approche (évaluation collaborative, participative, pluraliste), le type de questionnement (évaluation de la pertinence, de l'efficacité, d'impact...), la discipline ou les méthodes mobilisées (évaluation socioéconomique, évaluation qualitative ou quantitative), le mode opératoire (évaluation externe ou interne), la temporalité (*ex ante*, intermédiaire, finale, *ex post*) ou la nature de l'activité (recherche évaluative, suivi évaluatif). Cette liste évoque d'emblée la diversité des formes d'exercice, des approches et les multiples classifications dont l'évaluation a été l'objet.

L'objectif de ce chapitre est, à partir de quelques notions clés, de clarifier ce que recouvre l'évaluation, l'idéal type approché par les auteurs, d'essayer de tracer une ligne de partage entre ce qui est du registre de l'évaluation, de ce qui n'en est pas et de s'accorder sur quelques postulats. A cette fin, un détour sur les fondements théoriques de l'évaluation et sur les principales étapes de son évolution facilitera la compréhension des typologies et des traditions dans lesquelles s'inscrivent les évaluateurs mais également, leurs interlocuteurs.

1. L'évaluation, opération de jugement, étayée sur la méthode et orientée vers l'action

Au-delà de la diversité des définitions qui, selon les auteurs, mettent l'accent sur une composante particulière, on note une convergence sur la nature de l'évaluation : une opération qui consiste à construire un jugement (Mathison, 2005).

« L'évaluation est ce qu'elle est, la détermination du mérite ou de la valeur, et ce à quoi elle sert est une tout autre question » (Scriven, 1980, p.7).

« L'évaluation est « le processus pour décrire, obtenir et fournir de l'information pour juger des alternatives à la décision » (Stuffelbeam, 1973, p.129).

Par ailleurs, l'évaluation est définie en tant que méthode pour traiter l'information et ne peut être distinguée des finalités poursuivies.

« L'évaluation de programme est la collecte systématique de données au sujet des activités, des caractéristiques et des effets des programmes, afin de porter des jugements sur ces derniers, d'améliorer leur efficacité et/ou d'être en mesure de prendre des décisions plus éclairées au sujet des programmes futurs » (Patton, 1997, p.23).

Enfin, d'autres définitions mettent l'accent sur le questionnement.

« L'évaluation doit se préoccuper de l'utilité, de la mise en œuvre, de l'efficacité et de l'efficience des mesures qui ont pour but d'améliorer le sort des membres de la société » (Rossi et al., 2003, p.35).

2. Les fondements théoriques de l'évaluation

D'où vient l'évaluation ? Alkin (2004) utilise la métaphore de l'arbre pour présenter le développement de l'évaluation. Pour lui, les racines de l'évaluation sont doubles, d'une part, dans la recherche en sciences sociales, et, d'autre part, dans le contrôle et la reddition de comptes¹. En fonction de la croissance de l'une ou l'autre de ces racines, l'évaluation s'est développée dans trois directions (branches), selon qu'étaient privilégiées, la méthode, la notion de valeur ou l'utilisation de l'évaluation. Analysant leurs théories, Alkin positionne les théoriciens de l'évaluation, principalement nord-américains, (nous n'en citerons que quelques-uns) sur la branche qui représente le mieux le courant dans lequel ils s'inscrivent.

La branche du milieu, *Méthode*, dérive de la recherche en sciences sociales. Fondée sur le paradigme scientifique, l'évaluation, considérée comme une activité de recherche, est guidée par les méthodes en vue de produire de la connaissance et des données probantes, objectives, valides. La preuve de l'efficacité, de même que les questions de validité et de reproductibilité sont fondamentales. Tylor, Campbell, Rossi, Chen et Weiss sont positionnés sur cette branche : Campbell pour ses travaux sur les devis expérimentaux, Tylor pour le développement de l'évaluation basée sur les objectifs, Chen pour la théorie de programme visant à expliciter les relations de causalité. Selon cette orientation, l'évaluation est une activité scientifique et l'évaluateur adopte une posture de chercheur.

¹ Traduction de *accountability*.

La branche *Valeur* représente l'évaluation orientée vers le jugement. L'évaluateur joue un rôle fondamental car c'est lui qui choisit les critères et le *modus operandi* pour introduire du jugement dans les données. C'est le modèle de l'évaluation expertise au service des décideurs. Les théories sont inspirées des travaux de Scriven pour qui, « ce qui est mauvais est mauvais et ce qui est bon est bon, et c'est à l'évaluateur qu'il incombe de décider lequel est lequel » (Scriven, 1986, p.19). D'autres auteurs adoptent cette orientation avec toutefois des nuances. Pour House, l'évaluateur ne détermine pas ce qui est bon ou mauvais mais ce qui est juste. Guba et Lincoln considèrent que la détermination du mérite revient aux parties prenantes, dont les perceptions et interprétations traduisent les multiples aspects de la réalité ; le rôle de l'évaluateur est de faciliter la négociation des différents points de vue.

La branche *Utilisation* porte les théories de l'évaluation centrée sur l'utilisation. L'évaluation doit aider les utilisateurs à prendre des décisions pour leur programme. Stufflebeam a développé le modèle CIPP (Context, Inputs, Process, Products) centré sur le processus de décision. Selon son approche, l'évaluateur s'entoure d'un collectif avec qui définir les questions d'évaluation, le devis, réviser le rapport et en diffuser les conclusions. Il a par ailleurs identifié les quatre principaux standards (utilité, faisabilité, précision, propriété) devant guider la pratique des évaluateurs. Wholey qui définit l'évaluation comme une forme appliquée de la recherche en sciences sociales, insiste sur l'importance de s'accorder sur les buts du programme et les critères de performance ; il a développé la notion d'évaluabilité (*evaluability assessment*), exercice qui permet d'apprécier l'utilité et la faisabilité de l'évaluation avant de s'y engager. Patton (1997) est un des auteurs qui a le plus développé la théorie de l'utilisation (*utilization-focused evaluation*). Il considère que l'évaluation doit être entièrement orientée sur les besoins des utilisateurs et préconise d'identifier les véritables utilisateurs de l'évaluation, « *intended primary users* », l'intérêt des personnes envers l'évaluation étant le facteur le plus influent de l'utilisation. Plus récemment, Patton (2010) a introduit l'évaluation développementale (*developmental evaluation*) selon laquelle l'évaluateur est partie prenante de l'équipe du programme, son rôle étant d'en aider le développement. Alkin, l'auteur de l'ouvrage figure sur cette branche dans la lignée de Patton, ainsi que les auteurs, tels que Cousins, Preskill, et Fetterman, qui soutiennent les approches participatives et habilitantes. L'évaluateur aide les utilisateurs à choisir les modèles et approches en fonction de leur situation particulière.

L'arbre de Alkin, révisé depuis sa première édition, présente l'avantage de situer les auteurs et leurs théories dans une filiation, de comprendre les proximités et les différences entre ces théories et, transposé à notre contexte, leurs prolongements dans les représentations et pratiques d'évaluation actuelles.

3. L'évolution de l'évaluation

D'autres auteurs ont analysé la trajectoire de l'évaluation à travers son évolution dans le temps en reliant les principales étapes de son histoire aux contextes sociopolitiques.

3.1. L'évaluation sur quatre générations

Considérant l'évaluation dans sa chronologie, Guba et Lincoln (1984) ont proposé une division en quatre générations. Le résumé de cette évolution s'appuie sur la synthèse présentée par Dubois et al. (2009).

Héritage du siècle des Lumières, l'évaluation de première génération (début XIX^{ème} aux années 30) place la rationalité au cœur de la démarche. La recherche scientifique est la voie privilégiée pour accroître la connaissance indispensable à la construction des politiques destinées à l'amélioration des conditions de vie. Bénéficiant des apports de la statistique et de l'épidémiologie, l'évaluation élabore des outils de collecte et de mesure pour documenter, quantifier, comparer les conditions de vie des populations. C'est l'ère de la mesure et l'évaluation demeure l'apanage des techniciens et experts.

L'évaluation de deuxième génération (années 30-60) s'est développée en réaction aux insuffisances des méthodes précédentes, lesquelles ne rendraient pas compte des performances des programmes. Les travaux de Tyler sur la mesure de l'atteinte des objectifs sont alors déterminants dans la diffusion de cette approche. C'est la naissance de l'évaluation formative car l'évaluation est attendue sur des propositions d'amélioration des programmes. Cette période est marquée par deux convictions : l'obtention d'un progrès social avec les politiques menées et l'utilisation des connaissances pour améliorer les politiques. Outre les activités de mesure, l'évaluateur doit apporter des éléments de compréhension et des perspectives d'amélioration. Davantage descriptive, l'évaluation émerge comme activité propre et différente de la seule mesure.

L'évaluation de troisième génération (années 60-70) prend acte des critiques formulées sur les méthodes, jugées insuffisantes pour éclairer la décision. Il manque des éléments pour juger de la valeur et de la pertinence des objectifs des programmes et c'est à l'évaluateur qu'il revient d'accomplir cette fonction. C'est à cette période que se situent les travaux de Stufflebeam (modèle CIPP) et de Scriven sur la construction du jugement évaluatif. L'évaluation acquiert une légitimité scientifique pour la décision. Cependant, constatant qu'elle est peu utilisée, ou utilisée sans relation avec la qualité scientifique des travaux, plusieurs auteurs, dont Weiss (1997) vont approfondir la problématique de l'utilisation et développer l'approche centrée sur les utilisateurs. La production d'ouvrages pédagogiques, de revues, la création d'associations professionnelles, la mise en place de normes attestent de la construction d'un champ professionnel spécifique à l'évaluation.

L'évaluation de quatrième génération (à partir des années 90) accorde sa faveur à une approche différente, fondée sur le paradigme constructiviste. Guba et Lincoln (1984) s'inscrivent dans la tradition ontologique qui postule que la réalité en soi n'existe pas et soutient que la réalité est le produit d'une construction sociale. Dans cet esprit, la connaissance requise pour l'évaluation se construit avec les parties prenantes de l'évaluation. Cette approche élargit le cercle de l'évaluation à d'autres acteurs que les seuls experts, positionne l'évaluateur comme négociateur et facilitateur et préconise le pluralisme des méthodes. Elle sera le tremplin au développement des approches participatives. L'évaluation n'est pas seulement un outil au service des décideurs, c'est aussi un instrument d'émancipation.

Ces quatre générations, privilégiant des normes très différentes, coexistent encore aujourd'hui, comme l'illustre l'auteur suivant.

3.2. La diffusion de l'évaluation en quatre vagues

Plus récemment, Vedung (2010) a retracé l'évolution de l'évaluation des quarante dernières années, avec un focus particulier sur la Suède et, dans une moindre mesure, les pays du Nord. Pour cet auteur, l'évaluation est le produit de la sédimentation d'éléments déposés par vagues successives. Il associe, s'agissant de la Suède, le flux des différentes vagues aux idéologies politiques des gouvernements. Sa lecture de l'histoire, brièvement résumée ci-dessous, prolonge la présentation précédente.

Le première vague, *scientifique*, déferle dans les années 60. L'évaluation est basée sur le paradigme positiviste. L'évaluation, qui mobilise un arsenal de méthodes scientifiques, introduit de la rationalité dans les politiques. Le modèle d'ingénierie de l'action publique permet l'interaction entre la décision publique et le monde scientifique. L'évaluation est conçue dans une vision instrumentale au service de la décision. Selon la classification précédente, elle serait dans sa troisième génération.

La seconde vague, *orientée vers le dialogue*, présente les caractéristiques de l'évaluation de quatrième génération. L'évaluation produit de la connaissance et du sens avec les parties prenantes. C'est une arène pour l'expression de la démocratie.

La troisième vague, *néolibérale*, réintroduit la science, l'évaluation démocratique basée sur les idéologies étant jugée biaisée. Inscrite dans la réforme du *New Public Management*, l'évaluation met alors l'accent sur le management basé sur les résultats, la mesure de la performance, le développement des outils de suivi. La satisfaction des consommateurs (bénéficiaires) a relayé les points de vue des parties prenantes sur les opinions et besoins à prendre en compte.

La quatrième vague, *retour vers expérimentation*, (1995-2000) signe la résurgence de la science avec la quête de la preuve (*evidence*). La production croissante de revues systématiques, la hiérarchie des niveaux de preuve qui place les études randomisées au sommet donnent un autre visage à l'évaluation.

La métaphore des vagues est tout à fait illustrative de la constitution par strates successives de la culture d'évaluation, chaque vague apportant son lot de transformations sans pour autant emporter les sédiments de la vague précédente.

3.3. Les enjeux de l'évaluation aujourd'hui

Les enjeux de l'évaluation aujourd'hui témoignent de ces évolutions.

La sophistication des méthodes reste d'actualité. A côté des méthodes expérimentales, d'autres méthodes sont à l'honneur : l'approche *Realist* de Pawson et Tiley (1997), l'évaluation basée sur la théorie (Chen, 2010), la formalisation des méthodes mixtes (Guével and Pommier, 2012), la recherche de méthodes adaptées à l'évaluation des interventions

complexes (Walton, 2014). L'analyse du contexte comme cadre de référence fait désormais partie de l'arsenal des méthodes d'évaluation (Rog et al., 2012).

La recherche sur les formes d'évaluation participative, collaborative, habilitante qui mettent l'accent sur la participation des parties prenantes (Rodriguez-Campos, 2011) a capté l'intérêt des praticiens de l'évaluation.

4. L'évaluation, une activité spécifique

Les exposés précédents ont montré la pluralité des formes prises par l'évaluation et des théories qui la sous-tendent. L'idéal de la pratique s'incarne dans une chaîne logique entre les attentes envers l'évaluation, le questionnement, l'approche et la méthode choisie.

Les buts de l'évaluation sont multiples : juger de la valeur, rendre compte, juger de la conformité, améliorer le programme, développer la connaissance (Fitzpatrick et al., 2009). La finalité de l'évaluation, point de départ de la démarche, détermine la stratégie et l'approche de l'évaluation. La stratégie est définie comme « la direction générale pour atteindre le but d'une évaluation donnée² » et l'approche comme « l'ensemble des procédures et principes guidant l'évaluation pour satisfaire les attentes des parties prenantes³ » (Chen, 2004, p.144).

Le choix de l'approche dépend de la commande institutionnelle d'évaluation mais aussi de la pratique de l'évaluateur. Certains ajustent leur modèle au contexte, d'autres utilisent leur modèle favori quel que soit le contexte et une partie d'entre eux ne travaillent que dans des contextes qui leur conviennent (Contandriopoulos and Brousselle, 2009).

L'habilité de l'évaluateur s'apprécie dans la façon dont il combine connaissances, savoirs expérientiels, perception du contexte et affects pour appliquer sa théorie de l'évaluation sur un terrain à chaque fois nouveau et souvent imprévisible (Kundin, 2010). Stevahn et al. (2005) ont fourni un cadre d'analyse des compétences essentielles requises pour les évaluateurs de programme. Ces dernières associent des qualités méthodologiques, une connaissance du contexte et des aptitudes personnelles et relationnelles (Tourmen, 2009).

Ces particularités nous aident à distinguer l'évaluation d'autres formes de production de connaissances, que sont la recherche et l'expertise.

² Notre traduction.

³ Idem.

4.1. La distinction entre évaluation, recherche et expertise

L'évaluation présente à la fois des traits communs et des différences avec la recherche et l'expertise.

L'évaluation, une forme de recherche appliquée

Wholey (2004) considère l'évaluation comme une forme appliquée de recherche en sciences sociales. Son but principal est de regarder les résultats du programme pour rendre compte, améliorer la performance du programme et accompagner la décision.

Pour Weiss (1997) et Levin-Rozalis (2003), l'évaluation présente des similitudes et des différences avec la recherche. Levin-Rozalis (2003) soutient que l'évaluation dépend avant tout de l'action. Pour Weiss (1997), l'évaluation se rapproche de la recherche parce qu'elle fait appel aux mêmes méthodes de recherche et techniques d'enquête. Cependant, elle s'en distingue par le cadre de la commande et les compétences requises. Les questions évaluatives sont orientées par les décideurs et les praticiens en fonction de l'utilisation prévue de l'évaluation.

« Ce qui distingue l'évaluation de la recherche, ce n'est pas la méthode ou le sujet mais le but pour lequel elle est faite⁴ » (Weiss, 1997, p.15).

Cet auteur considère que l'évaluation exige un plus haut niveau de compétences que la recherche afin de produire une recherche rigoureuse, utile et adaptée à la complexité du monde réel.

Perret (2009c, p.55) insiste sur le but de l'évaluation qui est « non pas de produire des données scientifiques, mais de réduire l'incertitude et de renforcer la cohérence des décisions et des pratiques ».

⁴ Notre traduction.

L'évaluation, une forme d'expertise collective

L'évaluation est-elle synonyme d'expertise ? Tabuteau (2010) explorant les figures de l'expert et ses relations avec la décision en santé publique, débute son propos par un rappel de l'étymologie du terme. L'expertise est avant tout définie par celui qui la réalise. Le Littré définit les experts comme les « personnes qui ayant la connaissance acquise de certaines choses, sont commis pour les vérifier et en décider ». Tabuteau identifie les deux fonctions de l'expert : produire des connaissances nécessaires à la décision - l'expert « antidote de l'ignorance » ; produire les méthodes et cadres de réflexion en vue de les interpréter.

De ce point de vue, l'évaluation tend à se confondre avec l'expertise, rejoignant la conception de Sriven. Cependant, la reconnaissance par plusieurs auteurs du rôle des parties prenantes dans la négociation des référentiels de jugement en fait un exercice d'une autre nature. L'étendue des acteurs impliqués, leur degré de participation et le rôle de l'évaluateur dans l'animation de la démarche composent une palette de situations, de l'évaluation experte et dissociée de l'action à une activité clairement délibérative.

4.2. L'évaluation, une pratique composite, clivée, contrastée, pluridisciplinaire

« A la croisée de la connaissance et de l'action⁵ », l'évaluation est une pratique composite, contrastée, clivée et résolument pluridisciplinaire.

Composite parce qu'elle est façonnée par un ensemble de paramètres : les attentes des demandeurs, les finalités et enjeux, les types d'acteurs engagés, la nature de leur implication et leurs représentations de l'évaluation, les approches, les méthodes, la temporalité de l'évaluation et les contextes sociopolitiques ;

Clivée à cause des deux traditions philosophiques radicalement différentes qui la sous-tendent (Dubois and Marceau, 2005) ; ainsi, pour Trosa (2010), l'évaluation demeure « dans un équilibre difficile entre le positivisme et le constructivisme, entre la preuve et la construction de l'argumentaire ».

Contrastée parce qu'elle épouse des formes variées, selon la prédominance des dimensions (cognitive, instrumentale, normative, démocratique) qu'elle porte ;

⁵ Selon la formule de Guy Cauquil (Jabot, 2007).

Pluridisciplinaire parce qu'elle emprunte les méthodes d'une variété de disciplines pour recueillir et traiter les données qui lui sont utiles.

Le chapitre suivant situe l'évaluation dans le contexte français, précisant les étapes qui ont jalonné son histoire, les courants principaux qui la traversent et les modèles de référence, souvent implicites qui étayent les pratiques.

Chapitre 2– L'évaluation en France, un positionnement instable

En France, l'évaluation suscite depuis quelques années, un intérêt croissant. Inscrite dans les textes, imposée ou suggérée aux acteurs, elle s'insinue progressivement dans les pratiques. Cependant, elle n'est encore que peu intégrée dans le processus de prise de décision (Lacouette-Fougères et Lascoumes, 2013 ; Perret, 2009a). Dans le secteur de la santé, les décisions affectant les plans et programmes de santé, tant au plan national que régional, semblent peu corrélées avec les apports des travaux d'évaluation. Pourtant, face à des choix contraints par des ressources budgétaires limitées et à la nécessité d'en rendre compte aux citoyens, les décideurs sont amenés à orienter leur décision vers des actions les plus efficaces voire les plus efficaces et l'évaluation pourrait être un outil utile à la détermination des choix.

Dans une première partie, ce chapitre résume les principales étapes de la genèse de l'évaluation des politiques publiques en France et présente les institutions principalement engagées. La seconde partie s'attarde sur l'évaluation dans le domaine de la santé.

1. Repères historiques sur l'évaluation des politiques publiques

Même s'il est d'usage de rappeler que l'évaluation en France figure dans l'article 15 de la Déclaration des Droits de l'Homme et du citoyen de 1789, qui stipule que « la société a le droit de demander compte à tout agent public de son administration », l'histoire de l'évaluation est, comparativement au monde anglo-saxon, relativement récente en France (Barbier, 2010).

C'est l'expérience de plusieurs années de mise en place de la rationalisation des choix budgétaires (RCB) qui a créé l'amorce de la culture d'évaluation dans l'administration (Fontaine and Monnier, 2002 ; Perret, 2003). Inspiré du modèle américain, la RCB est fondée sur la recherche de l'optimisation des choix budgétaires et initie l'évaluation *ex ante*. Il s'agit, préalablement à la décision de programmation budgétaire, de définir des objectifs quantifiés et des indicateurs chiffrés et d'identifier les possibilités d'action et les alternatives à soumettre au décideur. Ainsi, entre 1970 et 1984, plusieurs centaines d'études ont été réalisées dans différents secteurs. Cette procédure lourde et jugée trop technocratique a rencontré des résistances de la part des administrations et a finalement été abandonnée.

Dans le courant des années 80, plusieurs événements marquent une curiosité, si ce n'est un intérêt, envers l'évaluation. Un premier état des lieux des pratiques est dressé à l'occasion d'un colloque organisé en 1983. Basé sur l'examen de trois cent évaluations *ex post*, il mène à plusieurs constats : l'évaluation souffre d'une faible visibilité, d'une conception floue, de travaux d'inégale qualité, notamment en raison d'un déficit de rigueur et d'objectivité, et d'une faible insertion dans le processus de décision. Les actes de ce colloque feront l'objet du premier ouvrage traitant de l'évaluation des politiques publiques en France (Nioche et Poinard, 1984). Deux ans plus tard, le rapport Deleau (1986) réalisé par un groupe de travail du Commissariat général au Plan relance la réflexion en défendant une évaluation de type managérial, orientée sur la mesure des effets réels de la politique. Selon ce point de vue, l'évaluation doit être objective, indépendante, scientifique, affranchie de jugement de valeur ; l'évaluateur est un expert, légitimé par ses compétences scientifiques et méthodologiques (Duran et al., 1995).

C'est en 1989 que la question de l'évaluation réapparaît à la faveur du mouvement de « modernisation de l'Etat » initié par le nouveau gouvernement. Le second rapport (Viveret, 1989) adopte une approche radicalement différente et va inspirer la parution du décret grâce auquel l'évaluation fera son entrée dans les institutions.

1.1. L'institutionnalisation de l'évaluation des politiques publiques

L'ancrage de l'évaluation dans les institutions en France relève d'une histoire rythmée par la mise en place et la transformation des dispositifs institutionnels en lien avec les réformes successives de l'Etat. Pour Jacob and Varone (2004), l'institutionnalisation se traduit par l'instauration de processus formels et informels, procéduraux et organisationnels, tels que la création d'instances, l'introduction de clauses évaluatives et le développement de pratiques en réseau.

1.1.1. La naissance de l'évaluation des politiques publiques : une évaluation de type démocratique

Le rapport Viveret (1989) entend rompre avec la conception de l'évaluation qui avait prévalu lors des années précédentes, à la fois technocratique et peu distincte du contrôle. Selon ce

rapport, l'évaluation définie comme « un jugement sur la valeur » de politiques publiques, doit répondre à une triple exigence : méthodologique, démocratique et d'indépendance. Il prône une évaluation construite à partir de la confrontation des points de vue et des intérêts divergents et génératrice d'un débat avec la société civile. Cette vision d'une évaluation démocratique relaye la conception de l'évaluation managériale défendue dans le rapport Deleau.

La circulaire du 23 février 1989 sur le renouveau du service public précise : « Il ne peut y avoir ni autonomie sans responsabilité, ni responsabilité sans évaluation, ni évaluation sans conséquence ».

Le Décret n° 90-82 du 22 janvier 1990

Avec le décret de 1990, l'évaluation acquiert une reconnaissance en tant qu'activité spécifique incarnée dans un dispositif institutionnel. L'évaluation est ainsi définie :

« L'évaluation d'une politique a pour objet de rechercher si les moyens juridiques, administratifs ou financiers mis en œuvre, permettent de produire les effets attendus de cette politique et d'atteindre les objectifs qui lui sont assignés ».

Deux structures sont dévolues à l'évaluation : le comité interministériel de l'évaluation (CIME) assure le pilotage les travaux d'évaluation ; le conseil scientifique de l'évaluation (CSE) est chargé de promouvoir une réflexion sur les méthodes et l'éthique de l'évaluation.

Le comité interministériel de l'évaluation a initié au cours de son mandat, de 1990 à 1998, vingt-trois projets d'évaluation parmi lesquels seize ont été achevés, dont l'évaluation du Revenu minimum d'insertion, de la politique du logement, de la maîtrise de l'énergie, de l'aménagement des rythmes de vie de l'enfant, de la politique de la montagne, ou encore de la loi Evin.

Le Conseil scientifique de l'évaluation a progressivement élaboré un corpus théorique traduit dans un guide qui demeure à ce jour une référence (CSE, 1996). L'évaluation est définie comme une activité spécifique, distincte d'autres activités, notamment du contrôle et de l'audit. Elle a vocation à apporter de la connaissance (dimension cognitive), produire un jugement sur la politique (dimension normative), contribuer à l'amélioration de la politique (dimension instrumentale) et faciliter le débat (dimension démocratique). Sur le plan méthodologique, le CSE insiste sur le pluralisme des disciplines et des méthodes, la rigueur de

l'argumentation, l'exigence de transparence ainsi que sur la « zone d'autonomie » à préserver vis-à-vis du commanditaire. Afin de garantir l'indépendance du processus d'évaluation, le CSE préconise la mise en place d'une instance d'évaluation, de nature à refléter la diversité des points de vue, sans être pour autant un lieu de négociation des intérêts des parties prenantes (Perret, 1997). Enfin, il suggère d'évaluer les politiques publiques à partir de six critères majeurs : pertinence, cohérence, efficacité, efficience, atteinte des objectifs et impact.

La lenteur des travaux, l'importance inégale des thèmes, l'intérêt insuffisant du gouvernement et le peu de suites donné à ces évaluations auront raison de ce dispositif qui sera rénové au terme de huit années de fonctionnement.

Le Décret n°98-1048 du 18 novembre 1998

Ce nouveau décret simplifie le dispositif et révisé la définition de l'évaluation, mettant l'accent sur l'identification des effets et l'efficience de la politique :

« L'évaluation d'une politique publique a pour objet d'apprécier, dans un cadre interministériel, l'efficacité de cette politique en comparant ses résultats aux objectifs assignés et aux moyens mis en œuvre ».

Le Conseil national de l'évaluation (CNE), assisté du Commissariat général au Plan qui en assure le secrétariat, reprend les compétences administratives du CIME et scientifiques du CSE. Son champ d'application est élargi au domaine des collectivités territoriales. Il s'agit désormais de produire des évaluations dans des délais brefs afin de répondre à des questions claires et conduisant à des décisions opérationnelles.

Entre 1999 et 2002, quinze évaluations ont été réalisées. Les derniers travaux d'évaluations concernaient : la politique de lutte contre le sida, la politique de préservation de la ressource en eau, le transport combiné rail-route, la sécurité routière ou encore la formation continue des agents de l'Etat. Bien que la durée des travaux ait été considérablement raccourcie, dix-huit mois en moyenne, elle était encore jugée trop longue.

Ce dispositif sera « mis en sommeil⁶ » en 2003 après plus de dix ans de fonctionnement avant d'être définitivement supprimé en 2007.

En 2004, un rapport du Sénat (Bourdin et al., 2004) fait une critique sévère de l'évaluation des politiques publiques en France. Il déplore le défaut de structuration des politiques publiques, handicap à l'évaluation, l'éparpillement des activités d'évaluation, la lourdeur du dispositif, l'effacement du Parlement, le déficit de capitalisation des expériences, le manque d'indépendance de l'évaluation vis-à-vis du pouvoir exécutif en raison de la pratique d'autoévaluation au sein des administrations. Cependant, le rapport approuve les principes d'indépendance et d'autonomie préconisés par le CSE et traduits dans le fonctionnement des instances pluralistes. Pour pallier les défauts du dispositif « cathédrale », les auteurs du rapport proposaient une institutionnalisation en « toile d'araignée » avec la création d'une Haute autorité de l'évaluation et la mise en place de commissions d'évaluation par domaines de politiques publiques. Ce rapport restera sans suites.

1.1.2. L'influence de la commission européenne et des contrats avec les collectivités

Parallèlement au mouvement de réforme de l'Etat, le double processus de décentralisation et d'intégration européenne a stimulé le développement de l'évaluation (Duran et al., 1995). Le développement des politiques contractuelles entre l'Etat et les collectivités territoriales (contrats de plan Etat Région, fonds structurels européens) et la présence de clauses évaluatives dans les actions financées sur des fonds européens ont contribué, au cours des années 2000, d'une part, à introduire l'évaluation au niveau régional (Duran et al., 1995 ; Le Bouler, 2003) et, d'autre part, à engager des évaluations partenariales. Cette obligation, accompagnée par la production de standards d'évaluation et la diffusion de guides, a jeté les germes d'une culture d'évaluation.

⁶ Selon l'expression de Stéphane Le Bouler qui assurait à cette période le secrétariat du Commissariat général au Plan.

1.2. L'évaluation à l'aune de la culture du résultat

Si le concept d'évaluation de politiques publiques a prévalu jusqu'alors, la réforme budgétaire engagée par la loi organique du 1^{er} août 2001 relative aux lois de finances (LOLF) introduit une nouvelle orientation, la mesure de la performance, et avec elle, le management par les résultats (Trosa and Perret, 2005).

1.2.1. Le virage de la performance avec la LOLF

La LOLF prescrit une nouvelle architecture selon laquelle le budget, présenté précédemment sous la forme de chapitres budgétaires, s'organise désormais autour de programmes et de missions. Les politiques publiques sont traduites en missions, elles-mêmes déclinées en programmes, eux-mêmes constitués d'un ensemble d'actions. Ainsi, l'action publique de l'Etat est découpée en quarante-sept missions, cent cinquante-huit programmes et cinq cent quatre-vingt actions. Chaque ministère est tenu, pour sa ou ses missions, de définir des programmes annuels de performance sur la base d'objectifs de performance en nombre limité, assortis des indicateurs adéquats, et d'établir des rapports annuels de performance. Ces derniers sont destinés à mieux informer le Parlement sur les résultats de l'action publique afin qu'il soit en mesure de faire des choix stratégiques et budgétaires plus éclairés.

Dans l'esprit de la LOLF⁷, la performance est définie comme « la capacité d'une organisation à tenir la totalité de la chaîne des résultats », à savoir traduire l'utilisation des crédits en réalisations/résultats/impacts. La mesure de performance consiste à apprécier les impacts de l'action de l'Etat, ce qui la différencierait de l'évaluation qui, d'une part, porte sur les politiques qui ne sont pas limitées à l'action de l'Etat et, de fait, impliquent d'autres acteurs, dont les collectivités locales, et, d'autre part, s'intéresse aux impacts lointains de ces politiques.

Les énoncés de cette réforme ont doublement brouillé les repères : d'une part, en réduisant la notion de programme à un ensemble de crédits, et, d'autre part, en redessinant le périmètre de l'évaluation ; limitée à l'appréciation des impacts à long terme, elle est tenue éloignée des pratiques de routine. La notion de performance devient synonyme d'efficacité, alors même que sa mesure se limite à une agrégation d'indicateurs dont la restitution ne dit rien du lien entre

⁷République française (2003) *Fiche 4, Différents types d'objectifs et d'indicateurs de la LOLF*, [consulté en ligne le 13 mars 2006], <http://www.performance-publique.gouv.fr/performance/etat/objectifs.htm>.

action mesurée et résultats. De ce fait, la mesure de performance ne saurait remplacer l'évaluation, seule apte à produire une intelligibilité de l'action mise en œuvre et entamer une réflexion sur les théories d'action qui supportent les politiques (Perret, 2006).

L'exercice promu par la LOLF, outil d'allocation budgétaire par excellence, a provoqué de nombreux débats sur le devenir de l'évaluation ; les uns redoutant de la voir érudite au profit des exercices purement budgétaires, les autres y voyant une voie de progrès. Avec un recul de plusieurs années, il s'avère que la LOLF n'a pas réussi à développer les activités d'évaluation, même si elle a contribué à une meilleure connaissance de l'action publique (Lacouette-Fougères et Lascoumes, 2013).

1.2.2. La révision générale des politiques publiques

Deux événements marquent l'année 2007: la création d'un éphémère secrétariat d'Etat de la prospective et de l'évaluation, lequel disparaîtra en 2009, et le lancement de la révision générale des politiques publiques⁸ (RGPP). Cette dernière consiste, à partir de sept questions clés puisées dans le répertoire évaluatif, à procéder à l'examen de grandes politiques publiques dans l'objectif de revoir les missions de l'Etat, rationaliser les dépenses publiques et identifier les leviers de réforme. Faute de temps, de concertation et de méthode, la démarche, plus proche de l'audit que de l'évaluation, n'a pas vraiment permis de juger de l'efficacité des politiques étudiées.

En réalité, la RGPP tout comme la LOLF ont emprunté la sémantique de l'évaluation plus qu'elles n'en ont appliqué la démarche (Lacouette Fougère and Lascoumes, 2014).

1.2.3. La modernisation de l'action publique

La modernisation de l'action publique (MAP) lancée en 2012, s'inscrit dans la lignée des démarches précédentes avec toutefois, quelques éléments de changement, liés d'une part, à l'élargissement du périmètre aux collectivités territoriales et aux associations de protection sociale et, d'autre part, à la recherche de concertation avec les parties prenantes. La MAP

⁸ République française (2007). *Révision générale des politiques publiques. Fiche de synthèse*, [consulté en ligne le 15 janvier 2009], <http://www.rgpp.modernisation.gouv.fr/>

prévoit que toutes les politiques seront évaluées d'ici la fin du quinquennat. Fin 2013, une vingtaine des quarante-neuf évaluations initiées, étaient achevées (Fouquet, 2014).

Un cadrage commun est donné dans le guide élaboré par une équipe d'inspecteurs généraux (Battesti et al., 2012). Pour ces évaluations, les auteurs prônent le pragmatisme et mettent en garde les évaluateurs des pièges de l'exercice : confusion des responsabilités, excès d'expertise au détriment de l'association des parties prenantes, exercice trop expert et déconnecté de la décision, préconisations irréalistes, capture par des intérêts particuliers.

Depuis 2013, un Commissariat général à la stratégie et à la prospective (CGSP), qui remplace le Centre d'analyse stratégique, a la double vocation, d'être un lieu d'échanges et de concertation et d'apporter son concours au gouvernement pour la construction de l'avenir. De même, inspiré des expériences étrangères, un Laboratoire de l'innovation publique (LIP) a vu le jour ; il aura la charge d'évaluer les politiques publiques, de faire remonter des expériences, de pratiquer des expérimentations et de diffuser des modèles.

Le récit des vingt-cinq dernières années témoigne des vicissitudes de l'évaluation des politiques publiques en France. Née sous le signe de l'évaluation démocratique, elle est aujourd'hui entraînée par la vague néolibérale et marquée du sceau des réformes successives de l'Etat.

1.3. Les acteurs de l'évaluation des politiques publiques

Les acteurs de l'évaluation des politiques publiques appartiennent à quatre catégories : les acteurs institutionnels relevant des corps de contrôle et des départements ministériels, les acteurs des collectivités territoriales, les consultants des organismes privés, les chercheurs des organismes universitaires.

« Promoteurs » ou « bénéficiaires » au cours du processus d'institutionnalisation (Jacob et Varone, 2004), ces acteurs, aujourd'hui, n'interviennent pas sur les mêmes segments de l'évaluation (Matyjasik, 2013).

1.3.1. Les acteurs institutionnels

Une place privilégiée pour les corps de contrôle

La construction des structures d'évaluation a accordé une place prépondérante aux grands corps de fonctionnaires traditionnellement dévolus aux tâches d'inspection, de contrôle administratif et autres formes d'expertise d'Etat (Barbier, 2010 ; Barbier et Matyjazik, 2010 ; Duran et al., 1995). L'unité de lieu dans l'exercice de ces activités explique en grande partie la difficulté à établir la frontière entre l'évaluation et les « activités voisines » (Société française de l'évaluation, 2007) et à affranchir l'évaluation d'une vision trop normative.

L'évaluation est explicitement inscrite dans les missions des trois principaux corps de contrôle (Inspection générale des Finances, Inspection générale de l'administration, Inspection générale des affaires sociales). En particulier, l'Inspection générale des affaires sociales (IGAS) est responsable du contrôle et de l'évaluation de la mise en œuvre des politiques publiques de la sécurité sociale et de la prévoyance sociale, de la protection sanitaire et sociale, du travail, de l'emploi et de la formation professionnelle. Depuis plusieurs années, elle accomplit, seule ou avec d'autres institutions, des missions d'évaluation d'actions publiques précises, ou de dispositifs législatifs ou réglementaires.

La Cour des comptes, principalement vouée au contrôle budgétaire, est chargée depuis la réforme constitutionnelle de 2008, d'une mission d'évaluation des politiques publiques. Son président, affiche pour cette activité, un positionnement distinct des autres travaux de la Cour, qu'il s'agisse des revues de programmes ou des audits de performance. Conscient des défis à relever pour produire des travaux de qualité et utiles pour la prise de décision, il appelle à une collaboration avec les organismes universitaires pour recueillir l'expertise scientifique nécessaire (Migaud, 2013). En matière de santé, la Cour a réalisé en 2010 une étude sur l'efficacité des politiques de lutte contre le VIH/Sida dans la perspective d'un prochain plan de lutte contre les infections sexuellement transmissibles et le virus de l'immunodéficience humaine ou encore, l'évaluation de la politique de lutte contre le tabagisme en 2012.

Les organisations publiques

A côté des corps de contrôle, plusieurs départements ministériels (Culture, Equipement, Emploi, Education, Justice, notamment) se sont engagés précocement dans des activités d'évaluation et ont pendant longtemps constitué l'essentiel des travaux conduits.

Progressivement la fonction d'évaluation s'est structurée dans les organisations publiques, tant dans les services de l'Etat que dans les collectivités territoriales qui ont créé des unités d'évaluation chargées soit de conduire des évaluations internes soit de piloter des démarches d'évaluation externalisées.

Le Parlement : une instance récemment reconnue

Malgré la mise en place d'un Office parlementaire des politiques publiques en 1996, disparu en 2001, le Parlement est resté absent de la scène de l'évaluation des politiques. Progressivement l'évaluation va intégrer le travail parlementaire (Le Bouler, 2003) : création en 1999 de la Mission d'évaluation et de contrôle (MEC) au sein de la Commission des Finances de l'Assemblée nationale et développement d'activités de contrôle et d'évaluation de la commission des finances du Sénat. Depuis 2008, elle est inscrite dans la Constitution et devient une responsabilité majeure du Parlement en partage avec la Cour des Comptes. Depuis 2009, un comité d'évaluation et de contrôle (CEC) permet à l'Assemblée nationale de mettre en œuvre la fonction de contrôle et d'évaluation qui lui est désormais explicitement reconnue par la Constitution. Ce comité conduit des évaluations de politiques publiques et apporte son expertise sur les études d'impact qui accompagnent les projets de loi déposés par le Gouvernement. Pour Lacouette-Fougères et Lascoumes (2013), il reste encore très en deçà des attentes, les élus manifestant peu d'intérêt pour les débats engagés à partir de travaux d'évaluation.

1.3.2. Le rôle fédérateur de la société française de l'évaluation

A l'instar des autres pays, une société professionnelle a vu le jour en France en 1999. La Société française de l'évaluation (SFE) s'est donnée la mission de « contribuer au développement de l'évaluation et promouvoir son utilisation dans les utilisations publiques et

privées⁹ ». Lieu de rencontres, de débats et de partages d'expériences, elle ambitionne de faire progresser les méthodes et de favoriser le respect des règles éthiques et des procédures propres à garantir la qualité des évaluations et le bon usage des résultats.

Actuellement, seize groupes travaillent sur des thèmes variés : évaluation des politiques culturelles, efficacité du développement, développement durable, évaluation et utilité sociale, standards de qualité et déontologie, politiques de développement territorial, politiques sociale, évaluation en santé, compétences et déontologie des évaluateurs. Sept clubs régionaux ou interrégionaux (Aquitaine, Bretagne Pays de Loire, Ile de France, Midi Pyrénées, Nord Pas-de-Calais, PACA, Rhône Alpes) complètent le panorama.

A l'initiative du groupe « Standards et déontologie », une charte de l'évaluation a été élaborée. L'évaluation est définie dans le préambule de la charte :

« L'évaluation vise à produire des connaissances sur les actions publiques, notamment quant à leurs effets, dans le double but de permettre aux citoyens d'en apprécier la valeur et d'aider les décideurs à en améliorer la pertinence, l'efficacité, l'efficience, la cohérence et les impacts. L'évaluation contribue ainsi à rationaliser la prise de décision publique, à moderniser la gestion de l'administration et des services publics et à rendre plus efficace la dépense publique ».

Conçue dans une visée non prescriptive, elle énonce sept principes (pluralité, distanciation, compétence, respect des personnes, transparence, opportunité, responsabilité) pour structurer et promouvoir de bonnes pratiques d'évaluation (Société française de l'évaluation, 2006). Document initialement prévu pour figurer en annexe des documents contractuels entre commanditaires et évaluateurs, cette charte constitue une référence avec une valeur symbolique d'engagement pour chacun d'eux (Barbier, 2009).

1.3.3. Le faible investissement du monde universitaire

Le monde universitaire est, en France, peu investi dans l'évaluation, ni comme activité de recherche appliquée ni comme objet de recherche. Barbier et Matyjasik (2010) évoquent plusieurs raisons. La première renvoie à la position épistémologique des chercheurs justifiant une prise de distance avec les institutions afin de conserver le caractère d'indépendance de la

⁹ www.sfe-asso.fr

recherche et se prémunir contre toute forme d'instrumentalisation. L'ajustement à la commande d'évaluation et au calendrier du commanditaire, la production de recommandations s'accordent peu avec les modalités de travail des chercheurs. La seconde serait liée à la suspicion que cette activité inspire aux universitaires du fait de l'investissement des consultants privés et du caractère commercial de l'activité. Enfin, la nature même de l'évaluation, par essence multidisciplinaire, ne favorise pas l'engagement des chercheurs dont la reconnaissance relève d'une production scientifique dans leur propre champ disciplinaire. Faute de cette reconnaissance, peu de chercheurs ont engagé des recherches en (et sur) l'évaluation.

Des trois disciplines « naturelles » de l'évaluation, les sciences économiques sont aujourd'hui plus présentes, eu égard à l'intérêt porté aux méthodes expérimentales (L'Horty et al., 2012), aux études de rendement ou aux études multicritères d'aide à la décision (Baslé and Malgrange, 2007), les sciences politiques et la sociologie, ayant pris leur distance. Cependant des initiatives récentes, telles la création en 2012 du Laboratoire interdisciplinaire d'évaluation des politiques publiques à Sciences-Po, le développement de travaux à l'Ecole d'économie de Paris ou encore la dynamique du réseau des chercheurs en évaluation des politiques publiques (RCEPP) créé en 2008 au sein de la SFE, semblent vouloir infléchir cette tendance (Fouquet, 2014).

1.3.4. La place du secteur privé

L'activité d'évaluation pratiquée dans le secteur privé s'est accrue ces dernières années en lien avec le développement des activités de conseil et d'ingénierie dans le secteur public (Matyjasik, 2013). Faute de formation universitaire structurée, les pratiques des consultants se sont forgées, au gré de la multiplication des expériences évaluatives et de l'incorporation de fondamentaux méthodologiques propres à l'évaluation à l'arsenal des outils issus de leur champ professionnel d'origine. A mi-chemin entre conseil et évaluation, ils se revendiquent non comme évaluateurs mais comme « consultants en évaluation des politiques publiques », activité où la fonction de médiation est considérée prédominante.

1.4. Des pratiques hétérogènes mais un fond culturel commun

Les pratiques d'évaluation sont hétérogènes et peu normées, à l'image de la diversité des praticiens.

Les activités d'évaluation sont de nature variée, entre évaluation, expertise et conseil. Barbier (2010) suggère une classification entre : des activités de recherche, incluant la recherche scientifique appliquée et des études de commande ; des activités d'évaluation des politiques publiques conduites sous l'égide d'instances autonomes ou à destination d'un commanditaire sur une base contractuelle avec un expert/consultant. C'est le degré de proximité entre l'évaluateur et son commanditaire qui, pour Barbier, permettrait de distinguer ce qui relève du conseil ou de l'expertise.

Les pratiques varient au regard des approches, des modèles théoriques ou des postures éthiques. L'évaluation est marquée, à la fois par une approche normative, compte tenu de sa filiation avec les institutions de contrôle de l'Etat, et par la doctrine impulsée par le CSE au sein duquel se côtoyaient membres des corps de contrôle et chercheurs en sciences politiques ; ces derniers ayant nourri de leurs apports théoriques, la réflexion en construction (Barbier and Matyjasik, 2010). Les travaux du CSE constituent une référence qui imprègne la culture évaluative française (Lacouette et Lascoumes, 2013). L'importance accordée à la formulation des questions évaluatives et au pluralisme des méthodes, la définition des étapes du projet d'évaluation, l'identification de critères majeurs d'évaluation représentent des repères essentiels et partagés par la communauté des évaluateurs (Perret, 2008). Parallèlement, l'investissement des consultants a introduit les outils du management et de la démarche de projet.

Même s'il existe de plus en plus de formations universitaires intégrant l'enseignement de l'évaluation, cet enseignement n'est pas uniforme. L'évaluation ne constitue pas en France un métier et le consensus sur l'accréditation d'évaluateurs est loin d'être obtenu.

Patchwork de pratiques contrastées selon l'appartenance institutionnelle et professionnelle des acteurs, formés au contrôle et à l'audit pour les uns, à la démarche de projet et au conseil pour les autres ou encore au maniement des statistiques, l'activité d'évaluation redevient l'objet de préoccupations méthodologiques. Le RCEPP poursuit l'objectif de développer la recherche interdisciplinaire en évaluation des politiques publiques. Il a récemment organisé un cycle de séminaires de recherche relançant notamment une réflexion sur les méthodes.

Le champ de l'évaluation est traversé de controverses et d'oppositions encore vives entre les tenants de l'évaluation managériale versus ceux de l'évaluation démocratique, les défenseurs d'approches participatives versus leurs détracteurs privilégiant les approches expertes et indépendantes, les partisans des méthodes qualitatives versus ceux des méthodes quantitatives. Pour autant, en dépit des clivages qui inclineraient à choisir des formes d'évaluation radicalement opposées, interne/externe, scientifique/profane, s'ouvre un espace d'échanges et de débats et des rapprochements sont perceptibles.

1.5. Peut-on toujours parler d'une « évaluation à la française ? »

Entre évaluation managériale et démocratique, l'évaluation en France a, en réalité, pris une troisième voie, l'évaluation pluraliste, compromis entre « nécessités techniques et exigences politiques » (Duran et al., 1995). Pour Monnier (1987), l'évaluation pluraliste est centrée sur les informations à produire pour améliorer la compréhension des effets d'un programme par le système social, sa première tâche étant l'identification des utilisateurs et la mise à jour des systèmes de valeur qui orientent l'évaluation. Fondé sur le devoir de transparence et la facilitation du débat avec les citoyens, le modèle français choisissait d'accorder une place importante à l'organisation de la confrontation des points de vue, en instaurant des espaces d'échanges au sein de collectifs pluralistes, les « instances d'évaluation », en charge du pilotage des évaluations. Dans ce modèle, l'évaluation n'est pas un événement ponctuel mais une activité intégrée au processus de construction de la politique, le rôle de l'évaluateur étant de créer les conditions de cette analyse collective. L'implication des universitaires dans cette fonction était un signe distinctif de « l'évaluation à la française ».

Si l'évaluation aujourd'hui conserve des éléments résiduels de ce modèle, elle ne bénéficie plus autant de la contribution du mode universitaire ; les chercheurs des sciences politiques et de la sociologie ayant, pour la plupart, délaissé le champ de l'évaluation. Pour autant, elle porte les marques de son histoire singulière, endossant, au gré des situations, des attentes plurielles en référence aux différents courants qui ont structuré le champ. Légitimer l'action publique, rendre compte, améliorer l'efficacité des politiques, faciliter la démocratie, mesurer la performance sont tous des mots d'ordre inscrits au répertoire de l'évaluation.

2. L'évaluation dans le domaine de la santé

Progressivement, et de façon accrue ces dernières années, l'exigence d'évaluation s'intensifie dans le secteur de la santé, comme en témoigne l'introduction de clauses évaluatives dans les textes législatifs et réglementaires¹⁰. Par ailleurs, les objets auxquels elle s'applique sont de plus en plus difficiles à appréhender. Les politiques de santé sont réputées complexes parce qu'elles s'attaquent à des problèmes ardues, nécessitant des actions conjointes, déployées à différents échelons territoriaux, parce qu'elles produisent des effets dans le long terme et intriquées aux effets d'autres politiques. Les interventions, objets à évaluer, sont parfois indéterminées, faites de l'imbrication de divers instruments de politiques. Enfin, elles se déploient dans un contexte frappé de réformes successives génératrices d'instabilité et de davantage de complexité.

L'exposé de la chronologie et de l'état des lieux des pratiques dans ce domaine s'appuie sur la documentation disponible et sur des travaux personnels.

Après avoir rappelé quelques événements à l'origine de l'émergence de l'évaluation dans le secteur de la santé (2.1.), nous aborderons les points saillants de la pratique de l'évaluation dans les institutions de la santé (2.2.) avant de conclure sur les principaux enjeux de l'évaluation dans ce domaine (2.3.).

2.1. L'émergence progressive de l'évaluation dans le secteur de la santé

C'est avec le développement des activités de planification et la création d'agences que l'évaluation acquiert une visibilité dans les trois domaines de la santé (soin, santé publique, médicosocial), chacun ayant suivi sa propre trajectoire. Pour dresser un tableau global de l'évaluation, nous évoquerons quelques éléments de son histoire dans les domaines du soin et du médicosocial. Néanmoins, nous recentrerons le propos au champ de la santé publique.

C'est dans le cadre des études de la RCB que l'évaluation fait son entrée avec la mise en place des premiers programmes de santé (Jabot and Demeulemeester, 2005b). L'expérience la plus remarquable concerne le programme périnatalité pour lequel une étude comparative systématique du rapport coût-efficacité avait été effectuée en vue de choisir les actions plus

¹⁰ Loi 2002-2 du 2 janvier 2002, loi 2004-806 du 9 août 2004 relative à la politique de santé publique et loi 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires ; pour ne citer que les principales.

coût-efficaces. Par la suite, l'évaluation s'est poursuivie dans plusieurs directions (évaluation des interventions et dispositifs de santé, pratique de soins, pratiques professionnelles) et, depuis les années 90, l'évaluation médico-économique (Tabuteau, 2013), avec des bifurcations entre ce qui relève du registre de l'évaluation de programme, des autres formes d'évaluation appliquées aux établissements de soin (accréditation, certification, démarche qualité) ; la démarche qualité étant une activité commune à plusieurs domaines, récemment développée dans le secteur de la promotion de la santé.

A partir des années 2000, le déploiement de la LOLF a introduit la pratique du *reporting* aux niveaux national et régional, une tendance confirmée par la loi de 2004 qui prévoit le suivi des objectifs de la politique de santé et de ses programmes, et le développement de la contractualisation de l'Etat avec ses établissements publics (Cases et al. 2009).

2.1.1. Le développement de la programmation en santé

En effet, depuis plus de dix ans, les travaux d'évaluation se sont multipliés à différents échelons (national, régional, local) en réponse aux obligations réglementaires ou à l'initiative des pouvoirs publics. Les objets sur lesquels ils portent, sont multiformes et d'un format plus ou moins bien adapté à l'exercice. Si les programmes et réseaux de santé ont une architecture compatible avec l'évaluation, les interventions de type plan ou politique de santé, dispositifs et *a fortiori* expérimentations, sont plus difficiles à appréhender, faute le plus souvent d'objectifs structurants. Ces travaux sont d'envergure nationale (lutte contre le sida, addictions, accès aux soins pour les plus démunis, santé environnementale, cancer), régionale (politiques de prévention, schémas d'organisation des soins, programmes spécifiques, réseaux) ou locale (projets ou programmes territoriaux, contrat local de santé).

2.1.1.1. Le rôle d'entraînement de l'échelon régional

La programmation en santé est une histoire récente en France. C'est à partir du rapport du Haut Comité de santé publique publié en 1994, que la démarche de programmation a été progressivement adoptée. Avec la parution des ordonnances de 1996¹¹ et du décret¹² qui lui a

¹¹ Ordonnance n° 96-345 du 24 avril 1996 relative à la maîtrise médicalisée des dépenses de soins.

¹² Décret n°97-360 du 17 avril 1997 relatif aux conférences régionales de santé.

fait suite, un cadre légal était donné au niveau régional pour esquisser une politique de santé. Ces dispositions réglementaires instaurent la mise en place de conférences régionales de santé, forums d'expression des partenaires en santé, au sein desquels étaient discutés les enjeux de santé afin de retenir les priorités sur lesquelles engager des programmes. Ainsi, le mouvement de généralisation des programmes s'est intensifié à la fin des années 90, amplifié par la loi de lutte contre les exclusions en 1998, qui généralisait la mise en place de programmes d'accès à la prévention et aux soins (PRAPS) spécifiquement dédiés aux personnes en situation de précarité. Progressivement toutes les régions ont construit des programmes. Fin 2003, on en dénombrait plus d'une centaine à l'échelle du territoire national. Au fil des années, des démarches d'évaluation ont été engagées, incitées pour bon nombre d'entre eux, par la clause évaluative des contrats de plan Etat région au titre desquels les conseils régionaux apportaient une contribution. Si quelques régions pratiquent l'évaluation de leurs programmes, d'autres considèrent qu'un état des lieux des réalisations est suffisant pour rendre compte de l'utilisation des fonds publics. La distinction entre activités de suivi et d'évaluation est loin d'être claire pour toutes. En 2003, une étude relevait que seulement un quart des programmes de santé avaient fait l'objet d'une évaluation (Jabot, 2004).

Le développement des appels à projet, mode privilégié de mise en œuvre des programmes, a relayé la démarche auprès des opérateurs locaux, conditionnant le renouvellement de leurs projets à la remise d'une évaluation, dont le format évoque plus un rapport d'activité qu'un véritable travail d'analyse. Cependant, et plus particulièrement dans certaines régions, des référentiels, des outils pédagogiques, des formations ont été construits et mis à la disposition des acteurs pour accompagner la démarche souhaitée et qualifier au mieux les pratiques.

La loi de santé publique de 2004, qui a instauré les PRSP, a introduit une clause évaluative pour ces plans. Cependant, un rapport du Haut Conseil de la santé publique (2008) relevait plusieurs obstacles à la mise en œuvre de l'évaluation des politiques régionales de santé, les uns d'ordre méthodologique, les autres d'ordre financier ou opérationnel, en raison du manque de clarté lié à la superposition des compétences. Ce rapport soulignait également l'hétérogénéité des méthodes rendant peu probable la construction d'une évaluation nationale.

A côté des programmes issus de la dynamique des conférences régionales de santé, d'autres interventions se sont développées, avec, pour certaines d'entre elles, un cahier des charges précis, tant pour leur construction que pour leur évaluation. C'est le cas des réseaux

de santé ou des programmes de dépistages du cancer (Haute Autorité de santé, 2004 ; Institut national de veille sanitaire, 2011).

Si l'essor de la programmation, associée aux obligations réglementaires propres à certains dispositifs, a constitué un levier pour faire progresser l'évaluation en santé publique, on ne relève pas les mêmes avancées dans le domaine de la planification de l'offre de soins. Les schémas régionaux d'organisation des soins (SROS) de première génération n'ont pas été évalués et l'évaluation des SROS de la génération suivante, pourtant prescrite par ordonnance¹³, était jugée décevante (Couty, 2006). Une enquête menée en 2006 (ENSP, 2006) montrait que seulement trois des vingt-quatre SROS étudiés avaient été évalués et rapportait que les agences régionales de l'hospitalisation chargées de cette évaluation n'avaient pas une vision claire de l'évaluation d'un schéma, souvent confondue avec l'évaluation des besoins, des activités ou des pratiques.

2.1.1.2. L'évaluation au niveau national

A l'instar du niveau régional, la multiplication des plans nationaux dans les années 2000, a eu un effet d'entraînement sur l'évaluation. Succédant au bureau de « l'évaluation et de l'économie de la santé », était créé en 2000 un « bureau de l'évaluation des programmes, recherche et prospective » au sein de la direction générale de la santé (DGS). La partie évaluation relevait d'une seule personne laquelle avait en charge plusieurs missions : évaluer l'impact des actions et programmes réalisés, passer des commandes et piloter des évaluations ; accompagner les bureaux thématiques désireux d'évaluer leurs propres programmes ; construire des repères pour les acteurs régionaux chargés de la mise en œuvre des évaluations et notamment, articuler évaluations nationales et régionales. Ainsi, plusieurs évaluations de programmes nationaux ont été réalisées : évaluation nationale des PRASP de première génération (PRASP 1) en 2003, évaluation *ex ante* des PRASP 2, évaluation de l'action des médiateurs santé dans le cadre de la lutte contre le sida en 2006, évaluation du programme Périnatalité en 2009, pour n'en citer que quelques-uns. Parallèlement, les bureaux « thématiques » de la DGS ou les autres directions avaient la possibilité d'engager des évaluations dans leur propre domaine.

¹³ Ordonnance 2003-850 du 4 septembre 2003.

En 2004, la loi confiait l'évaluation de la politique de santé et des plans nationaux au Haut conseil de santé publique (HCSP), instance d'expertise renouvelée à cette occasion. Quelques années plus tard, du fait de la recomposition des services du Ministère chargé de la santé consécutive à la loi de 2009, l'activité d'évaluation a perdu de sa visibilité.

Enfin, la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), service statistique du Ministère chargé de la santé, participe à la conception et à la mise en œuvre des méthodes d'évaluation des politiques sociales et développe un important programme de publications de ses travaux.

2.1.2. La mise en place d'agences spécialisées

A partir des années 1980, plusieurs agences ont été créées, principalement en réponse aux crises sanitaires, afin d'une part, de renforcer l'expertise et la gestion des risques sanitaires, et d'autre part, de séparer les niveaux d'expertise et de décision (Benamouzig et Besançon, 2007). La naissance des agences chargées de missions d'évaluation s'inscrit dans ce mouvement.

2.1.2.1. Dans le domaine des soins

Créée en 1990, l'*Agence nationale pour le développement de l'évaluation médicale* (ANDEM) est focalisée sur l'évaluation des techniques et stratégies médicales. Elle s'appuie sur les sociétés scientifiques et les groupes professionnels avec qui elle a développé un ensemble de méthodes et de recommandations pour la pratique clinique ainsi que des démarches d'évaluation de la qualité des soins dans les établissements et en médecine libérale.

En 1996 l'*Agence nationale d'accréditation et d'évaluation en santé* (ANAES) reprenait, dans une version élargie, les missions de l'ANDEM pour la partie évaluation. En matière d'accréditation, elle a construit et validé des méthodes et des référentiels pour l'accréditation des établissements.

Créée par la loi n°2004-810 du 13 août 2004, la *Haute autorité de santé* (HAS) succédait à l'ANAES. Instance d'expertise scientifique, consultative et indépendante, sa création relevait d'un triple objectif : améliorer la qualité des soins, renforcer les liens avec tous les acteurs du

système de santé et garantir la transparence de l'information médicale. Elle évalue l'utilité médicale des médicaments, les dispositifs médicaux et actes professionnels, coordonne les dispositifs d'évaluation et d'amélioration de la qualité des pratiques et des soins.

Bien que l'évaluation d'actions ou programmes de prévention figure dans ses missions, cette évaluation est centrée sur « l'opportunité de mettre en place des actions ou de modifier des programmes existants et d'en préciser les modalités afin d'émettre une proposition (la recommandation en santé publique) destinée en premier lieu aux décideurs publics (...) les recommandations en santé publique privilégient l'approche populationnelle des problèmes de santé appréciant le rapport bénéfices/risques des différentes interventions possibles à l'échelle de la population et intégrant de façon habituelle une évaluation économique et organisationnelle¹⁴ ». La HAS intervient plutôt en amont de la mise en place d'interventions, s'intéressant plus spécifiquement à leur pertinence au regard des données scientifiques ou encore au rapport bénéfices/risques ; elle peut également analyser des modes d'organisation du système de santé (Scemama and Rumeau-Pichon, 2009). Sa méthode de travail qui s'appuie sur l'analyse de la littérature et les avis d'experts, se différencie de la démarche d'évaluation de politiques publiques.

2.1.2.2. Dans le domaine de la santé publique

Le domaine de la santé publique est le terrain d'intervention de plusieurs organismes en matière de l'évaluation, certains ayant des missions formelles et exclusives, d'autres exerçant l'évaluation de façon plus marginale.

Le Haut conseil de la santé publique

Instance d'appui à la conception et à l'évaluation des politiques et stratégies de prévention et de sécurité sanitaire, le Haut conseil de santé publique (HCSP) est chargé depuis la loi du 9 août 2004, d'évaluer la politique de santé. A ce titre, il dispose d'une commission « évaluation, stratégie et prospective » qui réunit l'expertise nécessaire pour la réalisation des travaux¹⁵. Elle

¹⁴ http://www.has-sante.fr/portail/jcms/fc_1250012/fr/sante-publique-et-organisation-des-offres-de-soin, consulté le 22 juillet 2014

¹⁵ <http://www.hcsp.fr/Explore.cgi/groupe?clef=62> .

comprend vingt membres dont les directeurs généraux ou représentants des cinq agences¹⁶ les plus concernées. Depuis sa création, le HCSP a réalisé l'évaluation de l'atteinte des objectifs nationaux de santé publique ainsi que l'évaluation de plusieurs plans et programmes nationaux. Par ailleurs, il apporte un soutien méthodologique à l'élaboration de l'évaluation de plans en cours. Bien que ses missions soient distinctes de celles de la HAS, des « zones de recouvrement » existent sur des thèmes, tels que le dépistage ou la vaccination, mais le travail du HCSP a une dimension plus politique (Perret, 2009b).

Après quelques années de fonctionnement au cours desquelles une méthode de travail a progressivement été construite, la commission spécialisée a entamé une réflexion sur la démarche et les méthodes utilisées. En 2011, devant le nombre important de plans ou de programmes nationaux de santé susceptibles d'être évalués, le Haut Conseil de santé publique a souhaité se doter d'un outil d'aide à la décision pour la sélection des travaux d'évaluations. A sa demande, une étude a été réalisée avec l'appui d'un groupe d'experts (Deumelemeester and El Khoury, 2011) aboutissant à la mise à disposition d'une grille d'appréciation de l'évaluabilité des interventions. L'utilisation de cet outil permet de faciliter les échanges avec les promoteurs de l'évaluation pour formuler les questions évaluatives et identifier les méthodes appropriées pour y répondre mais elle peut aussi trouver des applications auprès des concepteurs des plans et programmes, en tant que levier d'amélioration de la qualité des programmes. En effet, ce travail montrait que les plans/programmes étaient rarement étayés par une logique d'intervention et que leur structuration n'était pas de nature à permettre leur évaluation. Par ailleurs, quelques principes méthodologiques et déontologiques, inspirés de la charte de la Société française de l'évaluation, étaient soumis à la réflexion de la commission du HCSP, notamment, l'association des parties prenantes à l'évaluation comme moyen d'appropriation des résultats et de contribution à une meilleure utilisation de l'évaluation. Depuis, deux séminaires de réflexion sur les méthodes ont été tenus au sein de l'audience élargie du HCSP.

L'Institut national de prévention et d'éducation pour la santé

L'Institut national de prévention et d'éducation pour la santé (INPES) créé également par la loi d'août 2004 n'a pas une mission explicite d'évaluation mais développe quelques actions

¹⁶ Institut de veille sanitaire, Agence nationale chargée de la sécurité sanitaire de l'alimentation, de l'environnement et du travail, Institut national de prévention et d'éducation pour la santé, Institut national de lutte contre le cancer, Haute Autorité de santé.

dans ce domaine (Guilbert et al. 2009). Cependant, promouvoir la recherche de données probantes et l'élaboration de stratégies adossées aux connaissances scientifiques fait partie de ses objectifs. Ainsi, la direction des Affaires scientifiques¹⁷ réalise, au sein de son département « Evaluation et expérimentation », des évaluations scientifiques d'efficacité sur les actions qu'il porte, expérimente et évalue de nouvelles formes d'actions potentiellement efficaces, développe une expertise sur les méthodes d'évaluation d'interventions. Ce département encourage le développement de la démarche « évaluation action », qui intègre l'évaluation à l'action, démarche utilisée à l'INPES, pour évaluer les diverses actions réalisées en son sein.

L'INPES contribue à la qualification des acteurs de la promotion de la santé avec la diffusion de matériaux pédagogiques : guide d'autoévaluation de la qualité, guide adapté du guide suisse Outil de catégorisation des résultats (Inpes, 2010). La promotion de ces outils a été accompagnée de formations relayées par les pôles de compétences régionaux en éducation pour la santé auprès des professionnels de terrain.

L'Inspection générale des affaires sociales (IGAS)

Les organismes de contrôle, Cour des comptes et IGAS, sont mandatés régulièrement seuls ou conjointement à d'autres organismes pour réaliser des évaluations.

L'IGAS est régulièrement saisie pour des évaluations de plans (Plan National Nutrition Santé, évaluation des mesures du Plan Cancer 2003-2007 relatives au dépistage et à l'organisation des soins), de politiques (soutien à la parentalité, prise en charge du diabète), de dispositifs (dispositif médicosocial de prise en charge des conduites addictives, financements des médicaments). Elle réalise parfois ces évaluations en complément des travaux conduits par le HCSP (évaluation du Plan Cancer, PNSE2), son périmètre étant généralement délimité par les questions relatives à la mise en œuvre et à l'utilisation des fonds. Certains travaux (Daniel et al., 2006) procèdent d'une forme hybride, traduisant « l'évaluation comme prolongement du contrôle » (Perret, 2009b).

Il n'est pas rare que plusieurs expertises ou évaluations soient demandées sur le même objet. L'exemple du Plan Cancer 2003-2007 est illustratif de la dispersion des commandes : le HCSP était saisi pour l'évaluation de l'impact du Plan ; l'IGAS évaluait les mesures du Plan relatives au dépistage et aux soins ; la Cour des comptes étudiait le degré de réalisation des mesures. Bien que les segments d'évaluation attribués aux différents organismes soient

¹⁷ <http://www.inpes.sante.fr/INPES/organisation.asp>, consulté le 22 juillet 2014.

adaptés à leur périmètre de compétences et permettent de distinguer la spécificité des apports de chacun, la coordination est parfois insuffisante. A l'instar de ce qui a été observé dans les autres secteurs de politiques publiques (Jacob et Varone, 2004), l'existence de dispositifs concurrents entre plusieurs types d'acteurs souhaitant demeurer présents sur la scène de l'évaluation démontre à quel point l'évaluation est perçue comme un enjeu d'influence (Perret, 2009b).

2.1.2.3. Dans le champ médicosocial

L'Agence nationale de l'évaluation de la qualité des établissements et services sociaux et médico-sociaux (ANESM) créée en 2007 a pour mission principale d'accompagner les établissements et services sociaux et médico-sociaux (ESSMS) dans la mise en œuvre de l'évaluation interne et externe, instituée par la loi du 2 janvier 2002¹⁸ qu'ils sont tenus d'effectuer pour le renouvellement de leur activité. Pour cela, l'Agence produit des procédures, références et recommandations sur les pratiques professionnelles et sur l'évaluation, de même qu'elle habilite les organismes chargés de l'évaluation externe.

L'évaluation dans ce domaine est une forme assez inédite dans la mesure où elle combine une évaluation interne, réalisée par les établissements eux-mêmes, proche de la démarche qualité et une évaluation externe ouvrant la voie à une forme d'accréditation.

Les autres agences sanitaires (Institut national de veille sanitaire, Institut national du Cancer par exemple), du fait de leur expertise et de leur capacité de synthèse de connaissances, contribuent à l'évaluation en fournissant des données.

La création d'agences spécialement missionnées pour l'évaluation a contribué à une plus grande visibilité et reconnaissance de l'exercice, favorisé la structuration de l'évaluation, tout particulièrement dans les domaines du soin et du médicosocial, la légitimité de l'Etat s'exprimant dans la définition des référentiels d'activité (Tabuteau, 2013).

¹⁸ <http://www.anesm.sante.gouv.fr/> .

2.2. La pratique de l'évaluation dans les institutions régionales de santé

La pratique d'évaluation dans les institutions de la santé est assez peu documentée. L'analyse qui suit, s'appuie sur des travaux personnels conduits à partir de plusieurs sources : enquête réalisée en 2005 à la demande de la mission d'animation des fonctions d'inspection en services déconcentrés de l'IGAS suivie d'une journée d'échanges avec les professionnels de ces services ; états des lieux réalisés par les élèves inspecteurs de l'action sanitaire et sociale (IASS) au cours de leur stage à la demande de l'ENSP/EHESP (années 2008 à 2012) ; animation des sessions de formation continue auprès des personnels œuvrant dans les services régionaux (années 2002 à 2014) ; formation et/ou accompagnement à la mise en œuvre d'évaluation dans cinq régions.

La pratique d'évaluation tient à plusieurs éléments : la place et la reconnaissance de l'évaluation dans l'institution, la densité et la fréquence des commandes ainsi que l'existence de référentiels explicites et partagés.

La loi HPST trace une ligne de partage entre deux époques, la création des ARS modifiant l'organisation interne et introduisant des repères culturels nouveaux du fait de la reconfiguration des services et du brassage de cultures institutionnelles.

2.2.1. L'évaluation dans les services déconcentrés

Un positionnement ambigu au sein des MRICCE

En 1999, étaient mises en place les Missions régionales d'inspection, de contrôle et d'évaluation (MRIICE), services destinés à développer chacune de ces trois fonctions avec l'appui d'une structure créée dans le même temps au sein de l'IGAS : la mission permanente d'animation des fonctions d'inspection (MAFI).

Après plusieurs années de fonctionnement, on constate qu'un travail substantiel a été élaboré sur le registre de l'inspection, avec une production conséquente d'outils méthodologiques (guides, grilles, protocoles) mais le chantier de l'évaluation n'a pas réellement été entamé (Israelian, 2008). En effet, les MRIICE ont concentré l'essentiel de leur activité sur les activités d'inspection et n'ont que très exceptionnellement investi le champ de l'évaluation, ces services ne disposant ni des moyens suffisants (de une à cinq personnes au

maximum) ni des compétences pour s'engager dans l'évaluation (Israelian, 2008). L'analyse des demandes qui leur étaient adressées témoigne d'une part, de la réalité d'un besoin émergent de la part de l'institution et, d'autre part, de la confusion entre l'évaluation et les activités de contrôle et d'audit (Jabot, 2005a). L'ambiguïté de ces services et le double positionnement des personnels, entre inspection et évaluation, participe de la défiance des opérateurs de l'Etat envers l'évaluation.

En dépit d'un besoin exprimé, ces services n'ont pas réussi à développer une véritable activité d'évaluation, faute de ressources, de formation et d'accompagnement.

La pratique existante, mal identifiée et peu soutenue dans les autres services

Si dans certaines régions, les MRIICE étaient les principaux destinataires de commandes d'évaluation, dans d'autres, les travaux étaient confiés aux services responsables de l'intervention concernée par l'évaluation¹⁹.

L'analyse des commandes montre la diversité des objets d'évaluation : réseaux (obésité infantile, troubles de l'apprentissage, addictologie), structures (Maison des adolescents), programmes (santé scolaire, PRAPS, toxicomanie), activités (coordination du centre de ressources autisme, soins infirmiers à domicile, activité IVG). Les contributions attendues de la part des professionnels étaient réparties entre maîtrise d'œuvre et appui méthodologique à une évaluation, donnant lieu à une grande variété des tâches : élaboration d'outils, remplissage de fiches, grilles, mise en place de comités de pilotage, coordination des acteurs.

Pour les programmes issus des conférences régionales, l'évaluation était habituellement confiée à des cabinets de consultants dont l'offre s'est considérablement accrue avec les demandes d'évaluation des PRSP, ou parfois aux observatoires régionaux de santé, certains ayant investi le domaine de l'évaluation.

Dépendant de la place accordée à l'évaluation, des compétences disponibles, de la motivation des acteurs à s'y impliquer, on constate que l'activité évaluative était très hétérogène d'une région à l'autre, quasi inexistante ou réelle, administrative ou partenariale, externalisée ou interne à l'institution.

¹⁹ Etude personnelle des rapports de stages des IASS de 2008 à 2012.

La pluralité des rôles à endosser - commanditaire, exécutant, contributeur - a progressivement estompé la frontière entre évaluation interne et évaluation externe, nécessitant la clarification des attendus de ces différents exercices et l'acquisition de compétences propres à chacun de ces rôles.

2.2.2. L'évaluation dans les ARS

La création de nouvelles structures répond à la volonté de décloisonnement des différents secteurs de la santé et vise une plus grande efficacité et efficience du système de santé. Reconfigurer les missions, traduites en « métiers », réexaminer l'organisation interne, repenser la gouvernance ont constitué des étapes préparatoires du chantier des ARS auquel de nombreux groupes de travail se sont attelés.

Une réflexion inexistante lors de la préparation de la loi HPST

Constatant l'absence de réflexion sur l'évaluation, un groupe de travail d'une vingtaine de personnes, les trois quart d'entre elles provenant de structures régionales²⁰, les autres relevant du niveau national, s'est constitué pour enclencher une réflexion sur la place de l'évaluation dans les futures ARS. Plusieurs réunions ont permis de dresser un état des lieux des pratiques et de formuler des scénarios pour l'organisation de la fonction évaluation dans les futures agences, dans une note adressée à l'attention du responsable de la préparation de la loi (Carbonnel et al., 2009). Par la suite, les agences ont construit leur organisation sans consignes particulières s'agissant des activités d'évaluation.

La place de l'évaluation dans les organigrammes des ARS

Notre analyse de la première version des organigrammes des ARS relève que l'évaluation était présente et visible dans presque la moitié des ARS (11/26) bien qu'elle y occupât une place inégale. L'évaluation figurait : au plus haut niveau hiérarchique (sous l'autorité directe directeur) dans quatre régions, deux niveaux sous le directeur dans six régions, trois niveaux sous le directeur dans une région. Dans sept cas sur onze, l'évaluation restait associée aux activités d'inspection/contrôle.

²⁰ Principalement des MRIICE.

Depuis leur création, plusieurs ARS ont revu leur organisation. Avec le déploiement du PRS, les programmes territoriaux et la signature de contrats locaux de santé, la place de l'évaluation a été rehaussée et figure majoritairement aujourd'hui dans les directions en charge de la stratégie.

2.2.3. Des pratiques hétérogènes

Le discours des professionnels et l'analyse des travaux d'évaluation, qu'ils soient produits par les institutions de la santé ou des prestataires externes, amènent plusieurs constats.

La vision normative de l'évaluation imprègne la culture d'évaluation et tend à faire de l'analyse de la conformité de la mise en œuvre du programme et de la vérification de l'atteinte des objectifs, le paradigme dominant. Pourtant, la référence aux critères classiques d'évaluation (pertinence, cohérence, efficacité, efficience, impact) est constante dans la phase de réflexion pré-évaluative.

Les méthodes d'évaluation sont rarement fondées sur des modèles théoriques. En particulier, la mise à jour de la théorie de l'intervention est assez peu recherchée, reflet des pratiques de programmation qui relèvent d'une démarche principalement implicite. En revanche, les acteurs du champ de la promotion de la santé sont plus perméables à ce type de démarche, la formation à des outils tels que l'outil de catégorisation des résultats les y aidant. En outre, les activités d'observation, de planification et d'évaluation conduites isolément et souvent par des acteurs et des services distincts, devraient désormais être mieux articulées en raison d'une part, de la structuration des activités de suivi liée à l'application des principes du *New Public Management* dans l'administration de la santé (Cases et al., 2009) et, d'autre part, de la création dans les ARS, de services regroupant ces trois activités.

Le Bouler (2009), ancien responsable de l'évaluation des politiques publiques au Commissariat au Plan, chef de la Mission recherche de la DREES et membre du HCSP, fait un constat similaire pour le niveau national, déplorant d'une part, l'absence d'explicitation de la chaîne de causalité et, d'autre part, un manque d'anticipation dans le recueil de données nécessaires à l'évaluation.

Influencé par le paradigme épidémiologique, les politiques et programmes sont aisément conçus comme des protocoles de traitement à appliquer sur des populations et pour lesquelles l'évaluation basée sur le contrefactuel représente un idéal à atteindre y compris pour des

interventions d'envergure. Pour autant, la prééminence du discours sur la démocratie sanitaire et le foisonnement de textes en la matière introduisent une préoccupation constante sur la prise en compte des usagers/citoyens, incitant à recourir aux approches qualitatives et enquêtes de satisfaction. En lien avec cette préoccupation, le modèle d'évaluation pluraliste est volontiers adopté, avec des variations significatives dans la gouvernance du dispositif d'évaluation et la participation accordée aux parties prenantes (Jabot et al., 2010).

L'évaluation peine à s'affranchir de la tradition de contrôle, activité pour laquelle l'expertise est présente et largement partagée, et ce d'autant, que dans certains domaines (évaluation des réseaux, évaluation dans le champ médicosocial), les frontières tendent à s'estomper entre audit, évaluation voire accréditation.

2.3. L'évaluation, un concept hybride en tension entre plusieurs visions

L'évaluation se construit progressivement comme une démarche autonome distincte d'autres activités avec lesquelles elle a été trop souvent confondue. La culture évaluative irrigue très inégalement les milieux institutionnels et professionnels et la diversité des pratiques reflète l'hétérogénéité des conceptions, des approches et des méthodes utilisées.

Reflet de la culture sociopolitique et historique française, la conception de l'évaluation dans le domaine de la santé demeure métissée de représentations contrastées, empreintes de la tradition de contrôle (vérification de conformité), de l'inspiration démocratique (espace de débat), de la rationalité budgétaire (performance) et, plus spécifiquement de son héritage endogène, le courant de *l'evidence-based medicine* (recherche de données probantes).

Dans un secteur marqué par une double influence, la culture du contrôle et le paradigme épidémiologique, l'évaluation en santé est en tension entre la tradition positiviste, qui conduit à des évaluations plus volontiers expertes avec recours aux approches quantitatives et expérimentales, et la tradition constructiviste qui privilégie la prise en compte des acteurs dans l'appréhension de la réalité. Pourtant, au-delà de ce clivage, les pratiques d'évaluation s'observent dans un continuum entre des formes *a priori* opposées, mais semblent plus évolutives que figées, à en juger par la prise de conscience de la variété des méthodes et approches, de l'appropriation de l'évaluation comme démarche sociale et processus d'apprentissage et de l'intégration des standards d'évaluation dans les pratiques.

Si la capacité d'évaluation est encore déficiente dans les institutions, une culture d'évaluation est en construction. Analysant quelques textes fondamentaux rédigés par la communauté internationale des évaluateurs, De Peuter et Pattyn (2008) se sont interrogés sur les relations entre ces deux concepts : la capacité d'évaluation est-elle une condition préalable au développement de la culture d'évaluation ou inversement, serait-il nécessaire de comprendre la culture d'évaluation pour renforcer les capacités ? Selon leur étude, la demande d'évaluation ainsi que son utilisation sont dépendantes de la culture d'évaluation et par extension du contexte, l'appropriation des résultats étant construite au cours du processus d'évaluation. Par contre, la phase de collecte et d'analyse des données, plus technique, est directement liée à la capacité d'évaluation. Toutefois, les deux concepts entretiennent des relations de dépendance ; la capacité d'évaluation est un facilitateur et un indicateur de la culture d'évaluation mais des éléments de la culture d'évaluation doivent être présents au stade très précoce de l'évaluation pour la construction des capacités. Enfin, les auteurs soulignent que l'obligation d'évaluation, l'ancrage institutionnel, le couplage évaluation/politique et l'organisation en réseau sont de nature à renforcer tant la culture que la capacité d'évaluation et des leviers pour l'institutionnalisation de l'évaluation.

Si la culture d'évaluation est, comme le définit Trosa (2009), « une façon de faire de l'évaluation non une discipline de professionnels « à part », mais une préoccupation constante des fonctionnaires », la culture d'évaluation dans le secteur de la santé est encore embryonnaire. Le décalage entre un discours exprimant des attentes élevées envers l'évaluation et la faible inscription de l'évaluation dans des mécanismes institutionnels, organisationnels, décisionnels est-il à voir du côté de la culture ou de la capacité ?

Chapitre 3 – La problématique

Le « paradoxe de l'évaluation » est un objet de questionnement commun aux chercheurs et praticiens de l'évaluation mais abordé selon les perspectives un peu différentes.

Pour Holjundt (2014), la non-utilisation de l'évaluation questionne la nature même de l'évaluation. Si l'utilisation est la raison d'être de l'évaluation, les évaluations non utilisées sont-elles des évaluations ?

Selon Jean Leca (1993), alors président du conseil scientifique de l'évaluation en France, c'est, du point de vue du décideur, la faisabilité de l'évaluation qui est en cause : « Le paradoxe de l'évaluation pour le décideur : quand elle est possible pour le décideur, elle est inutile ; quand elle est utile pour le décideur, elle est impossible à mener ».

Les praticiens de l'évaluation, quant à eux, portent leur attention sur la relation du décideur à l'évaluation : qu'est-ce qui explique l'abandon ou le désintérêt des commanditaires face à une évaluation qu'ils ont eux-mêmes sollicitée ?

1. Objectifs et questions de recherche

L'objectif principal de cette recherche est de décrire et analyser les liens qui unissent l'évaluation et la décision en santé. Il est décliné en deux objectifs secondaires avec un questionnement propre à chacun d'eux :

- **apprécier l'utilisation de l'évaluation pour la décision en matière de politiques de santé**

Cet objectif amène un questionnement sur les formes, les lieux, les modalités et les acteurs de l'utilisation :

- comment et où peut-on objectiver l'utilisation de l'évaluation ?
- par quoi se traduit-elle ?
- quels en sont les acteurs ?

- **identifier les facteurs et leviers associés à cette utilisation**

Cet objectif suggère d'aborder la question de deux angles : celui des décideurs et de l'évaluation :

- quelles sont les raisons qui motivent la demande d'évaluation et celles qui expliquent le désintérêt vis-à-vis des résultats de cette évaluation ?
- l'évaluation elle-même est-elle en capacité de soutenir la prise de décision ?

2. La revue de littérature

Pour approfondir cette problématique, une revue de littérature a été conduite en plusieurs étapes et selon deux méthodes.

Dans un premier temps, nous avons interrogé les bases de données en langue française (DOCTES de l'UCL) et anglaise (Scopus, Medline, Eric, Web of knowledge, Sciencedirect, Sage et Academic search, Premier d'Ebscohost) à partir de mots clés dont la liste a été progressivement étendue : decision-making, evidence-based policy, evidence-informed policy, evaluation institutionnalisation, knowledge management, knowledge translation, translational research, use, utilisation.

Une centaine d'articles ont été identifiés. A partir de la lecture de ces articles, d'autres références ont été identifiées, conduisant, de proche en proche, à identifier de nouveaux articles et à élargir le périmètre de la problématique.

Par la suite, des alertes ont été installées à partir de mots clés permettant de recueillir en continu les publications. De même, les sommaires des principales revues spécialisées en évaluation ont été consultés chaque mois, entre 2012 et 2014 : American journal of Evaluation, Evaluation and Program Planning, Evaluation Review, New Directions for Evaluation, Canadian Journal of Program Evaluation.

3. La problématique de l'utilisation au carrefour de plusieurs enjeux (article soumis)

La revue de littérature a permis d'enrichir la compréhension de la notion d'utilisation, d'identifier les facteurs d'influence et de faire émerger les enjeux propres aux relations entre la recherche et l'action. Ainsi, l'évaluation est interrogée dans sa capacité à produire du changement, à adapter ses méthodes à ses objets et à créer les conditions de l'appropriation des connaissances produites.

Par ailleurs, la question de la complexité, enjeu propre à l'évaluation des politiques, a été investiguée, conduisant à considérer l'évaluation elle-même comme une intervention complexe et à proposer un modèle théorique pour étudier le rôle et le poids des facteurs qui influencent l'utilisation ainsi que les interactions entre ces facteurs.

A conceptual framework for a better understanding of evaluation influence on policy decision making process

Françoise Jabot^{1,2*}, Jean Turgeon³, François Alla^{1,4}

¹ EA 4360 Apemac, Faculté de médecine, Université de Lorraine, France

² EHESP Rennes, Sorbonne Paris Cité, France

³ Ecole nationale d'administration publique, Canada

⁴ Inserm, CIC-EC, Centre hospitalier universitaire, France

*Corresponding author

Francoise.Jabot@ehesp.fr

Département Sciences humaines, sociales et des comportements de santé, Ecole des hautes études en santé publique, Avenue du Pr Léon Bernard, 35043 Rennes, France

Abstract

Although the use of evaluation for decision has been widely explored, it remains a recurrent concern for research. This article approaches the issue from three angles: firstly, it clarifies the concept of use and its determining factors; it then offers a summary of the contributions of research with respect to the relationship between the production and use of knowledge; lastly, it examines the challenge of complexity for evaluation. Given that evaluation is itself a complex intervention, the authors suggest a model based on a systemic approach to provide a better understanding of the effects of evaluation on the decision-making process. This model aims simultaneously to study the dynamic between the components of the evaluation system and the interplay of interactions between the categories of determining factors affecting use.

Key words

Evaluation influence, evaluation use, decision-making, complexity, health policy, knowledge dissemination

Introduction

Evaluation, defined as ‘the process of delineating, obtaining, and providing information for judging decision alternatives’ (Stuffelbeam, 1973, p129) is an activity present in institutions, in the political arena and in policy and programme planning cycles. Although it is enshrined in professional practices and in legislation, there appears to be little correlation between decisions affecting policy and programmes and the conclusions of researchers and evaluators ((Marmot, 2004; Muir Gray, 2004; Sheldon, 2005; Weiss, 1998). However, faced with choices that are constrained by limited budget resources and the need to be accountable to citizens, decision-makers have to direct their decisions towards the most effective and the most efficient actions – which is what evaluation is supposed to help them to do (Blake and Ottoson, 2009).

There therefore appears to be a paradox between a stated will for evaluation, and decision-making reasonings that are disconnected from evaluation. Research into the use of knowledge offers a major resource for understanding the issue of the use of evaluation (Blake and Ottoson, 2009). A function common to evaluation and research is that of producing knowledge. However, evaluation is different in that it is mandated, and focuses on a policy/programme of which it is supposed to assess the value, for which it must make proposals for improvement (Levin-Rozalis, 2003; Donaldson et al., 2009). The reasons put forward to explain the lack of use of knowledge include firstly, the complex, non-linear nature of the decision-making process (Hunter, 2009) and secondly, the gap between the world of research and that of decision-making, each of them governed by their own values, rules and reasoning (Choi et al., 2005; Jansen et al., 2010; Lomas, 2007). Research has provided only unsatisfactory, partial answers to the preoccupations of decision-makers confronted with the complexity of policies and the decisions to be taken (Green et al., 2009). What is more, knowledge-sharing processes have been insufficiently developed (Hunter, 2009; Ellen et al., 2013).

However, in the health sector, our field of study, the development of systematic reviews have favoured the use of knowledge in medical clinical practice, *evidence-based medicine* (EBM). That said, the application of knowledge to the field of health policy, *evidence-based policy* (EBP), has not experienced the same increase, although since the 1990s, initiatives have been developed (in particular, under the aegis of the World Health Organisation) to improve the contribution of research to the decision-making process. In this context, *evidence-informed policymaking* (EIPM) is an approach aimed at ensuring that decision-making is informed by

the best available evidence (Oxman et al., 2009). However, policies are complex in nature (Rogers, 2008) ; decisions applied on the scale of an entire population need to include many different parameters (Dobrow et al., 2004).

Given these findings, the question of the usefulness of evaluation arises. An evaluation is deemed to be useful if it affects policy decisions (Grob, 2003) and produces relevant knowledge for the policy being evaluated (Leviton and Hughes, 1981), which confirms previous results, provides fresh understanding of the reasons for the success of a programme, puts forward new hypotheses about a particular intervention (Ginsburg and Rhett, 2003).

The issue of evaluation use which has been thoroughly explored for over thirty years (Cousins and Leithwood, 1986; Johnson et al., 2009; Leviton and Hughes, 1981; Patton, 1997; Shulha and Cousins, 1997; Weiss, 1998), has allowed the understanding of the concept of use to be broadened, the main factors at play to be identified and the specific issues relating to the relationship between research and action to be highlighted. This means that evaluation is examined in terms of its capacity to produce change, adjust its methods to its goals and create the right conditions for the knowledge produced to be appropriated. Although there is plenty of literature on each of these subjects, the interaction between these three issues has not yet been widely explored.

The aim of this article is to offer, on the basis of this threefold interrogation, a framework for analysing the factors influencing the use of evaluation applied to health policies. Firstly, we will summarise the contributions of the literature on which we have based our hypotheses in order to establish our model, secondly present the model, and thirdly conclude by examining the outlooks offered by this approach.

The use of evaluation – a multidimensional concept

It is now accepted that use takes on a number of different forms. *Instrumental use* is a basis for action and decision-making that follows on from the findings and recommendations of evaluation. In this sense, it is an activity which is beneficial for public management, in particular to improve the performance of programmes (Wholey, 2004). *Conceptual use* is expressed individually (in terms of better understanding of the programme, its effects and improvement margins) and collectively and institutionally (through transformation of how the issue is perceived and reformulated). Evaluation is used to inform judgement (Weiss, 1998). It can also play a role in supporting positions and giving them legitimacy; this constitutes

symbolic use: for the purposes of legitimacy and persuasion. In addition, evaluation results in lessons being learned by the individuals or organisations that have taken part in the process; as a result, it is likely to produce conceptual, behavioural and cultural changes in them (Patton, 1997). This constitutes *process use*. Lastly, it may be mandatory to take results into account – this constitutes *imposed use* (Weiss, 2005). These categories do not apply simply to use (as opposed to non-use), but more importantly to usage.

We believe that these various dimensions of use and the consequences of evaluation can be incorporated within a single framework by recourse to the concept of influence, defined as the ‘capacity or power of individuals and things to produce effects on others by non-tangible or indirect means’ (Kirkhart, 2000). Kirkhart analyses influence according to its sources (actors or processes of change), their intentions and the timeframe within which the mechanisms operate. Henry and Mark (Henry and Mark, 2003) based themselves on her work and on programme theory to develop a model enabling the impact of evaluation to be appraised. In this development of influence theory, evaluation activities can be linked to short, medium and long-term consequences: both those that are direct and easily objectified and those that are indirect and more difficult to track. Their theoretical framework allows effects to be appreciated individually, interpersonally and collectively, along with the mechanisms at play in the production of changes at each of these levels. While use supposes intention and awareness, and relates to the exercise of power by those making the decisions, influence is indirect and takes many different pathways (Henry and Mark, 2003, p 305).

This makes it possible to exploit not only the conclusions and recommendations of the evaluation but also the ideas, the event of evaluating itself, and the data produced, including the type of method chosen (Weiss, 1998). Consequently, potential users include a wide range of groups which, due to their interest and position, have the capability to introduce change as a result of the evaluation (Patton, 1997; Weiss, 1998). The process of use is a complex phenomenon (Weiss, 1998) – a ‘black box’ between the evaluation being mandated and its unpredictable outcome.

Factors influencing use

Reviews of the literature on the use of evaluation (Blake and Ottoson, 2009; Cousins and Leithwood, 1986; Johnson et al., 2009; Lavis et al., 2005; Leviton and Hughes, 1981; Shulha and Cousins, 1997) have underscored the multiple factors at play. These include the type of

information produced, the characteristics of the users, the knowledge production process and the social, political and institutional context of the evaluation.

The nature of data

There is a consensus, particularly in the health sector, on the need to produce more evidence to ensure more effective and more efficient systems and policies (Donaldson et al, 2009) and to better use research-based evidence to inform decision-making (Oxman, 2009; World Health Organization, 2004). However, authors disagree as to what constitutes evidence (Donaldson et al., 2009). For Rychetnik, ‘evidence comprises the interpretation of empirical data derived for formal research or systematic investigations, using any type of science or social science methods’ (Rychetnik et al., 2002, p.119). The reliability and quality of evidence depends on the methods used to produce it, experimental methods being the gold standard for supplying the highest level of evidence as to the effectiveness of interventions (Rychetnik et al., 2002). The applications of the evidence-based approach are aligned with the concept of EBM and have led to other sources of evidence being neglected, whereas as far as EBP is concerned, decisions apply to populations and have to take into account the differing interests of stakeholders (Dobrow et al., 2004). Thus Schwandt rejects any definition of evidence which is too narrow – one that implies that evidence can only be scientifically put together and that it exists only in terms of probative value of causal effectiveness (Schwandt, 2009). The decision-making process relies on a set of considerations that extend beyond simple scientific knowledge and require other types of evidence (Bowen and Zwi, 2005; Head, 2010; Lavis, 2005; Oxman et al., 2009). The debate highlights two conceptions of evidence: in one, evidence expresses an objective reality and can be generalised; it is said to be *context-free*; in the other, evidence is informal, *colloquial*, embedded in reality and directed towards practical applications; it is said to be *context-sensitive* (Lomas, 2006). For Berriet-Sollicec et al. (2014), the quality of evidence depends on the relevance and strength of the evidence in question, as well as its suitability in view of the aims of the evaluation.

Furthermore, Dobrow et al. (2004) insist on the three stages (identification, interpretation, application) of the process of use of evidence. Indeed, the definition of evidence is an interpretive undertaking combining the evaluation of quality, validity and the potential for generalisation, the conclusions of which depend both on the stakeholders involved and on the goal of the policy (Dobrow et al., 2004; Oxman et al., 2009). A number of studies indicate that the views of researchers/evaluators and stakeholders may be highly

divergent when it comes to the interpretation of data (Chen, 2010; Pineault et al., 2007; Petticrew et al., 2004; Whitehead et al., 2004). The credibility of the evidence depends on more than simple notions of validity and method. Credibility is also built up with regard to the questions asked, the context and the interpretive framework (Donaldson et al., 2009).

The scientific quality of work, although essential, is not sufficient; a number of studies point to the weakness of the link between the production of convincing data and its use in public decision-making (Black, 2001; Marmot, 2004; Sheldon, 2005). Decision-makers are faced with many different political, economic, organisational, and media constraints and must satisfy the different interests of the population (Black, 2001). They seek to know whether the actions undertaken produce changes and why, in order to be able to reproduce them elsewhere or in other populations (Hunter, 2009).

Decision-makers are sensitive to all types of evidence, be they organisational, political or experimental (Pineault et al., 2007). They need data that address the whole gamut of political questioning, include contextual components (Lewin et al., 2009), shed light on complexity (Greene, 2009), allow uncertainty to be reduced (Contandriopoulos et al., 2008; Patton, 2010), and are applicable in their actual environment (Hunter, 2009; Greene, 2009; Owman et al., 2009). Decision-makers need data which are diverse, both quantitative and qualitative in nature (Schwandt, 2009), and derived not only from research but also from experts' judgements and citizens' preferences (Head, 2010; Hunter, 2009). They also need data which are opportune, enabling them to take decisions quickly (Innvaer et al., 2002). To be used, data must be relevant, complete, meaningful and credible (Chen, 2010).

The characteristics of users

The multiple uses of evaluation suggest a wide variety of users. The literature mentions three categories of actor, with a certain degree of overlap: users, stakeholders and decision-makers. The category of users covers all the individuals concerned by the intervention evaluated (decision-makers, financial backers, managers, professionals engaged with the public and beneficiaries), as well as more distant individuals (individuals in charge of similar programmes, lawmakers and the community of researchers and evaluators) (Weiss, 1998). Stakeholders are defined as the individuals concerned, i.e. they are preoccupied, interested or potentially affected by the intervention being evaluated and its evaluation (Greene, 2005; Patton, 1997, Weiss, 1998). Greene (2005) classifies stakeholders into four groups: individuals who have decision-making power, those who have some responsibility for the programme, the beneficiaries of the programme, and individuals who are disadvantaged by the programme. On

the basis of these definitions, the distinction between the users and stakeholders is not very clear. As regards stakeholders, inasmuch as the evaluation cannot answer all their questions, Patton (1997) suggests identifying a more restricted list of individuals – *intended users* – with whom to conduct the evaluation as regards its future uses. In this article, we shall use the three terms according to the posture of actors with reference to the three situations (decision-making, use and interest). The term ‘decision-maker’ is applied to individuals who have decision-making power, or the potential to introduce change into the system. The term ‘user’ designates individuals likely to benefit from the contributions of the evaluation. We have reserved the term ‘*stakeholders*’ for individuals who have a genuine stake in the evaluation, whether or not they are in a position to use it. These categories do not form homogenous entities. They bring together individuals driven by particular interests and strategies in the evaluation. The acceptability of evaluation is dependent on these interests, as is the credibility ascribed to the knowledge; this credibility is known to be influenced by their preferences and pre-existing opinions (Christie, 2007; Leviton, 2003; Patton, 1997; Sabatier, 1988, 2010; Weiss, 1998).

The context of the user’s decision depends on their mental model, i.e. the way they see the nature of the social problem and the way in which policies work (Leviton, 2003). According to the latter author, in order to meet users’ needs it is essential to understand their knowledge-building process, identify its shortfalls and their level of awareness of the limits of this knowledge. Encouraging use involves building up a body of evidence by doing more research (in order to reduce uncertainty) and conduct evaluation tailored to the users’ construct of reality and the expertise they provide. Decision-makers adopt programmes not only because of tested theories but also because of their opinions and perceptions, which they hold to be true. Christie’s work on the preference of decision-makers for data sources and particular types of information shows the link between their beliefs as to the effectiveness of programmes and the type of information required; they have a preference for information that corresponds the most closely to their own representation of the problem (Christie, 2007; Sabatier, 1988).

Patton (1997) underscores the role of other factors relating to political issues (especially power) and personal factors such as interest, engagement and determination. Given that decision-makers receive large amounts of different information that is the result of different opinions, the process of use of data from the evaluation competes with the interest of decision-makers (Black, 2001, Contandriopoulos and Brousselle, 2012), which in turn is

determined by what corresponds the most closely to their worldview and what appears to them as being the most plausible and reasonable (Schwandt, 2009).

The evaluation process

However, the credibility of data is not enough to establish a basis for the credibility of the evaluation itself. This depends on the argument put forward to reach the conclusions (Schwandt, 2009), the acceptability of this argument for stakeholders (Chen, 2010), and the ability to generate trust and belief (Hurteau et al., 2012). The credibility of evaluative judgement is the result of a complex process combining a twofold process of scientific and social legitimisation, brought into play in accordance with six principles. These include the iterative construction of the argument, the involvement of stakeholders in the process, and the competencies of the evaluator – the latter being essential to establishing both the appraisal and their own credibility (Hurteau et al., 2012).

Stakeholder involvement has been the subject of many published works (Brandon and Fukunaga, 2014) is considered to be a major factor in the use of evaluation results (Cousins and Leithwodd, 1986; Johnson et al., 2009; Leviton and Hughes, 1981; Patton, 1997; Plottu and Plottu, 2009). It forms the basis of a number of evaluation models, all of which are aimed at increasing the participation of the stakeholders concerned in the decisions (Rodriguez-Campos, 2011). According to these approaches, the evaluator plays an essential role in creating dialogue, facilitating discussion and enabling lessons to be learned (Morabito, 2002; Skolits et al., 2009). In addition, the breadth of actors and the nature of their participation depend on the institutional and organisational context in which the evaluation takes place. For instance, Preskill et al. (2003) suggest exploring the learning process using five categories of variable: capacity of the evaluator as a facilitator, managerial support, characteristics of the individuals concerned, communication process and characteristics of the organisation. In addition, the quality of the evaluator's relationship with the stakeholders, in particular those mandating the evaluation, is deemed vital to the success of the evaluation (Owen, 2012).

The context of evaluation

The role of the context in any evaluation has gained increased recognition in recent years (Fitzpatrick, 2012). Greene suggests analysing context from the point of view of its demographic, material, economic, institutional and organisational characteristics, as well as in

terms of the relational and political dynamics that drive it (Greene, 2005). Indeed, context forms a framework of social rules, standards and values that affect the mechanisms of the policy/programme (Stame, 2010) and determine the conditions of the evaluation's implementation (Rog, 2012) and use (Hojlund, 2014).

Furthermore, the institutional and organisational context plays a role in the implementation of evaluation. Depending on whether there is specific governance for the evaluation, expressed in procedure, resources, an action plan and suitable management within the mandating institution, the evaluation may or may not be conducted effectively and achieve its aims, whether these involve decision-making support, improved intervention or accountability (Hanberger, 2011). According to Hojlund (2014), research into use focuses in the main on the evaluation itself and does not sufficiently take into account the impact of the organisational context. He argues that the potential for use of the evaluation resides in a combination of an organisation's propensity to evaluate, the outside pressure to engage in evaluation and the means of institutional adaptation to change.

It follows that a number of variables should be taken into account for an in-depth analysis of context: the original problem, the evaluation, the decision, the overall environment (which often involves many layers and sites) (Rog, 2012), the time needed for change, resources, leadership, and the circumstances in terms of social support (Blake and Ottoson, 2012). It is important to analyse the context of the evaluation during all three main stages (project, implementation and use) (Conner et al., 2012).

Favourable conditions for the appropriation of knowledge

For some ten years now, one field of research has focused on the interface between the production and use of knowledge (Blake and Ottoson, 2009; Contandriopoulos et al., 2010; Lavis et al., 2003, 2005; Ross et al., 2003). Research findings have highlighted the gap between decision-makers and researchers, who have difficulty understanding their respective versions of reality (Choi et al, 2005; Innvaer et al., 2002; Jansen et al., 2010; Lomas, 2007; Macintyre, 2012; Petticrew et al., 2004; Whitehead et al., 2004). Indeed, decision-makers bemoan the lack of knowledge (and sometimes even naivety) they perceive on the part of researchers themselves, in the way organisations operate, and in institutional and decision-making mechanisms. Researchers and decision-makers have differing perceptions of research, the nature and strength of evidence and the use of the knowledge produced. Research has also

highlighted the lack of trust and recognition between these two categories of actor, each with its own mindset relating to its own system of values and beliefs, and concludes that action must be taken to facilitate these two worlds drawing closer together.

Knowledge-for-action theories have a useful contribution to make when it comes to understanding the use of evaluation (Ottoson, 2009). Different theories relating to different disciplines explore the conditions in which knowledge can be passed on and applied, depending on the forms of knowledge in question (ideas, teaching, innovations), the means of communication, and the relationships established between producers and users of knowledge (Ottoson, 2009). A number of models linking research and action (*push, pull, exchange/linkage activities*) have been drawn up in an attempt to analyse the various means of use depending on the origin of the research, the type of transmission and interaction between actors, with knowledge being taken either as a product moving between transmitters and receivers or being seen as the result of negotiations between stakeholders (Blake and Ottoson, 2009). Following on from many years that saw little success in terms of extending research in the direction of policy (*push*), an opposing model has emerged, according to which initiative for the direction of research should be taken by decision-makers (*pull*), as well as an interactive model that calls for knowledge management (World Health Organisation, 2006).

The interactions between knowledge producers and decision-makers are determining factors in the appropriation of research findings by the latter (Contandriopoulos et al., 2010; Lavis et al., 2003, 2005; Lomas, 2007; Pineault et al., 2007). The more accurately evaluators identify types of use, the better their evaluation method and the better chance there is of their work being used (Shulha and Cousins, 1997). This means that evaluators need to work with potential users to identify the contextual components that are germane to knowledge sharing, better understand the decision-making process (Lavis et al., 2003), target values, and transform beliefs (Black, 2001; Leviton, 2003).

This has a number of consequences when it comes to evaluation. On an operational level, this can be expressed in terms of the development of knowledge transfer mechanisms (Cambon and Alla, 2013; Ellen et al., 2013), along with mechanisms for establishing room for dialogue between the evaluator and potential users. This can make it possible to provide better understanding of key issues and needs, draw up adequate evaluation designs, and communicate in a way that is understandable (Christie, 2007). This approach interrogates the role of the evaluator: some authors are of the opinion that in addition to examining how

knowledge is used, evaluators should also investigate how it is adopted and transformed (Blake and Ottoson, 2009).

Favouring the use of the results of evaluation encourages the production of work that is in line with users' preoccupations, creating a favourable framework by including decision-makers and other stakeholders in the thinking and having evaluators who play the role of mediators, responding to the context. The development of *knowledge brokering* is proving to be a promising avenue (Lomas, 2007; Munerol et al., 2013).

Evaluation and the complexity of interventions

Complexity has been a growing preoccupation for evaluators in recent years. The approach to complexity is based on systemic analysis developed in a number of areas in order to understand how systems work, in particular health systems (De Savigny et al., 2009). Complexity relates to interventions themselves, as well as the systems of which they are a part (Shiell et al., 2008). Interventions are complex because they include a number of interdependent components that interact with one another (Craig et al., 2008; Rogers, 2008). Systems are complex because they are made up of embedded sub-systems that are constantly changing in order to endure and adapt to the environment (Rickles et al., 2007). They are also characterised by the non-linear, unpredictable nature of the changes they produce, the non-proportionality of changes to a component in the system compared to their influence, and the presence of internal tensions to maintain the survival of the system. In social terms, complex adaptive systems are defined as 'a collection of individual agents that have freedom to act in ways that are not always totally predictable and whose actions are interconnected so that one agent's actions change the context for others' (Plsek and Greenhalgh, 2001). Thus Hawe, Shiell, et al. (2009b) consider interventions to be events that alter a system by creating new relationships, transforming roles and representations, and causing new opportunities, thus allowing change.

Policies/programmes, especially health policies, have the characteristics of complex interventions. They are complicated because they include a number of components, require governance that is both vertical and horizontal and combine several strategies, simultaneously implemented by different actors operating at different levels and enshrining different representations and reasoning. They are complex due to the many different interactions and their uncertain future, which is dependent on the unpredictable contributions of actors and

subject to unforeseen circumstances in the context. Interventions are all the more complex in that they deal with *wicked issues*, calling for combined actions and leveraging at the level of individuals, communities or systems (Barnes et al., 2003).

Faced with complexity, the approaches to evaluation need to be renewed and adjusted. First of all, systemic thinking leads to a paradigm shift in the way interventions are seen: no longer as fixed, normative frameworks but as moving structures, calling into question approaches that focus on ‘program fidelity’ (Hawe, Bond, et al., 2009b). Indeed, theories of complexity challenge positivist standards (Sanderson, 2000) and approaches built on hypotheses based on stable, balanced, linear relationships between variables and the proportionality of changes in response to causal influences are not appropriate when it comes to understanding complex systems. And so it is that experimental methods are no longer deemed to be the most relevant for the evaluation of complex interventions (Green et al., 2009; Mackenzie et al., 2010; Stame, 2010).

The challenges raised by the systemic approach have a number of consequences on evaluation (Walton, 2014): approaching interventions as a whole and analysing the interactions between the components and the context; understanding programme theory; addressing the implementation process; identifying the mechanisms that explain change; taking uncertainty into account; and using mixed methods. These call for the design of tailor-made evaluations for each case (Mackenzie et al., 2010), developing innovative models to collect probative data. A number of strategies built on programme theory, *realist* approaches, and configuration analysis have been put forward to rise to this challenge (Contandriopoulos et al., 2011 ; Hawe et al, 2009a; Ling, 2012; Patton, 2010; Pawson, 2002; Rogers, 2008; Stame, 2004; Touati, 2011). They illustrate the choices made (in terms of evaluation design) by evaluators confronted with the main methodological challenges identified (Walton, 2014) when dealing with questions relating to multiples, uncertainty, unpredictability, paradoxes inherent to complexity, and long-term temporality.

We consider that evaluation itself may be deemed to be a system of complex actions. Breaking down evaluation into a series of specific activities at a number of stages in the process entails as many components and mobilises a broad range of actors, with various responsibilities, who do not share the same expectations or even the same reference frameworks. Each evaluation takes place in a specific environment which shapes the organisation of the process. The effects of the evaluation are neither linear nor totally predictable, since the participation, acceptance and involvement of the actors is subject to a set

of conditions that are neither stable nor controlled. Lastly, evaluation is a space inhabited by multiple representations, a wide range of approaches and practices, and numerous (sometimes contradictory) expectations.

The future of evaluation remains uncertain, in spite of the increased number of initiatives aimed at making policies better-informed (Bowen and Zhi, 2005; Oxman et al., 2009). The obstacles or factors favouring the incorporation of knowledge in the decision-making process have been identified (Jewell and Bero, 2008) but the combination of many different influences is still a matter for research (Mark, 2011).

A model for analysing the influence of evaluation on policies

Apprehending evaluation as a complex intervention leads us to consider it in a non-segmented manner, analysing the interactions between the evaluation process, the actors and the context, identifying change-generating mechanisms and taking into account the constraints relating to paradoxical situations, divergences, uncertainty and timescale.

We thus suggest a model (see Figure 1 below) that draws inspiration from the work of Johnson (1998) and Mark and Henry (2004). Johnson's approach offers a systemic perspective for the analysis of the effects of evaluation. Indeed, it sees the use of evaluation as an ongoing, diffuse process in an open system, with many different interrelations between variables. The model developed by Mark and Henry enables the possible consequences of the evaluation – its *influence pathways* – to be tracked, along with the levels on which it is exerted, and details the underlying mechanisms (Mark and Henry, 2004). We suggest a model that combines Johnson's systemic view and the contributions of Mark and Henry's influence theory; the configuration of our model providing greater visibility of the ongoing dynamic of interactions than the linear chain-of-effects model.

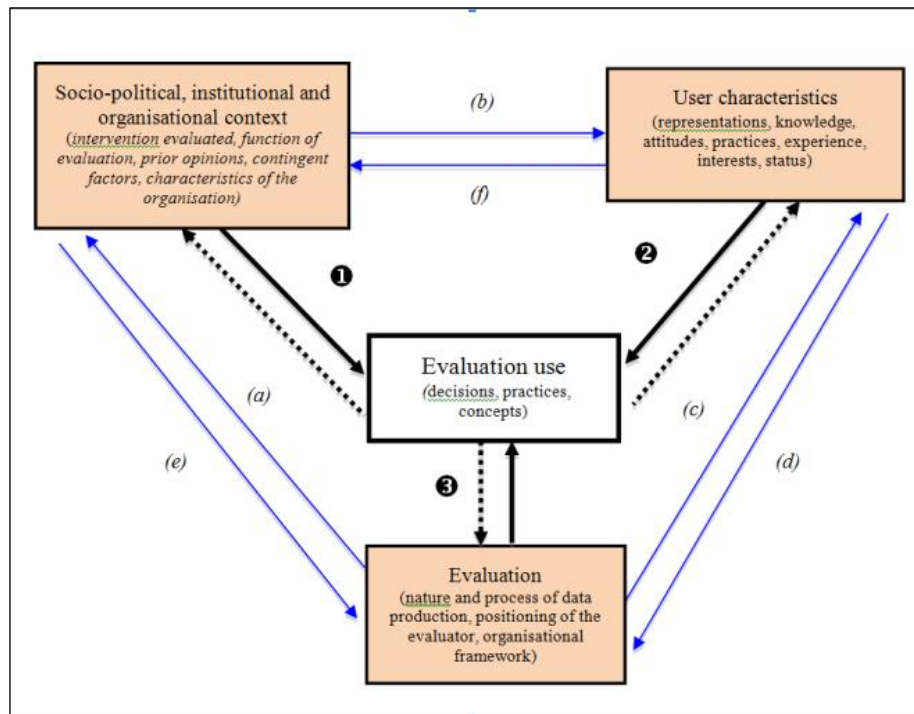


Figure 1. Conceptual framework

Our model aims to study the consequences of the evaluation (*what*), shown in the centre of the model; the reasons (*why*) for which they occur (three shaded squares); and the interactions between these components (arrows). The consequences are influenced by three main factors: the evaluation, insofar as it produces useful, credible, knowledge that can be appropriated; the characteristics of the individuals liable to make use or have use made of it (users); and the context, facilitating or hindering change. The mechanisms explaining change based on these three categories of determining factors are shown by unbroken arrows. The dotted arrows in the opposite direction indicate the possible feedback of use on these determining factors. Indeed, we advance the hypothesis that an evaluation which is used reinforces the relevance of this function, demonstrates its validity and makes the evaluation exercise credible, transforming the attitudes of users towards the evaluation. Lastly, the three categories of determining factors may influence each other (two-way arrows).

The changes brought about by evaluation (what)

For use to be tracked, it is necessary to identify what is used and how (Blake and Ottoson, 2009). What may be used involves knowledge and innovations and is expressed in the implementation of policies (DeGroff and Cargo, 2009). The notion of use should be taken in the broadest sense, including many different dimensions.

It follows that cognitive, behavioural and organisational changes (Henry and Mark, 2003) may be identified in the short, medium and long term, on the basis of the following:

- concrete decisions: modification, extension, renewal of the intervention, administrative or financial measures, new regulations;
- the practices of actors and/or organisations: planning, implementation, governance of programmes and policies (architecture, priorities, intervention strategies, rules, procedures), restructuring;
- actors' representations, opinions, attitudes and concepts.

The determining factors of change (why)

The components for initiative in change fall into three categories of determining factors for use, which can be divided into sub-categories and variables (see Table 1 below).

The evaluation process. The evaluation process has been studied as to its capacity for producing useful knowledge and favouring the use of knowledge by future users. We shall focus on the evaluation mandate, the existing organisation, the positioning of the evaluator, the data produced, the data production process and how the evaluation is conducted.

We have chosen the components identified in the literature as favouring the use of evaluation: clarity of expression in the evaluation mandate (Patton 1997, Perret, 2008; Owen, 2012), the scope and involvement of stakeholders, the existence of interaction between the evaluator and the stakeholders (Patton 1997), the ability of the evaluation to favour dialogue between stakeholders (Skolits, 2009), the suitability of the method given the complexity of the subject (Hawe et al, 2009a), the quality and credibility of the data (Schwandt, 2009), and the capacity of the evaluation to shed light on choices and meet expectations (Greene, 2009; Hawe et al, 2009a).

Characteristics of users. As the use of the evaluation takes many pathways, we include in this category the intended users (Patton, 1997), i.e. decision-makers and preparers, the individuals with the ability to influence the decision-maker (Sabatier, 1988) and all those who have the

potential to alter the action and introduce change in the system or organisation, in particular individuals in charge of the interventions evaluated.

User characteristics are observed firstly on an individual level, with respect to their motivations, representations, knowledge, attitudes and practices, and secondly on an interpersonal level, through their ability to influence and interact with other actors. The variables chosen are as follows: representations of problem-solving methods (Christie, 2007; Leviton, 2003; Weiss, 1983), expectations, status in the organisation, scientific knowledge, opinions on the benefits of the evaluation, interest in the subject evaluated, experience and/or expertise in evaluation or in the subject evaluated (Preskill and Torres, 2003), leadership (Patton, 1997). This involves appraising the level of engagement of users and identifying the reasons for which individuals will have the contributions of the evaluation adopted within the organisation, as well as how this is done.

The context. The context in which the evaluation is performed is determining in a number of respects: the consideration given to the problem and the policy concerned; the level of complexity of the intervention in question; and the organisational and institutional climate of the intervention and of the evaluation (Rog, 2012).

The contextual variables adopted relate to the following: a) the intervention (importance of the issue and policy it involves, complexity of the intervention); b) the institutional and organisational environment of the evaluation (recognition of the function, resources, propensity to evaluate, procedures of the organisation); c) the contingent factors that interfere with decision-making (Appleton-Dyer et al., 2012; Conner and Rog et al., 2012; Firtzpatrick, 2012; Hanberger, 201; Hojlund, 2014).

Table 1. Determining factors in evaluation use

Socio-political context	Evaluation process	Characteristics of users
<p>Intervention evaluated</p> <ul style="list-style-type: none"> social significance of the problem nature of the problem key issues relating to the problem degree of complexity of the intervention number of components number of actors / institutions degree of uncertainty about the action theory degree of consensus about the problem-solving method <p>Evaluation in the institution</p> <ul style="list-style-type: none"> degree of recognition view of evaluation organisation of the function of evaluation <p>Other factors</p> <ul style="list-style-type: none"> other priorities on the agenda budget resources organisation's capacity to change internal communication time 	<p>Mandate</p> <ul style="list-style-type: none"> - clarity of aims and expectations - detailed questions <p>System</p> <ul style="list-style-type: none"> - variety of stakeholders - organisation of relationships between actors and agencies - nature of exchanges - communication <p>Evaluator</p> <ul style="list-style-type: none"> - role - positioning adopted - approaches and models <p>Method</p> <ul style="list-style-type: none"> - coherence of stages - nature of data produced - data production process - timeframe - feedback and dissemination of findings 	<p>Users</p> <p>(individual level)</p> <ul style="list-style-type: none"> - representations of problem-solving methods - interest in the subject evaluated - scientific knowledge - opinions about the benefits of evaluation - experience and/or expertise in evaluation or the topic being evaluated <p>(interpersonal level)</p> <ul style="list-style-type: none"> - status in the organisation - leadership - interest in the results of the evaluation - evaluator-decision-maker interface

Interactions between determining factors and the mechanisms at work

The socio-political and organisational context within the institution determines the framework (type and approach of evaluation) and governance of the evaluation (a), and models the representations, interests and expectations of users, particularly those who are within the institution (b). The latter mandate an evaluation expressing the respective positions of these standards and their own set of representations of the evaluation (c). As for evaluators, they adapt their practice to the expectations expressed whilst taking into account the requirements fixed by the institution. The evaluation mandate – which is determining for the future of the evaluation – is the result of negotiations between these three components. In return, the evaluation process may, due for example to the engagement of the evaluator, contribute to raising awareness in some key actors in the system (d), encourage the sharing of a reference

framework for the evaluation, and thus alter the evaluation (as a function) within the organisation (e). Similarly, users, whatever their position in the system, may influence these changes, commensurately with their status or leadership (f).

In our model, use does not appear as the outcome of a linear process but is instead positioned at the centre of the model. This arrangement makes it possible to view converging influences from the three categories of factors chosen (unbroken arrows), the feedback effect following use of the evaluation on the categories of determining factors (dotted arrows), and the circular nature of relationships between the three categories of determining factors.

Conclusion

The proposed model puts forward an overall analysis and aims to answer a series of questions. To what extent does evaluation, as a process and a product, meet users' expectations? To what extent does the evaluation process encourage knowledge to be shared with users? To what extent is the institutional and organisational environment favourable to the implementation of useable evaluation, i.e. evaluation which is high-quality and credible? To what extent will users encourage the contributions of the evaluation to be adopted? Answering each of these questions involves identifying the change-generating mechanisms: on an individual, interpersonal and collective level; and on the basis of the sub-categories studied (Mark and Henry, 2004). Did a social or political event increase interest in the subject being evaluated? Did knowledge that emerged during the course of the evaluation challenge the representations of a key actor in the system? Why – and how – does an actor become an agent of change?

Our model aims to study, simultaneously, the dynamics between the components of the evaluation system (activity conducted by actors in a given environment) and the interplay of compensations that keep the whole in balance. Indeed, from the systemic perspective, altering one component of the system causes repercussions on other components, activating mechanisms to maintain cohesion and stability (Sicotte et al., 1998). To fulfil the evaluation and its repercussions, one avenue of investigation would be how individuals' commitment and the level of responsiveness of the organisation on the one hand balance with the scope of the evaluation process and the level of recognition of the evaluation function on the other.

This model fits with two recent publications on evaluation use. First, Saunders (2012) emphasizes a distinction between 'use' and 'usability'. 'Use' refers to the capacity of stakeholders to make use of evaluation (see interactions between boxes 'users' and 'context')

while ‘usability’ refers to the capacity of evaluation design to facilitate or inhibit its potential use (see interactions between boxes ‘evaluation’ and ‘context’). Second, Lederman (2012) addresses this issue in combining mechanisms of evaluation use, context conditions, actors perceptions (see triangle between boxes ‘use’, ‘users’, and ‘context’). Relying on the conception of evaluation as a complex phenomenon, that need a global and systemic approach, our framework is a way to link the three dimensions of the issue as studied by Saunders and Lederman.

This article constitutes a further step towards increased understanding of the effects of evaluation on the conception and implementation of policies. It is a challenge for research to objectivise the medium and long-term effects of evaluation, since the traceability of these effects is diluted at a number of levels. This challenge must be pursued for two reasons: firstly, to develop better understanding of the interplay between interactions, in order to adjust practice and create favourable conditions for use; and secondly, to increase visibility of the usefulness of evaluation and promote the development of relevant, high-quality evaluations.

Références

- Appleton-Dyer S, Clinton J, Carswell P, et al. (2012) Understanding Evaluation Influence Within Public Sector Partnerships: A Conceptual Model. *American Journal of Evaluation*, 33(4), 532–546.
- Barnes M, Matka E and Sullivan H (2003) Evidence, Understanding and Complexity Evaluation in Non-Linear Systems. *Evaluation*, 9(3), 265–284.
- Berriet-Sollicec M, Labarthe P and Laurent C (2014) Goals of evaluation and types of evidence. *Evaluation*, 20(2), 195–213.
- Black N (2001) Evidence based policy: proceed with care. *BMJ: British Medical Journal*, 323(7307), 275–279.
- Blake SC and Ottoson JM (2009) Knowledge utilization: Implications for evaluation. *New Directions for Evaluation*, 2009(124), 21–34.
- Bowen S and Zwi AB (2005) Pathways to ‘evidence-informed’ policy and practice: a framework for action. *PLoS medicine*, 2(7), e166.

- Brandon PR and Fukunaga LL (2014) The State of the Empirical Research Literature on Stakeholder Involvement in Program Evaluation. *American Journal of Evaluation*, 35(1), 26–44.
- Cambon and Alla (2013) Transfert et partage de connaissances en santé publique : réflexions sur les composantes d'un dispositif national en France. *Santé Publique*, Vol. 25(6), 757–762.
- Chen HT (2010) The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning*, 33(3), 205–214.
- Choi BCK, Pang T, Lin V, et al. (2005) Can scientists and policy makers work together? *Journal of Epidemiology and Community Health*, 59(8), 632–637.
- Christie CA (2007) Reported Influence of Evaluation Data on Decision Makers' Actions An Empirical Examination. *American Journal of Evaluation*, 28(1), 8–25.
- Conner RF, Fitzpatrick JL and Rog DJ (2012) A first step forward: Context assessment. *New Directions for Evaluation*, 2012(135), 89–105.
- Contandriopoulos D and Brousselle A (2012) Evaluation models and evaluation use. *Evaluation*, 18(1), 61–77.
- Contandriopoulos D, Brousselle A and Kêdoté N (2008) Evaluating Interventions Aimed at Promoting Information Utilization in Organizations and Systems. *Healthcare Policy | Politiques de Santé*, 4(1), 89–107.
- Contandriopoulos D, Lemire M, Denis J-L, et al. (2010) Knowledge Exchange Processes in Organizations and Policy Arenas: A Narrative Systematic Review of the Literature. *Milbank Quarterly*, 88(4), 444–483.
- Contandriopoulos A-P, Rey L, Brousselle A, et al. (2011) Evaluer une intervention complexe: enjeux conceptuels, méthodologiques et opérationnels. *Canadian Journal of Program Evaluation*, 26(3), 1-16.
- Cousins JB and Leithwood KA (1986) Current empirical research on evaluation utilization. *Review of educational research*, 56(3), 331–364.
- Craig P, Dieppe P, Macintyre S, et al. (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, 337(sep29 1), a1655–a1655.

- De Savigny D, Adam T, Alliance for Health Policy and Systems Research, et al. (2009) *Systems thinking for health systems strengthening*. Geneva: Alliance for Health Policy and Systems Research: World Health Organization, Available from: <http://public.eblib.com/EBLPublic/PublicView.do?ptiID=476146> (accessed 22 April 2014).
- DeGross A and Cargo M (2009) Policy implementation: Implications for evaluation. *New Directions for Evaluation*, 2009(124), 47–60.
- Dobrow MJ, Goel V and Upshur REG (2004) Evidence-based health policy: context and utilisation. *Social science & medicine (1982)*, 58(1), 207–217.
- Donaldson SI, Christie CA and Mark MM (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* SAGE.
- Ellen ME, Léon G, Bouchard G, et al. (2013) What supports do health system organizations have in place to facilitate evidence-informed decision-making? A qualitative study. *Implementation science: IS*, 8, 84.
- Fitzpatrick JL (2012) An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, 2012(135), 7–24.
- Ginsburg A and Rhett N (2003) Building a Better Body of Evidence: New Opportunities to Strengthen Evaluation Utilization. *American Journal of Evaluation*, 24(4), 489–498.
- Green LW, Glasgow RE, Atkins D, et al. (2009) Making Evidence from Research More Relevant, Useful, and Actionable in Policy, Program Planning, and Practice. *American Journal of Preventive Medicine*, 37(6), S187–S191.
- Grob GF (2003) A Truly Useful Bat is One Found in the Hands of a Slugger. *American Journal of Evaluation*, 24(4), 499–505.
- Guillemette F, Hurteau M and Houle S (2012) *L'évaluation de programme axée sur le jugement crédible*. PUQ.
- Hanberger A (2011) The real functions of evaluation and response systems. *Evaluation*, 17(4), 327–349.
- Hawe P, Bond L and Butler H (2009) Knowledge theories can inform evaluation practice: What can a complexity lens add? *New Directions for Evaluation*, 2009(124), 89–100.

- Hawe P, Shiell A and Riley T (2009) Theorising Interventions as Events in Systems. *American Journal of Community Psychology*, 43(3-4), 267–276.
- Head BW (2010) Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society*, Evidence-based Policy, 29(2), 77–94.
- Henry GT and Mark MM (2003) Beyond Use: Understanding Evaluation’s Influence on Attitudes and Actions. *American Journal of Evaluation*, 24(3), 293–314.
- Hojlund S (2014) Evaluation use in the organizational context - changing focus to improve theory. *Evaluation*, 20(1), 26–43.
- Hunter DJ (2009) Relationship between evidence and policy: A case of evidence-based policy or policy-based evidence? *Public Health*, 123(9), 583–586.
- Innvaer S, Vist G, Trommald M, et al. (2002) Health policy-makers’ perceptions of their use of evidence: a systematic review. *Journal of health services research & policy*, 7(4), 239–244.
- Jansen MW, Oers HA van, Kok G, et al. (2010) Public health: disconnections between policy, practice and research. *Health Research Policy and Systems*, 8(1), 37.
- Jewell CJ and Bero LA (2008) ‘Developing good taste in evidence’: facilitators of and hindrances to evidence-informed health policymaking in state government. *The Milbank quarterly*, 86(2), 177–208.
- Johnson K, Greenesid LO, Toal SA, et al. (2009) Research on Evaluation Use A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
- Johnson RB (1998) Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21(1), 93–110.
- Kirkhart KE (2000) Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000(88), 5–23.
- Lavis J, Davies H, Oxman A, et al. (2005) Towards systematic reviews that inform health care management and policy-making. *Journal of Health Services Research & Policy*, 10(suppl 1), 35–48.

- Lavis JN, Robertson D, Woodside JM, et al. (2003) How Can Research Organizations More Effectively Transfer Research Knowledge to Decision Makers? *Milbank Quarterly*, 81(2), 221–248.
- Ledermann S (2012) Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159–178.
- Levin-Rozalis M (2003) Evaluation and research: Differences and similarities. *Canadian Journal of Program Evaluation*, 18(2), 1–32.
- Leviton LC (2003) Evaluation Use: Advances, Challenges and Applications. *American Journal of Evaluation*, 24(4), 525–535.
- Leviton LC and Hughes EF (1981) Research On the Utilization of Evaluations A Review and Synthesis. *Evaluation review*, 5(4), 525–548.
- Lewin S, Oxman AD, Lavis JN, et al. (2009) SUPPORT Tools for evidence-informed Policymaking in health 11: Finding and using evidence about local conditions. *Health Research Policy and Systems*, 7(Suppl 1), S11.
- Ling T (2012) Evaluating complex and unfolding interventions in real time. *Evaluation*, 18(1), 79–91.
- Lomas J (2007) The in-between world of knowledge brokering. *BMJ*, 334(7585), 129–132.
- Lomas J (2006) In WHO Bridging th "Know-Do" Gap, Meeting on knowledge translation in Global Health, 10-12 October 2005, Geneva: World Health Organization, Available at http://www.who.int/kms/WHO_EIP_KMS_2006_2.pdf
- Macintyre S (2012) Evidence in the development of health policy. *Public health*, 126(3), 217–219.
- Mackenzie M, O'donnell C, Halliday E, et al. (2010) Evaluating complex interventions: one size does not fit all. *BMJ. British medical journal*, 340(7743), 401–403.
- Mark MM (2011) Toward better research on—and thinking about—evaluation influence, especially in multisite evaluations. *New Directions for Evaluation*, 2011(129), 107–119.
- Mark MM and Henry GT (2004) The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10(1), 35–57.

- Marmot MG (2004) Evidence based policy or policy based evidence? *BMJ (Clinical research ed.)*, 328(7445), 906–907.
- Mays N, Pope C and Popay J (2005) Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research & Policy*, 10(suppl 1), 6–20.
- Morabito SM (2002) Evaluator Roles and Strategies for Expanding Evaluation Process Influence. *American Journal of Evaluation*, 23(3), 321–330.
- Muir Gray JA (2004) Evidence based policy making. *BMJ: British Medical Journal*, 329(7473), 988–989.
- Munero L, Cambon L and Alla F (2013) Le courtage en connaissances, définition et mise en œuvre : une revue de la littérature. *Santé Publique*, Vol. 25(5), 587–597.
- Ottoson JM (2009) Knowledge-for-action theories in evaluation: Knowledge utilization, diffusion, implementation, transfer, and translation. *New Directions for Evaluation*, 2009(124), 7–20.
- Oxman AD, Lavis JN, Lewin S, et al. (2009) SUPPORT Tools for evidence-informed health Policymaking (STP) 1: What is evidence-informed policymaking? *Health Research Policy and Systems*, 7(Suppl 1), S1.
- Patton MQ (1997) *Utilization-Focused Evaluation: The New Century Text*. SAGE Publications.
- Patton MQ (2010) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.
- Pawson R (2002) Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, 8(3), 340–358.
- Perret B (2006) De l'échec de la rationalisation des choix budgétaires (RCB) à la loi organique relative aux lois de finances (LOLF). *Revue française d'administration publique*, n°117(1), 31–41.
- Perret B (2008) *L'évaluation des politiques publiques*. Paris: La Découverte.
- Petticrew M, Whitehead M, Macintyre S, et al. (2004) Evidence for public health policy on inequalities: 1: The reality according to policymakers. *Journal of Epidemiology and Community Health*, 58(10), 811–816.

- Pineault R, Tousignant P, Roberge D, et al. (2007) Involving Decision-Makers in Producing Syntheses: The Case of the Research Collective on Primary Healthcare in Quebec. *Healthcare Policy | Politiques de Santé*, 2(4), Available from: <http://www.longwoods.com/content/18867> (accessed 25 May 2014).
- Plottu B and Plottu E (2009) Approaches to Participation in Evaluation Some Conditions for Implementation. *Evaluation*, 15(3), 343–359.
- Plsek PE and Greenhalgh T (2001) The challenge of complexity in health care. *BMJ: British Medical Journal*, 323(7313), 625–628.
- Preskill H, Zuckerman B and Matthews B (2003) An Exploratory Study of Process Use: Findings and Implications for Future Research. *American Journal of Evaluation*, 24(4), 423–442.
- Rickles D, Hawe P and Shiell A (2007) A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health*, 61(11), 933–937.
- Rodriguez-Campos L (2011) Stakeholder involvement in evaluation: three decades of the American journal of evaluation. *Journal of MultiDisciplinary Evaluation*, 8(17), 57–79.
- Rog DJ (2012) When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation*, 2012(135), 25–40.
- Rogers PJ (2008) Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29–48.
- Ross S, Lavis J, Rodriguez C, et al. (2003) Partnership experiences: involving decision-makers in the research process. *Journal of Health Services Research & Policy*, 8(suppl 2), 26–34.
- Rychetnik L, Frommer M, Hawe P, et al. (2002) Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*, 56(2), 119–127.
- Sabatier Paul (1988) An advocacy coalition framework of policy change and the role of policy-oriented learning therein - Springer. Available from: <http://link.springer.com/article/10.1007/BF00136406#page-1> (accessed 9 April 2014).
- Sanderson I (2000) Evaluation in Complex Policy Systems. *Evaluation*, 6(4), 433–454.

- Saunders M (2012) The use and usability of evaluation outputs: A social practice approach. *Evaluation*, 18(4), 421–436.
- Sheldon TA (2005) Making evidence synthesis more useful for management and policy-making. *Journal of health services research & policy*, 10 Suppl 1, 1–5.
- Shiell A, Hawe P and Gold L (2008) Complex interventions or complex systems? Implications for health economic evaluation. *BMJ*, 336(7656), 1281–1283.
- Shulha LM and Cousins JB (1997) Evaluation Use: Theory, Research, and Practice Since 1986. *American Journal of Evaluation*, 18(3), 195–208.
- Sicotte C, Champagne F, Contandriopoulos AP, et al. (1998) A Conceptual Framework for the Analysis of Health Care Organizations' Performance. *Health Services Management Research*, 11(1), 24–41.
- Skolits GJ, Morrow JA and Burr EM (2009) Reconceptualizing Evaluator Roles. *American Journal of Evaluation*, 30(3), 275–295.
- Stame N (2004) Theory-Based Evaluation and Types of Complexity. *Evaluation*, 10(1), 58–76.
- Stame N (2010) What Doesn't Work? Three Failures, Many Answers. *Evaluation*, 16(4), 371–387.
- Stuffelbeam D.L (1973) An introduction to the PDK book: Educational evaluation and decision-making. In B.R. Worthen & J.R. Sanders (Eds), *Educational evaluation: Theory and practice*. Belmont, CA: Wadsworth(pp 128-142).
- Tabuteau D (2010) L'expert et la décision en santé publique. *Les Tribunes de la santé*, 27(2), 33–48.
- Touati N (2011) L'évaluation des interventions complexes: quelle peut être la contribution des approches configurationnelles? *Canadian Journal of Program Evaluation*, 26(3), 17-35.
- Walton M (2014) Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning*, 45, 119–126.
- Ward VL, House AO and Hamer S (2009) Knowledge brokering: Exploring the process of transferring knowledge into action. *BMC Health Services Research*, 9(1), 12.

- Weiss CH (1998) Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, 19(1), 21–33.
- Weiss CH (2005) An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. *American Journal of Evaluation*, 26(1), 12–30.
- Whitehead M, Petticrew M, Graham H, et al. (2004) Evidence for public health policy on inequalities: 2: Assembling the evidence jigsaw. *Journal of Epidemiology and Community Health*, 58(10), 817–821.
- Wholey (2004) Using Evaluation to improve Performance and Support Policy Decision Making. In: *Evaluation Roots*, Sage Publications, pp. 267–275.
- World Health Organization (2004) *World report on knowledge for better health: strengthening health systems*. Geneva: World Health Organization.

Chapitre 4 : La méthode

La revue de littérature et le cadre conceptuel ont permis d'affiner les hypothèses et la méthode de recherche.

1. Les hypothèses de recherche

Pour que l'évaluation soit utilisée et utile, nous faisons les hypothèses suivantes :

- les connaissances produites (et le processus de production) doivent être crédibles, c'est à dire valides sur le plan scientifique et porteuses de sens ;
- il existe une interface entre le monde des décideurs et celui des chercheurs/évaluateurs propice à la construction d'une compréhension commune du problème et des solutions, condition indispensable à l'appropriation des résultats ;
- le contexte dans lequel s'inscrit l'évaluation doit être favorable : politique (opinion publique, lobbies, valeurs..), culture institutionnelle (histoire des institutions comme obstacle au développement de l'évaluation), rôle des leaders (experts influents).

Le questionnement est ainsi décliné :

- Les données produites sont-elles des données pertinentes, utiles, c'est à dire, exploitables pour guider l'action publique, et crédibles pour les décideurs ?
- Existe-t-il des relations entre le producteur et l'utilisateur des connaissances ? Sont-elles de nature à faciliter une appropriation des connaissances ?
- Quelle est la capacité de l'évaluation à s'adapter aux besoins des utilisateurs en conservant une légitimité méthodologique et scientifique ?
- Dans quelle mesure le contexte est-il un atout ou un obstacle pour l'utilisation des travaux d'évaluation, c'est-à-dire, créé des conditions satisfaisantes augurant d'un bon potentiel d'utilisation ?

2. Démarche de recherche et recueil de données

Le cadre conceptuel ayant été présenté, le travail de recherche pour le mettre à l'épreuve a été conduit en plusieurs étapes : analyser les évaluations des PRSP pour estimer leur contribution à la compréhension des politiques régionales de santé (2.1.) ; analyser de façon approfondie le processus d'évaluation dans une région (2.2.) ; apprécier l'utilisation à court terme des évaluations dans neuf régions (2.3.) ; identifier les facteurs qui influencent l'utilisation de l'évaluation et leurs mécanismes d'action (2.4.). L'approche méthodologique de ces quatre investigations est résumée ci-après. Leur description approfondie figure dans chacune des publications.

2.1. Etape 1 : analyser les évaluations régionales

L'obligation réglementaire faite aux régions d'évaluer leur plan de santé publique a permis de générer une connaissance des politiques de santé publique à l'échelle nationale. Ce matériau disponible couvrant un ensemble de régions a été exploité avec plusieurs visées.

Premièrement, l'analyse globale permettait de dresser un portrait national de la mise en œuvre des politiques de santé publique, de faire un bilan des forces et faiblesses de ce nouvel instrument de planification et d'apprécier les adaptations aux spécificités régionales.

Deuxièmement, cette analyse était l'occasion de mieux connaître les pratiques d'évaluation en France dans le domaine des politiques de santé.

Enfin, ce travail ouvrait la voie, à partir des similitudes et des différences, à des questions plus précises sur la façon dont les apports de l'évaluation pouvaient être considérés.

2.2. Etape 2 : analyser le processus d'évaluation et sa relation avec l'utilisation de ses conclusions

Nous avons fait un focus sur l'évaluation du plan régional de santé de la région Provence Alpes Côte d'Azur (PACA), que nous avons réalisée, et exploré les dimensions instrumentale, légitimatrice et conceptuelle de l'utilisation.

Pour cela, nous avons analysé les composantes du processus d'évaluation afin d'apprécier la capacité de l'évaluation à satisfaire les finalités managériale (générer du changement),

démocratique (mettre en débat l'action publique avec les citoyens) et scientifique (produire de la connaissance).

2.3. Etape 3 : identifier les résultats des évaluations dans neuf régions

Après la mise en place de la réforme instituant les agences régionales de santé, nous avons procédé à une analyse approfondie dans neuf régions afin de repérer la manière dont les enseignements tirés de l'évaluation des instruments de planification précédents avaient été exploités dans la construction des plans suivants et d'estimer le compromis réalisé entre les enseignements régionaux et les dispositions formulées par le niveau national.

En parallèle, nous avons tenté d'apprécier dans quelle mesure la connaissance produite par la somme des évaluations régionales avait été capitalisée au niveau national et retraduite dans les nouvelles orientations.

2.4. Etape 4 : étudier les facteurs d'influence de l'utilisation dans cinq régions

Cette dernière étape, réalisée avec un recul de quatre à cinq années après la fin des évaluations, prolonge l'étude précédente pour juger de leur utilisation à long terme et analyser les interactions entre les trois catégories de déterminants en utilisant le cadre théorique élaboré. L'étude porte sur cinq régions des neuf précédemment étudiées.

En dernier lieu, les résultats de ces travaux successifs ont été mis en perspective avec les données de la littérature.

Chapitre 5 : L'influence de l'évaluation sur les politiques régionales

Ce chapitre présente les résultats des travaux réalisés aux différentes étapes de la recherche ; chaque partie, correspondant à chacune de ces étapes, a donné lieu à la rédaction d'une publication.

La première partie qui présente l'analyse de seize rapports d'évaluation de PRSP, dépeint le contexte de mise en œuvre des PRSP et identifie leur contribution à la construction de la cohérence des politiques régionales de santé (chapitre d'ouvrage accepté, publication prévue en décembre 2014).

La deuxième partie, centrée sur le processus d'évaluation du PRSP de la région Provence Alpes Côte d'Azur, illustre la relation entre les finalités, la démarche et les utilisations de l'évaluation (publication).

La troisième partie, examinant les suites des évaluations dans neuf régions, cerne les formes d'utilisation à court et moyen terme, les principaux facteurs intervenant et les différences d'exploitation interrégionales (publication).

La quatrième partie, explorant l'ensemble des conditions influençant l'utilisation de l'évaluation, explicite la complexité des interactions entre l'évaluation, le milieu dans lequel elle se déroule et les acteurs impliqués, et souligne la singularité de chaque situation (article soumis).

1. Les enseignements des évaluations des PRSP sur l'évolution des politiques régionales de santé (publication acceptée)

La loi n°2004-806 du 9 août 2004 relative à la politique de santé publique créé de nouveaux instruments au service d'une plus grande cohérence de l'ensemble des actions de santé publique. Parallèlement, elle affirme très fortement le rôle de chef de file de l'État et réintroduit l'application d'un certain nombre de priorités nationales, affaiblissant, de ce fait, la prise en compte des spécificités et problèmes de santé régionaux. L'élaboration des plans régionaux de santé publique, dont le contenu est défini par décret, met les acteurs de l'Etat face à un double défi : faire converger les logiques de tous vers des objectifs de santé et de bien-être de la population tout en identifiant des priorités ; faire un plan adapté aux spécificités des territoires sans négliger pour autant les préoccupations nationales. Les groupements de santé publique associent les institutions engagées dans sa mise en œuvre mais pour autant, accordent une place prépondérante à l'Etat.

L'obligation réglementaire a créé une situation jusqu'alors inédite puisque quasiment toutes les régions françaises ont procédé à l'évaluation de leur plan. L'analyse de ces évaluations est éclairante sur le contexte de mise en œuvre de cette nouvelle planification, révèle les points forts et points faibles de ces nouveaux instruments et apprécie leur capacité à structurer la santé publique en région.

Regards sur la performance du système de santé régional en France

Françoise Jabot, Annie-Claude Marchand

In « *L'amélioration continue de la performance des systèmes et services de santé : un enjeu incontournable* », sous la direction de J.C. Suárez-Herrera, A.-P. Contandriopoulos, Z. Hartz et F. Cupertino, Presses de l'Université de Montréal & Conselho Nacional de Secretários de Saúde do Brazil. A paraître décembre 2014.

Dans de nombreux pays, la performance est devenue une question centrale pour la gouvernance publique. Elle s'est imposée en France comme mode de gestion, avec la réforme budgétaire du début des années 2000 qui prescrit à chaque administration la traduction de l'ensemble de son activité en un programme assorti d'objectifs et d'indicateurs afin de rendre compte annuellement des résultats obtenus au regard du budget alloué²¹. La performance est estimée sur la base de critères de qualité, d'efficacité socioéconomique et d'efficience. La notion de performance est, d'une part, intimement liée à la dépense budgétaire et, d'autre part, englobe l'ensemble des activités de contrôle de gestion et d'évaluation.

Le contexte

Performance et évaluation : des relations équivoques

Dans le secteur de la santé, l'évaluation a été introduite au sein de l'administration déconcentrée en 1999 avec la création de services régionaux qui avaient vocation à développer tant des activités d'inspection et de contrôle que d'évaluation. Plusieurs années après, force est de constater que la fonction d'évaluation n'a été que peu exercée par ces services, voire confondue avec les activités d'inspection et de contrôle. Le déploiement des programmes régionaux de santé de la dernière décennie a largement contribué à promouvoir la démarche d'évaluation ; cependant, cette dernière répond plus à des obligations réglementaires qu'à des initiatives propres. Quelques régions se sont engagées dans cette voie, en construisant une dynamique de concertation et d'accompagnement à la démarche évaluative mais, en 2004, seul un quart des programmes régionaux de santé avaient fait l'objet d'une évaluation (Jabot 2004). Les méthodes, les approches ainsi que les pratiques et les savoir-faire entre les régions sont très hétérogènes.

²¹ Loi organique du 1^{er} août 2001 relative aux lois de finances (LOLF)

Au niveau national, un bureau spécifiquement chargé de l'évaluation des programmes a été créé en 2000 à la direction générale de la santé du Ministère de la santé. Progressivement, les travaux d'évaluation de plans et programmes nationaux se sont multipliés avec des modalités institutionnelles de suivi et de pilotage témoignant d'une ouverture plus grande aux divers groupes d'intérêt et d'un travail plus collégial et coopératif entre les différentes catégories d'acteurs (Jabot et *al* 2010). Depuis 2004, le Haut conseil de santé publique, instance d'expertise du Ministère, a un mandat explicite d'évaluation des politiques de santé et dispose pour cela d'une commission spécialisée. Concourent également aux évaluations des politiques nationales, des institutions traditionnellement en charge du contrôle (Inspection générale des Affaires sociales, Cour des comptes) ainsi que certaines directions des administrations centrales. Par ailleurs, l'assurance maladie a fourni un cadre pour l'évaluation des réseaux de santé et encourage l'exercice pour quelques programmes spécifiques. Progressivement, des savoir-faire se sont construits ; il existe notamment un nombre non négligeable de guides, d'outils, de procédures, toutefois encore insuffisamment diffusés et partagés. Malgré ces avancées, l'évaluation reste une pratique fragmentée, peu reconnue, non programmée, dont les différents travaux, conçus et réalisés isolément, sont mal articulés, ce qui ne facilite pas leur exploitation à l'échelon régional ou national.

Tandis que les démarches d'évaluation se structurent dans les domaines du soin (évaluation des pratiques professionnelles, démarches qualité) et du médicosocial (évaluation des prestations des établissements médicosociaux) avec le soutien d'organismes experts tels que la Haute autorité de santé et l'Agence nationale d'évaluation sociale et médicosociale, elles peinent à se construire une légitimité dans le domaine de la prévention et de la promotion de la santé.

La logique de la performance introduite par la LOLF côtoie une culture évaluative encore récente et fragile. La place centrale de l'indicateur, variable emblématique de l'évaluation, et le caractère englobant de la mesure de performance ont créé de la confusion chez bon nombre d'acteurs au point de voir cet exercice se substituer parfois aux démarches d'évaluation engagées. Aujourd'hui, le concept de performance conserve des contours encore flous, souvent synonyme de contrôle de gestion ou d'efficience, entre reporting et évaluation ; ses relations avec l'évaluation restent ambiguës.

L'émergence du système de santé au niveau régional

Depuis une quinzaine d'années, un processus de réforme du système de santé a été engagé afin de concevoir et mettre en œuvre une politique de santé adaptée aux territoires. Le mouvement de régionalisation amorcé en France à partir des années 1970 avec la planification hospitalière s'est amplifié de façon notable ces dernières années. Progressivement, le cadre législatif et l'implication de plus en plus affirmée de l'ensemble des acteurs ont contribué à faire de l'échelon régional le lieu de la construction et du pilotage de la politique de santé. En 1996²², un ensemble de dispositions réglementaires ont permis la création d'instances régionales (agences régionales de l'hospitalisation, unions régionales des caisses d'assurance maladie notamment) et la mise en place de lieux de débat (conférences de santé) pour identifier des priorités régionales à partir desquelles construire des programmes de santé. En 1998, la loi de lutte contre les exclusions²³ prescrivait la mise en œuvre de programmes régionaux à destination des personnes en situation de précarité et créait des comités pour la concertation interinstitutionnelle et le pilotage des politiques de santé²⁴. En France, les lois de décentralisation, en opérant de nombreux transferts de compétence, ont permis aux collectivités locales, municipalités en premier lieu puis aux départements et aux régions, de se saisir de la question de la santé (Jabot et Loncle 2010). Les collectivités de niveau régional, bien que n'ayant à cette époque aucune compétence en santé, ont apporté leur concours en cofinçant les programmes de santé dans quelques régions. Les collectivités départementales notamment chargées des actions de protection de la mère et de l'enfant et d'insertion sociale ainsi que des vaccinations, étaient depuis longtemps associées aux actions de l'Etat. Enfin les villes, sollicitées par les orientations de la politique de la ville ou membre des réseaux Villes Santé OMS, développaient leur propre programme de santé. Progressivement, s'instaurait une nouvelle gouvernance fondée sur une recomposition des relations entre les partenaires et une gestion plus transversale de l'action publique.

Face à la dispersion des compétences institutionnelles, à la pluralité des niveaux d'intervention et des acteurs et au foisonnement de programmes, d'actions et de dispositifs, la loi votée en 2004²⁵ dotait les régions de processus formels pour construire une action publique globale et cohérente. Les plans régionaux de santé publique (PRSP) étaient conçus pour regrouper dans un même cadre l'ensemble des interventions de santé et les groupements

²² Ordonnance n° 96-345 du 24 avril 1996 relative à la maîtrise médicalisée des dépenses de soins

²³ Loi d'orientation du 29 juillet 1998 relative à la lutte contre les exclusions

²⁴ Décret n°98-1216 du 29 décembre 1998

²⁵ Loi n°2004-806 du 9 août 2004 relative à la politique de santé publique

régionaux de santé publique (GRSP), associant les institutions engagées dans le champ de la santé publique, avaient la mission d'en assurer la mise en œuvre et le suivi. En 2009, une étape de plus sera franchie avec la création d'agences régionales de santé (ARS) regroupant au sein d'une même entité les domaines de la prévention, du soin et du médicosocial²⁶.

En même temps qu'elle instituait la mise en place de plans régionaux de santé publique, la loi de 2004 prévoyait leur évaluation. Pour accompagner les acteurs dans cette démarche, un groupe de travail piloté par la direction générale de la santé, avait, avec la contribution de représentants régionaux, élaboré un document méthodologique (Ministère de la santé 2008). Fait inédit en France, toutes les régions ont procédé, entre 2007 et 2010, à l'évaluation de leur plan.

Sur la base de ces travaux, nous nous proposons d'apprécier la performance du système de santé régional en matière de santé publique, en utilisant le modèle d'évaluation globale et intégrée de la performance du système de soins (EGIPSS) et en l'adaptant au dispositif français déployé entre 2004 et 2010 puis nous discuterons de son application aux organisations de santé publique.

La méthode

Notre étude porte sur le dispositif PRSP/GRSP²⁷ dédié à l'élaboration et à la mise en œuvre de la politique régionale de prévention/promotion de la santé. Elle questionne les améliorations apportées par la nouvelle planification régionale en matière de gouvernance de l'action publique et de structuration de la politique de santé.

Le cadre d'analyse : le modèle EGIPSS

Le modèle EGIPSS, inspiré de la théorie de l'action sociale de Parsons, évalue la performance sur la base des quatre fonctions qu'un système doit assurer et sur les équilibres à maintenir entre ces fonctions pour assurer sa survie (Sicotte et al 1999, Champagne et Contandriopoulos 2009). La première fonction vise l'atteinte des buts, à savoir la capacité du système à remplir sa mission principale traduite en termes de valeur ajoutée sur la santé. La seconde concerne la capacité du système à s'adapter à son environnement pour assurer sa pérennité et se transformer en fonction des changements. La troisième porte sur la production en termes de biens et services et le processus de production. Enfin, la dernière fait référence aux valeurs et

²⁶ Loi n°2009-879 du 21 juillet 2009 portant l'hôpital et relative aux patients, à la santé et aux territoires

²⁷ Nous avons choisi ce terme désignant l'organisation conçue comme un tout, l'instrument de la politique ne pouvant se distinguer de sa gouvernance.

au climat qui sous-tendent le fonctionnement du système et lui confèrent sens et cohésion. Selon ce modèle, la performance d'une organisation s'apprécie à l'aune de son aptitude à régler et sauvegarder des équilibres entre ces quatre fonctions. Etudiant les interactions entre les fonctions deux à deux, six types d'équilibres sont proposés : tactique (entre atteinte des buts et production), stratégique (entre atteinte des buts et adaptation), légitimatif (entre atteinte des buts et maintien des valeurs), allocatif (entre production et adaptation), opérationnel (entre production et maintien des valeurs), contextuel (entre adaptation et maintien des valeurs).

Nous avons adopté le modèle EGIPSS pour notre étude principalement pour deux raisons. La première tient à la conception pluridimensionnelle de la performance sur laquelle il repose ; en effet, celle-ci intègre les notions d'efficacité, d'efficience, de rendement mais aussi d'accessibilité, de qualité et d'équité. Cette approche, plus globale que la vision française, plus volontiers centrée sur l'atteinte des résultats et l'efficience, rendra mieux compte des effets générés par la réforme en place. La seconde tient à sa capacité à évaluer des systèmes complexes, la performance étant vue comme le « résultat de compromis et tensions entre différentes dimensions » (Sicotte et *al* 1999) ; nous préciserons plus loin en quoi le dispositif PRSP/GRSP répond aux caractéristiques d'un système complexe.

Nous avons traduit les quatre fonctions décrites dans le modèle EGIPSS en variables pertinentes pour notre objet et identifié des indicateurs de réussite pour chacun des items (cf. tableau). Le but poursuivi par le dispositif PRSP/GRSP (F1) vise la mise en cohérence de l'ensemble de l'action publique en santé. La gouvernance du plan assuré par le GRSP concerne l'ensemble des procédures pour la mise en œuvre de ces plans (F2). La fonction d'adaptation (F3) renvoie à la flexibilité du dispositif pour répondre au mieux aux besoins et attentes des acteurs et des territoires. Les principes et valeurs sont traduits dans F4.

Le matériau de l'étude : 16 régions françaises

Les seize régions²⁸ pour lesquelles un rapport d'évaluation était disponible, constituent le matériau de cette étude. Les évaluations ont été achevées en 2008 pour 4 régions, en 2009 pour 7 régions et en 2010 pour 5 régions, sachant que l'installation officielle des ARS était le 1^{er} avril 2010.

²⁸ Aquitaine, Bretagne, Centre, Champagne Ardenne, Corse, Franche Comté, Ile de France, Languedoc Roussillon, Limousin, Midi Pyrénées, Pays de Loire, Poitou Charente, Provence Alpes Côte d'Azur, Picardie, Réunion, Rhône Alpes

Le travail repose sur l'analyse des documents relatifs aux plans et à leur évaluation, d'entretiens conduits avec une vingtaine d'acteurs clés au niveau régional ou infrarégional (responsables du plan et acteurs des programmes) et sur un examen plus approfondi de quatre évaluations dans lesquelles nous avons été personnellement impliquées, comme évaluateur, membre du groupe de suivi de l'évaluation ou aide à la maîtrise d'ouvrage.

L'analyse des données

Dans un premier temps, une grille d'analyse a été construite afin d'explorer pour chaque situation régionale : les caractéristiques régionales (géographie, démographie, économie, état de santé, historique des politiques de santé) ; le processus d'évaluation (finalités, enjeux, champ, questions, organisation, modalités de travail, suites données) ; les modalités d'élaboration du plan (concertation, structuration, nature des priorités) sa mise en œuvre (procédures, territorialisation, ressources allouées, relations entre les groupes d'acteurs, suivi) ; les effets sur les pratiques des acteurs (compétences, collaborations) et le partenariat institutionnel et associatif. Ensuite, nous avons analysé ces données (issues des documents, des entretiens et de nos observations) au regard des fonctions attendues du système PRSP/GRSP et des interactions entre ces fonctions.

Le dispositif PRSP/GRSP : un système organisé d'action

La structure

La loi de santé publique a formalisé le cadre de la politique régionale de santé publique, en définissant la méthode d'élaboration et le contenu des plans. Le PRSP est l'instrument de la politique de santé d'une région ; il englobe un ensemble de programmes et d'actions au regard des priorités et des objectifs définis et permet de structurer une politique en opérant des agencements entre les différentes interventions du PRSP et avec celles qui sont conduites en dehors de ce cadre. Par ailleurs, il favorise la mise en synergie des différents acteurs.

Le GRSP est un groupement d'intérêt public composé de représentants institutionnels prévus par la loi (Etat, organismes de protection sociale, agences de l'hospitalisation et, à titre facultatif, les collectivités territoriales). Il est géré par un conseil d'administration présidé par le Préfet et dispose d'un comité technique, le comité des programmes. Il peut constituer une entité identifiée, disposant de locaux distincts et d'une équipe, ou exister à travers un réseau d'acteurs demeurant chacun dans leur institution d'appartenance. Il est chargé de la mise en

œuvre des programmes du plan et pour cela il dispose d'un budget alimenté pour 85% de crédits de l'Etat et pour 25% de crédits de l'assurance maladie.

Les acteurs

Les acteurs sont nombreux : pluri institutionnels et pluri catégoriels (administratifs, responsables des programmes, référents territoriaux), engagés à des niveaux différents (équipe rapprochée, réseau interne élargi, partenaires externes des autres secteurs) et investis de responsabilités différentes.

Le processus

Le déploiement du PRSP se traduit par la mise en œuvre de multiples interventions d'envergure variable et requiert des formes d'organisation incarnées par des procédures et des mécanismes de coordination et de régulation.

L'appel à projets, généralisé en France depuis quelques années, est l'outil de mise en œuvre du PRSP. Il repose sur la sollicitation des opérateurs, à un rythme annuel, aux fins de développer des actions répondant aux objectifs des programmes, conformément aux orientations données en termes de ciblage des populations et/ou de territoires. Chaque promoteur financé par une ou plusieurs institutions publiques du GRSP doit traduire son activité en un ensemble d'actions répondant aux objectifs des différents programmes et les soumettre dans le cadre de l'appel à projets.

La finalité du système

Le dispositif a été instauré afin de structurer la politique de santé publique au niveau régional et de rallier les collectivités territoriales dans ce mouvement. L'ensemble des partenaires ont à s'accorder sur la lecture des problèmes à résoudre dans la région, le bilan des réalisations et des ressources mobilisables, les priorités à retenir, les objectifs à fixer en vue de l'amélioration de la santé de la population et la façon d'y parvenir.

La performance du système de santé régional à l'épreuve du PRSP/GRSP

L'analyse comparative des travaux d'évaluations a permis d'apprécier l'évolution des politiques de santé publique en région avec la mise en place des plans régionaux de santé publique et de faire l'état des forces et faiblesses de cette nouvelle planification. Dans la perspective de l'évaluation de la performance, nous allons d'abord évaluer l'accomplissement

des quatre fonctions d'un système, puis nous analyserons les tensions et la dynamique d'équilibre entre chacune d'elles.

L'analyse des quatre fonctions du système

Nous allons aborder successivement : l'atteinte des buts (mise en cohérence des interventions de santé et mobilisation des institutions), la fonction de production (déploiement des programmes et actions du plan), l'adaptation à l'environnement (adaptation aux besoins des populations dans les territoires et innovation), le maintien des valeurs et le climat (consensus sur les principes et cohésion organisationnelle).

- *L'atteinte des buts : des institutions regroupées au service d'une politique de santé structurée*

La construction et le maintien d'une cohérence d'ensemble entre les différentes interventions de santé, qu'elles soient ou non inscrites dans le PRSP, repose sur l'organisation des articulations entre les orientations issues des différents échelons, les différents programmes et l'organisation des services de soins, formalisée dans le document de planification (architecture du PRSP) puis opérationnalisée avec la mise en œuvre des actions (adéquation entre priorités identifiées et financements alloués, convergence entre PRSP et autres politiques).

L'élaboration des PRSP a été encadrée par un ensemble de textes fixant des règles ou des principes à respecter. Les acteurs régionaux devaient relever plusieurs défis : faire un plan sur-mesure, adapté aux spécificités des territoires tout en incluant les priorités nationales, intégrant les actions préexistantes et les priorités de chaque institution adhérente tout en ciblant quelques priorités. L'usage du mot priorité recouvre au moins deux acceptions ; il est autant employé dans son sens littéral, ce qui doit passer en premier, que pour désigner des problèmes ou besoins de santé. De ce fait, la représentation des priorités d'un plan n'est pas univoque : choix à opérer parmi les programmes du plan pour les uns, thématiques de santé abordées dans les programmes du PRSP, voire programmes eux-mêmes pour les autres. La nécessité d'intégrer tous les thèmes, le découpage à partir de portes d'entrée multiples (populationnelles, déterminants de santé, pathologies, stratégies d'intervention, programmes, volets obligatoires), la propension à créer des programmes pour chaque problème de santé jugé prioritaire, a généré une multiplication de programmes, induisant de multiples recoupements. L'architecture des plans, souvent découpée en thématiques multiples et

enchevêtrées a rendu sa lecture difficile pour les différents acteurs, notamment les opérateurs engagés dans des actions transversales. Afin de décloisonner les champs de la prévention/promotion de la santé et du soin, les PRSP ont souvent adopté les territoires de proximité des schémas régionaux d'organisation des soins (SROS) pour la mise en œuvre des programmes. Réciproquement, leurs objectifs prioritaires figurent dans les orientations des schémas. Cependant, au-delà des intentions affichées, les deux systèmes restent le plus souvent centrés sur les préoccupations propres à chaque secteur (santé publique versus sécurité sanitaire) à l'exception de situations favorables où une seule et même personne assure le double pilotage des actions de santé publique (programme du PRSP) et de l'organisation des soins (volet du SROS) pour un même thème (addictions, santé mentale, cancer, etc.). Par ailleurs, le PRSP rassemble et organise des programmes hétérogènes au regard de plusieurs caractéristiques (volume et nature des actions, importance politique, rattachement à un plan national, approche plus ou moins transversale). Ces derniers seront plus ou moins aptes à créer des passerelles avec d'autres dispositifs ou politiques publiques.

La nouvelle organisation avait l'ambition de mobiliser les collectivités territoriales afin de les faire adhérer au groupement. Le bilan établi par la direction générale de la santé (Ministère de la santé 2007) après deux ans de mise en œuvre, faisait état de l'adhésion de plus de la moitié des conseils régionaux au GRSP, des deux tiers des conseils généraux et de la présence croissante des villes. Si les collectivités ont volontiers participé aux réflexions sur le diagnostic régional, elles se sont prudemment engagées dans le processus, leur implication dépendant tant de leur statut territorial que de l'histoire locale. Les collectivités participant au financement des actions ont conservé leur autonomie de gestion, à l'exception d'une seule collectivité régionale (sur 26) qui a versé une contribution directe au GRSP. Les collectivités locales, quant à elles, ont endossé plusieurs rôles. En tant que décideur, elles participaient aux délibérations au sein des groupes de travail. Parfois, elles ont apporté un financement complémentaire à des actions soutenues par le GRSP. Dans d'autres cas, elles se sont positionnées comme opérateur, mobilisant les budgets du PRSP comme des leviers pour développer et amplifier leurs politiques, au-delà de leur capacité financière.

Le bilan reste mitigé ; le dispositif en place a contribué à la structuration de la politique régionale de santé mais celle-ci n'en demeure pas moins perfectible. Malgré les imperfections relevées, les plans ont permis une meilleure lisibilité de l'offre, plus de transparence dans l'attribution des financements et ont généré des rapprochements avec d'autres politiques (éducation, environnement, travail).

- *La production de services : une gouvernance organisée à différents niveaux et régie par un ensemble de procédures*

La fonction de production s'apprécie dans la confrontation entre le cadre d'orientation du PRSP et les procédures (gestion de l'appel à projets), le mode de pilotage et d'animation du plan et de ses composantes, les relations entre les différentes catégories d'acteurs (interactions formelles et informelles, qualité de la communication).

Le processus d'appel à projets a été rendu systématique pour toutes les actions financées par une ou plusieurs institutions adhérant au GRSP. Ainsi, chaque opérateur associatif a dû trouver des voies d'entrée dans le plan et segmenter son activité en plusieurs projets ou actions pour les inscrire dans les programmes. En raison de la multiplicité et des zones de chevauchement entre les programmes, certains projets étaient soumis à plusieurs programmes engendrant un surcroît de travail de tri et de coordination. Ainsi, ce sont plusieurs centaines d'actions qui ont dû être analysées chaque année par les divers groupes de travail mis en place à l'échelon départemental et régional pour l'étude et la sélection. La gestion de l'appel à projets a constitué l'activité centrale de la mise en œuvre du PRSP, impliquant dans un continuum, l'équipe régionale du GRSP, les responsables des programmes ou des axes du PRSP, les équipes territoriales et les opérateurs pour l'étude et le suivi des dossiers. En dépit des procédures balisant les interventions de différents acteurs, chaque étape du parcours des dossiers était susceptible d'engendrer des ruptures, faute du respect des règles établies (critères d'analyse et de priorisation des dossiers, calendrier pour la transmission des documents) et du principe de subsidiarité entre les échelons mais aussi en raison d'un déficit de communication entre les échelons. Entre responsables régionaux de programmes et instructeurs territoriaux, des différences d'appréciation des projets étaient parfois notables, l'affrontement entre les logiques propres aux positionnements des acteurs dépassant la volonté de planification rationnelle. Les efforts pour introduire davantage de méthode et harmoniser des pratiques hétérogènes entre les territoires ont complexifié le fonctionnement, entraînant des répercussions sur la charge de travail des équipes régionales et territoriales et sur la lisibilité du processus par les opérateurs. Le poids des formalités administratives a dissuadé les associations de moindre envergure de poursuivre une demande de financements, d'autant que les équipes du GRSP n'étaient pas en mesure de les épauler, étant trop absorbées par la gestion administrative.

- *L'adaptation à l'environnement : territorialisation et innovation*

Les PRSP ont vocation à prendre en compte les spécificités locales afin d'apporter une réponse plus adaptée aux besoins de la population. Pour cela, la contribution des acteurs locaux devait être recherchée à chaque étape : diagnostic initial élaboré avec les partenaires locaux, étude des projets au niveau territorial, sélection des projets et répartition des budgets étayées sur une base territoriale, solutions adaptées aux besoins et caractéristiques des territoires et de leurs populations.

La nécessité d'une action publique territorialisée s'est progressivement imposée en France depuis les années 80. Cadre structurant de l'organisation des soins, le territoire est devenu l'élément de référence à partir duquel concevoir et organiser les politiques de santé (Coldery et Lucas-Gabrielli 2012). Ainsi, dans la majorité des régions, la territorialisation a figuré au rang des préoccupations si ce n'est des priorités. Pour autant, le concept reste flou et hétérogène recouvrant des acceptions ambiguës. La sémantique utilisée est fluctuante ; pour signifier la mise en œuvre d'une intervention au niveau local, on parle indifféremment de déclinaison territoriale ou de démarche territorialisée alors même que les deux expressions renvoient à des approches assez opposées. Les interprétations du concept sont multiples et la traduction dans les faits variable selon les régions. Si le territoire réfère toujours à un espace géographique, son périmètre est variable, selon qu'il s'agit du lieu de pouvoir politique (commune, département, région), d'une intervention liée à une organisation particulière (réseau de santé) ou d'un milieu de vie (quartier). Ainsi, l'animation territoriale, modalité emblématique de la territorialisation, évoque tour à tour le principe de l'étude des projets par les équipes institutionnelles du niveau départemental, incluant ou non, au gré des partenariats locaux, une effective collaboration avec les collectivités territoriales, ou des modalités de coordination avec tous les acteurs du territoire. Les responsables, dénommés référents, animateurs ou coordinateurs territoriaux de santé, sont tantôt des représentants institutionnels du GRSP (le plus souvent de l'Etat), tantôt des opérateurs ou encore des acteurs extérieurs engagés pour assurer la fonction, avec des missions plus ou moins bien définies. La conception de la territorialisation procède de différentes logiques : géographique (découpage fin pour l'organisation de l'offre de soins et de services), technique (prise en compte du niveau territorial à différentes étapes de programmation), démocratique (espace de concertation, rôle actif des collectivités, affectation d'une marge d'autonomie administrative et financière), éthique (équité d'accès et d'offre), de santé publique (lieu de mise en cohérence). Chaque région puisant dans ces différents registres, la « territorialisation » revêt

des configurations variées dans les PRSP, entre déclinaison des politiques régionales et construction de politiques locales.

Le PRSP devait encourager l'innovation sous toutes ses formes : nouvelles stratégies d'intervention, nouveaux modes de coopération entre acteurs, y compris ceux issus de secteurs connexes à la santé. Bien qu'affiché, le soutien à l'innovation s'est d'abord heurté aux contraintes budgétaires, l'obligation de reconduire la majorité des actions existantes ainsi que certains programmes (dépistage des cancers par exemple), laissant peu de marge pour financer des actions nouvelles. Enfin, la lisibilité réduite du PRSP et de ses procédures n'a pas facilité l'inscription de nouveaux acteurs, d'autant que les acteurs institutionnels n'avaient pas toujours la disponibilité requise pour promouvoir et accompagner des initiatives nouvelles.

- *Le maintien des valeurs et le climat organisationnel*

Les acteurs engagés dans le PRSP partagent *a priori* les mêmes valeurs de la santé publique, tout particulièrement équité, solidarité et intérêt général. En revanche, les stratégies d'implantation ne font pas autant l'unanimité et illustrent les divergences entre les visions des différentes catégories d'acteurs sur les enjeux et le rôle du PRSP.

Les divergences de doctrine sont présentes dès la conception du plan. L'approche transversale (populationnelle ou stratégique) s'affronte à l'approche par thématiques ou pathologies, relayant l'opposition entre les partisans de l'approche holistique et de promotion de la santé versus les acteurs de tradition fondamentalement biomédicale. La phase de mise en œuvre révèle également des différences quant aux relations instaurées entre institutionnels et opérateurs (contrôle versus accompagnement), aux choix stratégiques (prise de risque avec les actions innovantes, soutien d'opérateurs plus fragiles et/ou moins qualifiés versus pratiques de routinisation). La diversité des modes de mise en œuvre dans les territoires reflète des visions contrastées dans la répartition des responsabilités et le partage du pouvoir entre l'Etat et les autres institutions et entre les différents échelons (hiérarchie versus collégialité).

Du côté des opérateurs, le découpage de leur projet pensé comme une globalité, en une somme de morceaux de projet, représente une contrainte d'autant plus forte qu'il altère la cohérence et le sens de leur action. Malgré cela, ces acteurs reconnaissent le PRSP comme un document de référence. Dans certaines régions sont appréciées les opportunités de dialogue à partir de l'évaluation des actions. Les acteurs institutionnels de leur côté soulignent que le PRSP offre un cadre qui favorise l'harmonisation de certaines pratiques de gestion des

programmes de santé publique, renforce et valorise des pratiques déjà éprouvées avec la dynamique des programmes régionaux de santé, notamment de formalisation des partenariats.

Des équilibres en tension pour la gestion des paradoxes

L'analyse précédente des quatre fonctions a déjà relevé des tensions entre les fonctions et mis en évidence des paradoxes inhérents à la construction et à la mise en œuvre du PRSP. Nous allons examiner chacun des équilibres décrits dans le modèle.

Le dispositif PRSP/GRSP est fondé sur un ensemble de paradoxes, politiques et méthodologiques, qui mettent en tension la gouvernance du système (équilibre tactique). Alors que la responsabilité de l'Etat en matière de politique de santé est clairement affirmée, la participation des collectivités, jugée indispensable, est recherchée, mais pour autant, elle reste facultative. De fait, les institutions adhérentes conservent leur autonomie financière voire leurs propres orientations. Par ailleurs, les PRSP englobent tout un éventail d'actions et de programmes dont une partie se prolonge dans des dispositifs gérés par d'autres institutions sur lesquels le GRSP ne dispose pas de leviers. Ils sont conçus comme un cadre d'accueil permettant de tout intégrer mais au sein desquels des priorités doivent être dégagées. Le mode de sollicitation par appel à projets, destiné à susciter des initiatives nouvelles, se heurte à une double contrainte : reconduire annuellement des actions existantes, tout en favorisant des opérateurs et des projets nouveaux qui répondraient aux objectifs actualisés des programmes. Les dispositions (équilibre opérationnel) et les choix de planification opérés relayent ces contradictions (équilibre légitimatif) : architecture verticale versus lecture transversale et territoriale des problèmes ; priorités affichées versus thèmes des appels à projets des programmes ; recherche de cohérence régionale versus morcellement des projets locaux des opérateurs ; concurrence entre opérateurs induite par les appels à projets versus efforts de rapprochement amorcés les années précédentes ; soutien des grosses associations bien implantées versus renouvellement des opérateurs.

En contrepoint, les acteurs développent des stratégies en lien avec les valeurs de santé publique qu'ils défendent, afin d'introduire des modes d'action adaptés aux difficultés des territoires (équilibre contextuel) et permettre des ajustements malgré des cadres formels pas toujours propices au changement (équilibre stratégique). En effet, les PRSP font l'objet d'un arrêté préfectoral après avoir été approuvés par différentes instances de concertation et toute modification est soumise à ce même processus de validation. Pour les faire évoluer en évitant ces procédures lourdes, les acteurs ont mis en place des modes de régulation interne, par exemple en révisant annuellement les appels à projets ; mais ces actualisations régulières,

parce qu'elles donnent une impression d'instabilité, génèrent de l'incertitude et troublent les opérateurs et les acteurs institutionnels proches du terrain. Pour introduire de la transversalité dans des plans d'architecture généralement verticale, les responsables de programmes ont suscité des collaborations entre interventions du PRSP ou avec des programmes extérieurs. De la même façon, les équipes locales se sont efforcées d'inventer des formes adaptées aux spécificités du territoire. Pour résoudre le dilemme budgétaire (reconduire l'existant et financer des actions nouvelles), les acteurs ont souvent préféré la dispersion des financements à l'ensemble des opérateurs au ciblage des quelques projets, pour ne pas mettre en péril l'une ou l'autre des associations (équilibre allocatif). Enfin, dans plusieurs régions, pour pallier le manque de disponibilité des acteurs institutionnels, des fonds ont été alloués aux organismes régionaux de promotion de la santé afin d'assurer l'accompagnement des opérateurs (équilibre opérationnel).

Pertinence du modèle EGIPSS pour évaluer la performance du système régional de santé

Un modèle plutôt bien adapté à l'évaluation au regard de la complexité du dispositif

Le modèle EGIPSS permet l'évaluation de systèmes complexes (Dubois et Ouchene 2010) et, pour cette raison, il est adapté à notre objet. En effet, le PRSP répond aux caractéristiques des interventions complexes (Glouberman 2002, Rogers 2008, Stame 2004, Zimmerman et *al*, 2011). Il est constitué d'une multitude d'interventions, obligatoires ou issues des volontés locales, destinées à différents groupes de populations, mises en œuvre par une multitude d'acteurs issus de culture variées, opérant à différents niveaux et chargés de responsabilités différentes (régulation, décision, accompagnement, mise en œuvre). Il comprend des sous-systèmes (programmes ou territoires d'intervention) obéissant à des logiques propres (objectifs, populations, stratégies) avec des chaînes de causalités en interactions réciproques et des effets attendus dans des horizons temporels à court, moyen, et long terme. La gouvernance mobilise plusieurs niveaux : régional pour la décision, la régulation et l'allocation de ressources ; départemental et parfois local pour l'étude des interventions à subventionner, l'accompagnement des opérateurs et l'animation des acteurs et dispositifs. Le leadership de l'Etat, la diversité des points de vue sur les objectifs à atteindre ou les stratégies à déployer, la mise en concurrence entre les opérateurs sont à l'origine de relations voire de tensions entre les groupes d'acteurs. Enfin, les acteurs ont affronté quelques paradoxes : le

caractère facultatif d'inscription dans les PRSP pour les collectivités, la nécessité de faire à la fois du consensus et des priorités, l'obligation de maintenir l'existant et soutenir l'innovation à moyens constants.

S'agissant des évaluations réalisées, les cahiers des charges avaient été rédigés en s'alignant sur les repères proposés par le niveau national (Ministère de la santé 2008), à savoir une série de questions portant sur les différentes composantes d'une intervention. Le modèle EGIPSS invite à une lecture du système sur la base de quatre fonctions auxquelles peuvent être aisément reliées les quatre axes de questionnement des évaluations régionales françaises. La question relative à la capacité du PRSP à produire de la cohérence renvoie à l'atteinte des buts, celle relative à sa gouvernance interroge la fonction de production. La capacité d'adaptation figure dans les questions sur la flexibilité du PRSP ou sur la dynamique de territorialisation. Les valeurs et le climat sont sous-jacentes au questionnement sur les changements induits par le PRSP sur les acteurs. Bien que relevant d'une autre approche, nous constatons que les interrogations formulées par les acteurs régionaux rejoignent les points relatifs aux quatre fonctions du modèle, à l'exception de la problématique sur le maintien des valeurs et le climat organisationnel, plus approfondie dans le modèle EGIPSS.

Enfin, l'analyse sur les équilibres entre les fonctions constitue un apport intéressant dans la mesure où les tensions et les paradoxes sont abordés comme des faits consubstantiels au système et ouvrent des pistes de réflexion pour en améliorer le fonctionnement.

Difficultés et limites d'utilisation du modèle EGIPSS

Ce modèle a été conçu principalement pour analyser des organisations de soins et son utilisation pour évaluer la performance d'un système de santé, aux frontières plus perméables avec l'environnement, est moins aisée.

Les fonctions étant étroitement reliées, il est souvent difficile d'attribuer les constats à l'accomplissement de telle ou telle fonction. Les variables sont parfois communes à plusieurs fonctions, attendu notamment le caractère transversal de la fonction liée au maintien des valeurs. L'outil de la politique (le PRSP) est indissociable de sa gouvernance (le GRSP). L'architecture du plan conditionne la mise en œuvre et si la cohérence recherchée relève de l'atteinte des buts, elle dépend étroitement des modalités de déploiement des plans. Par ailleurs, nous l'avons montré, les valeurs et principes sont sous-jacents à l'élaboration des plans. Dans le même esprit, l'adaptation du système relève de la volonté des acteurs, de la

représentation de leur fonction et de leur positionnement dans le système. Ainsi, pour analyser les équilibres entre les fonctions, nous avons été confrontées à des dilemmes s'agissant de prendre position entre telle ou telle fonction.

Au-delà des tensions liées à l'accomplissement des fonctions, la mise en œuvre des PRSP met à jour la confrontation entre les logiques de plan, de programmes et de territoires, portées par les acteurs à leurs différents niveaux. La logique planificatrice vise la centralisation des procédures autour de l'appel à projets avec une animation et une gestion régionale des projets locaux et des dynamiques inter-programmes. Les responsables de programmes défendent l'approche thématique, identifient des critères de sélection conformément à leurs orientations et développent des modes d'animation spécifiques en fonction des caractéristiques du programme. Enfin, les acteurs des territoires obéissent à des logiques de protection/défense des territoires, dépendant de leur perception des potentialités et besoins du territoire, des ressources disponibles, des dynamiques créées entre les territoires et l'échelon régional. Les logiques de planification, qu'elles s'expriment dans le pilotage du plan ou de ses programmes, sont tendues vers les fonctions de production et d'atteinte des buts, tandis que les logiques de territoire répondent à la fonction d'adaptation du système. Au croisement de ces trois logiques, la fonction orientée vers le maintien des valeurs aboutit à un ensemble de convergences et d'oppositions grâce auquel le système va fonctionner en dépit de ses imperfections. Cependant, au-delà de la volonté des acteurs, les politiques demeurent contraintes par les cloisonnements inhérents au système de santé, structurels ou budgétaires, pour lesquels la marge de manœuvre est limitée.

Perspectives pour l'évaluation des futurs projets régionaux de santé

La loi du 21 juillet 2009, dans la continuité des réformes engagées, constitue une nouvelle étape dans la consolidation de l'échelon régional avec la création des agences régionales de santé chargées d'élaborer une politique de santé globale incarnée dans le projet régional de santé. Cette réorganisation poursuit plusieurs objectifs, améliorer la performance du système de santé afin d'augmenter l'espérance de vie en bonne santé, réduire les inégalités de santé et accroître l'efficacité et l'efficience du système de santé. Le projet régional de santé est l'instrument unique de la politique. Etayé sur un modèle global de santé, il vise, en cohérence avec le parcours de vie des personnes, le développement des coopérations transversales entre des secteurs historiquement cloisonnés (promotion/prévention, soins, prises en charge médico-

sociales). Il est conçu en plusieurs strates : un plan stratégique qui définit les priorités de santé de la région et les objectifs qui en découlent, des schémas d'organisation dans les trois secteurs et des programmes d'actions, des programmes territoriaux pour intégrer l'ensemble.

Trois défis sont à relever pour améliorer la performance du système de santé : restaurer de la transversalité malgré des segmentations résiduelles, réussir les liens avec les autres politiques malgré un recentrage sur les agences, atteindre les objectifs en termes de santé malgré un contexte de restrictions budgétaires. Les tensions sont, là encore, présentes entre les principes affichés (cohérence, réduction des inégalités de santé) et les exigences de production efficiente avec un impératif de réduction des dépenses. Les acteurs régionaux sont aujourd'hui engagés dans les travaux de conception de leur système d'évaluation, partagés entre devoir d'imputabilité et préoccupation d'amélioration de la qualité des services rendus. La clarification des enjeux à partir de l'examen des fonctions du système et de leurs interactions pourrait être un guide dans la réflexion.

Remerciements

Les auteures remercient Lauren Saunders, interne de santé publique au Centre hospitalier de Rennes pour sa contribution aux travaux préliminaires sur les premières évaluations ainsi que Nathalie Dubois, chercheure à la direction de la santé publique de Montréal et affiliée à l'ENAP pour la relecture de ce texte et ses commentaires pertinents.

Tableau 1 : traduction du modèle EGIPSS au système de santé régional français

Fonctions d'un système (Contandriopoulos et al, 2008)	Traduction PRSP/GRSP	Indicateurs de réussite
<p><i>F1 : atteinte des buts</i></p> <p>(capacité à réaliser sa mission, en termes d'amélioration de la santé)</p>	<ul style="list-style-type: none"> - mise en cohérence des différentes interventions de santé - mobilisation des institutions - mutualisation des ressources - amélioration de la santé de la population 	<ul style="list-style-type: none"> - logique architecturale, hiérarchie des priorités (cohérence interne) - articulations avec autres secteurs et autres politiques (cohérence externe) - adhésion des institutions au groupement de santé publique - participation aux activités du groupement - financement des projets par les institutions adhérentes - indicateurs de santé (mortalité prématurée évitable, mortalité par pathologies...)
<p><i>F2 : production avec productivité de biens et services de qualité</i></p> <p>(production de soins, continuité, coordination, collaboration des acteurs)</p>	<ul style="list-style-type: none"> - déploiement des programmes de prévention et promotion de la santé /priorités retenues - respect des bonnes pratiques - développement des partenariats entre acteurs - articulations avec le secteur du soin 	<ul style="list-style-type: none"> - nombre et nature des projets - qualité des projets (étayés sur données probantes, stratégies pertinentes, évaluation réalisée) - compétences des professionnels - collaborations formalisées (contrats, réseaux)
<p><i>F3 : adaptation à l'environnement</i></p> <p>(capacité d'attirer les clientèles, s'adapter aux besoins, de mobiliser, d'innover, d'acquérir et de contrôler les ressources)</p>	<ul style="list-style-type: none"> - meilleure réponse aux besoins - adaptation aux spécificités territoriales - mobilisation de tous les opérateurs concernés - soutien aux projets innovants - rééquilibrage des financements en faveur de la prévention 	<ul style="list-style-type: none"> - projets sur nouveaux axes - projets construits localement et financements adaptés - inscription de nouveaux acteurs (secteurs connexes à la santé) dans le PRSP - financements de projets nouveaux - marge de manœuvre financière dans les territoires
<p><i>F4 : maintien des valeurs et climat favorable à la réussite des trois autres fonctions</i></p> <p>(production de sens, de cohésion, d'organisation)</p> <p>(équité, liberté, efficience)</p>	<ul style="list-style-type: none"> - consensus sur les principes et les objectifs du PRSP - répartition équilibrée des responsabilités entre acteurs - équité entre les territoires et entre les groupes - reconnaissance du potentiel local - autonomie d'action - utilisation rationnelle des ressources 	<ul style="list-style-type: none"> - critères partagés au regard des options retenues - explicitation des missions - répartition des ressources selon les besoins des territoires et des groupes - prise en compte des analyses et initiatives locales - financements de projets en conformité avec les besoins

NB : Pour faciliter la compréhension du système étudié, nous abordons la fonction de production avant la fonction d'adaptation.

Références

- Champagne, F. et A.P. Contandriopoulos, 2009. « Eléments d'architecture des systèmes d'évaluation de la performance des services de santé », dans AP. Contandriopoulos AP, Z.Harz, M. Gerbier, A. Nguyen (dir.), *Santé et citoyenneté. Les expériences du Brésil et du Québec*. Les presses universitaires de Montréal, 251-274.
- Coldefy, M. et V. Lucas-Gabrielli, 2012. « Le territoire, un outil d'organisation des soins et des politiques de santé ? Évolution de 2003 à 2011 », *Questions d'économie de la santé*, IRDES, n°175, p. 1-8.
- Contandriopoulos, AP. et al, 2009. « L'évaluation de la performance au service d'un pilotage décentralisé du système de santé », *Actualités et dossier en santé publique. Dossier Evaluation en santé publique*, n°69, p.21-26.
- Dubois, N. et H. Ouchene, 2010. *Cadre de référence pour le développement conjoint (CSSS/DSP) d'un système d'appréciation de la performance du Plan d'action régional (PAR) et des Plans d'action locaux (PAL) en santé publique du territoire montréalais*. Agence de la santé et des services sociaux de Montréal, Direction de la santé publique, Québec.
- Glouberman S, Zimmerman B and Commission on the Future of Health Care in Canada (2002) *Complicated and complex systems what would successful reform of medicare look like?* [Saskatoon]: Commission on the Future of Health Care in Canada, Available from: http://epe.lac-bac.gc.ca/100/200/301/pco-bcp/commissions-ef/future_health_care-ef/discussion_paper-e/no08/8_e.pdf (accessed 25 May 2014).
- Jabot, F. et P. Loncle, 2010. « Politiques locales » dans D.Fassin et B. Hauray (dir.), *Santé publique : l'Etat des savoirs*. Paris, Éditions La Découverte, p.372-382.
- Jabot, F. et al., 2010. « Dispositifs d'évaluation et appropriation des résultats » dans F. Mouterde et S. Trosa (dir.), *Les nouvelles frontières de l'évaluation, 1989-2009, vingt ans d'évaluation des politiques publiques en France et demain ?* L'Harmattan, p. 217-230.
- Jabot, F., L. Saunders et E. Rush, 2009. « L'évaluation des plans régionaux de santé publique: quels enseignements sur les politiques régionales de santé ? », *Congrès de la Société française de santé publique*, Nantes 1-3 octobre 2009.
- Jabot, F. 2004. « L'évaluation des programmes régionaux de santé », *Actualités et dossier en santé publique*, n°46. p.48-55.

- Ministère de la Santé, de la Jeunesse et des Sports et de la Vie associative, direction générale de la santé, 2008. *Évaluation des plans régionaux de santé publique : quelques repères pour les acteurs*, Paris : non publié.
- Ministère de la Santé, de la Jeunesse et des Sports et de la Vie associative, direction générale de la santé, 2007. *Situation des GRSP : bilan au 31 juillet 2007*, non publié.
- Rogers, P. 2008. « Using Program Theory to Evaluate Complicated and Complex Aspects of Intervention », *Evaluation: the international journal of theory, research and practice*, 14, 1, p.29-48.
- Sicotte, C., F. Champagne et A.-P. Contandriopoulos, 1999. « La performance organisationnelle des organismes publics de santé », *Rupture*, 6, 1, p.34-46.
- Stame N., 2004. « Theory-Based Evaluation and Types of Complexity », *Evaluation*, 10, 1, p. 58-76.
- Zimmerman, B.J., N. Dubois et al, 2011. « En quoi la complexité a-t-elle des répercussions sur l'évaluation ? », *Revue canadienne d'évaluation de programme*, 26, 3, p.xiii-xx.

L'espace régional entre enjeux nationaux et territoriaux

Deux éléments ressortent principalement de cette analyse : le caractère commun des difficultés rencontrées par les acteurs régionaux, contraints pour la mise en œuvre des PRSP, par les règles et procédures impulsées par le niveau national ; la convergence des constats et recommandations des évaluations. Ces constats interrogent l'équilibre des forces entre l'influence des orientations nationales versus l'adaptation aux spécificités régionales. Le poids du cadrage national était-il contraignant au point d'effacer l'inscription des exercices d'évaluation dans une spécificité régionale ?

Des expériences antérieures ont montré que les contrastes interrégionaux, manifestes dans les configurations administratives ou géographiques, s'incarnent dans la mise en œuvre des politiques ou des programmes. Par exemple, la diffusion d'orientations communes à l'ensemble du territoire national²⁹ pour le déploiement des Programmes d'accès à la prévention et aux soins (PRAPS) n'avait pas empêché une traduction diversifiée selon les régions, tant de la mise en œuvre des programmes que de leur évaluation (Gerhart et Pelé, 2004). La situation propre au PRSP serait-elle différente au point de faire exception ?

L'exemple de l'évaluation menée en région PACA, relatée dans la section suivante, apporte des éléments de réponse à cette question, illustrant la manière dont l'évaluation peut être adaptée aux enjeux régionaux.

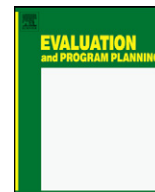
²⁹ Circulaire DGS/DAGPB/DAS/98/568 du 8 septembre 1998.

2. L'influence du contexte institutionnel sur le processus d'évaluation : l'exemple de l'évaluation du PRSP de la région PACA (publication)

L'évaluation du PRSP de la région Provence Alpes Côte d'Azur (PACA) est intéressante à plusieurs égards. Pionnière dans le lancement de l'évaluation du PRSP, elle a eu un effet d'entraînement sur les autres régions et a influencé le format de leur commande d'évaluation. Le responsable de cette évaluation, investi dans des réflexions nationales et régionales, notamment au sein de la SFE, a configuré le processus d'évaluation pour en faire un modèle le plus conforme possible aux standards de la SFE. Enfin, cette évaluation s'inscrit dans la tradition d'évaluation formative de la région.

Premier principe de la charte de la SFE (2006), référence de cette évaluation, le principe de pluralité énonce que « l'évaluation s'inscrit dans la triple logique du management public, de la démocratie et du débat scientifique ». La capacité de l'évaluation à s'aligner sur ces trois logiques fait écho au principe d'opportunité qui conditionne la décision d'évaluation à son potentiel d'utilisation.

L'examen de ce cas d'évaluation est mis à l'épreuve du modèle théorique que nous avons élaboré. Les interactions entre le contexte et l'évaluation elle-même sont étudiées, en observant tout particulièrement l'influence des procédures et valeurs du dispositif PRSP/GRSP sur le cadre, la gouvernance et les attributs de l'évaluation, et leur rôle en faveur de l'utilisation des résultats.



The evaluation of the PACA regional public health plan: Reconciling the managerial, scientific and democratic finalities

Françoise Jabot^{a,*}, Jean Turgeon^b, Léopold Carbonnel^c

^a Département Sciences humaines, sociales et des comportements de santé, Ecole des hautes études en santé publique, Avenue du Pr Léon Bernard, 35043 Rennes, France

^b Ecole nationale d'administration publique, 555 Boulevard Charest Est, Québec, Qc, Canada G1K9E5

^c Direction régionale de la cohésion sociale, de la jeunesse et des sports, Provence Alpes Côte d'Azur 23/25 Rue Borde-CS 90002 13285 Marseille Cedex 8, France

ARTICLE INFO

Article history:

Received 2 April 2010

Received in revised form 24 February 2011

Accepted 27 February 2011

Keywords:

Democracy

Debate

France

Management

Regional health planning

Pluralism

Evaluation process

ABSTRACT

For more than a decade now, evaluation has developed considerably in France, thanks in particular to the Société Française de l'Évaluation, whose charter sets out a number of principles designed to guide the work of evaluators.

This article examines how the evaluation process surrounding a regional public health plan (referred to as PRSP) – itself being a new instrument for regional planning in France – accords with one of these principles, which specifies that evaluation must be framed according to “a three-fold logic involving public management, democracy and scientific debate.”

Our analysis shows that while this evaluation was driven primarily by managerial concerns (i.e., assessing the capacity of the plan to structure health policy in a region), it also provided an Opportunity for restoring dialogue with a range of actors by opening up a space of cooperation and consultation. In addition, in order to ensure the legitimacy of the evaluation's conclusions, the knowledge produced by the evaluators had to rest on an irreproachable methodology.

This example illustrates how evaluation, in the French tradition, is a process that strives to reconcile the viewpoints and expectations of managers, scientists and the general public; it is also a process that brings out lines of tension and areas of complementarity between these three logics.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Modern social program evaluation emerged in the United States in the 1960s and in several English-speaking countries in the years following. On the other hand, public policy evaluation – a considerably much more loosely defined approach, given its objectives – was what, in France, mobilized the attention of the public authorities beginning in the mid 1980s. Then, from the early 2000s (decade), the budgetary reform of the French central government (RF, 2001) ushered in a logic of performance and results measurement, imposed the translation of public action into programs, and, in the process, threw into question the received conception of evaluation, with greater emphasis being placed on budgetary performance. Year 1999 saw the founding of the *Société française de l'évaluation* (French Evaluation Society, hereafter referred to as SFE), a decade after the International Research Group on Policy and Program Evaluation (INTEVAL), gathered around Ray Rist, first attempted to establish an evaluation society in France

(Jacob, 2005, p. 99). For several years now, the SFE has strived to promote quality evaluations and has set out in its charter a number of principles designed to guide the work of evaluators (SFE, 2006).

In the health sector, France's Act on public health policy (RF, 2004) requires of each region that it adopt a regional public health plan (*plan régional de santé publique*, hereafter referred to as PRSP) covering all public actions and programs and that it evaluate this plan. Such was the context in which, in 2008, the evaluation of the regional health plan of the Provence Alpes Côte d'Azur (PACA) region was conducted, constituting a first in France. This evaluation was, in consideration of the plan's recent implementation, considered to be a formative intermediate evaluation (RF, 2006) – in other words, its purpose was to “recapitulate the progress achieved to date in view of potentially reframing the action undertaken or the *référentiel*”¹ (RF, 2007).

¹ Loosely translated in English as “system of reference,” *référentiel* has been defined by the French political scientist Pierre Muller to mean a representation of reality serving as the basis for making a policy (Boussagué, Jacquot, and Ravinet, 2004, p. 371). A *référentiel* is both a cognitive process (simplifying reality in order to grasp it) and a prescriptive process (making it possible to act on reality) and, as such, can be viewed as a structure of meaning delineating the interconnections between different levels of perception.

* Corresponding author. Tel.: +33 02 99 02 28 42.

E-mail addresses: francoise.jabot@ehesp.fr (F. Jabot), Jean.turgeon@enap.ca (J. Turgeon), Leopold.carbonnel@drjcs.gouv.fr (L. Carbonnel).

The objectives of this article are first to describe and thereafter to analyze, from the perspective of the first principle set out in the SFE charter; the process involved in the evaluation of a new planning tool in France – namely, the regional public health plan (PRSP). This charter leverages both the evaluation lessons acquired in France and the methodological reflections developed in various countries to identify seven principles – including one pertaining to pluralism, whereby “evaluation is framed according to a three-fold logic involving public management, democracy and scientific debate” (SFE, 2006). These three logics in fact represent three forms of legitimacy. Working from the underlying legitimization processes outlined by the SFE to examine an evaluation process is in line with the suggestion of Tourmen (2009), for whom “Analyzing a situation with concepts related to legitimacy would be helpful to anticipate the kind of evaluation process and conclusions that may be accepted, acknowledged, and able to produce effects” (p. 24).

This article examines how the PRSP evaluation process in the PACA region accords with each of the three underlying logics. It also identifies the tensions and complementarities between them.

2. The context of evaluation

Evaluation has developed considerably in France for more than a decade, particularly in the health sector. A strong increase in health-related planning, particularly at the regional level, has constituted a lever contributing to the advancement of evaluation. Evaluations have been gradually initiated in the country’s regions, not only as part of the planning process but also in keeping with a dynamic of local consultation (Jabot, 2004). The conception of evaluation as a collective process and approach has spread alongside the model of expert evaluation as an independent process reserved exclusively for specialists. Thus, an abundance of research has been conducted at various levels (national, regional, local) in response to various regulatory requirements or local initiatives and covering a range of subjects (e.g., policies, plans, programs, organizations, networks, and pilot projects). All in all, the broad array of evaluations methods and approaches used to date have undergone little if any standardization, aside from a few situations in which evaluation mandates for evaluation have been accompanied by a fairly specific frame of reference.

2.1. The situation of evaluation in France

2.1.1. A few milestones

Notwithstanding a few prior initiatives, for the most part evaluation began to develop in France from the early 1990s on (Fontaine & Monnier, 2002, p. 63). At that time, France established an institutional evaluation apparatus – that is, a set of organizations and rules that “contribute to the development and sustainability of the evaluative practice in a given space” (Jacob, 2005, p. 107). At the national level, the adoption of a government order (RF, 1990), following a decade of debates, marked the official beginning of public policy evaluation. Drawing on the conclusions of a report (Viveret, 1989) drafted at the request of the government as part of its efforts to modernize the public service, the authors assigned an important role to organizing the confrontation of the viewpoints of all stakeholders; specifically, they established spaces of discussion and debate with the pluralistic evaluation bodies charged with conducting evaluations (Viveret, 1989). In addition, the 1990 government order both defined evaluation² and established an institutional framework consisting, on the one

hand, of an interministerial committee mandated to manage evaluation work and, on the other hand, a scientific council (Conseil Scientifique de l’Evaluation – CSE) charged with fostering reflection about evaluation methods and ethics. The work of this council (CSE, 1996) constitutes a theoretical reference that has since permeated France’s evaluation culture; indeed, among the community of evaluators there is a shared emphasis on such key values as the wording of evaluation questions, the plurality of viewpoints, the delineation of phases in an evaluation project, and the identification of main evaluation criteria (Perret, 2001). The CSE selected six criteria for evaluating policy quality: coherence, achievement of objectives, effectiveness, efficiency, impact and relevance. In 1998, the CSE was replaced by the *Conseil national de l’évaluation* (RF, 1998), which in turn was abandoned in 2003 and definitively abolished in 2008, coinciding with the implementation of the general review of public policies conducted alongside the reform of the central government (RF, 2008a).

In June 1996, the French Parliament established a public policy evaluation body answering to both the National Assembly and the Senate known as the *Office parlementaire d’évaluation des politiques publiques* (OPEPP, parliamentary public policy evaluation office) (RF, 1996). It seems that this experience proved to be inconclusive, as the OPEPP was abolished in 2001 (Jacob, 2005, p. 74). During the same period, the National Assembly founded the *Mission d’évaluation et de contrôle* (MEC, an evaluation and control delegation) in 1999 for the purpose of improving control over the use of moneys allocated under the appropriations bill. For its part, the Senate in 2000 created the *Comité d’évaluation des politiques publiques* (public policy evaluation committee) answering to the standing committee on finance and whose evaluations were conducted with the assistance of external experts (Jacob, 2005, p. 74). Fontaine and Monnier note that despite these initiatives, a degree of skepticism toward evaluation continues to linger at the political level (2002, p. 74).

At the regional level, evaluation developed rapidly during the 1990s as the result of a combination of actions: the 1982 law on decentralization, an increase in “State-Region Contractual Plans,” the implementation of European structural programs beginning in 1989, and the 1993 circular on partnership evaluation of co-financed policies (Fontaine & Monnier, 2002, p. 69).

At the dawn of the third millennium, Furubo, Rist, and Sandahl (2002), in an inventory of 22 countries, note that evaluation in France, “has not yet reached the full dimension of ‘maturity’ that can be expected from this type of exercise,” despite considerable internal and external pressures in this direction (p. 74). Derlien (cited by Jacob, 2005, p. 50) associates France with a wave of countries in which the institutionalization of evaluation is recent. One of the main reasons for this situation would appear to stem from how, in France, the relevance of government intervention comes in for little questioning, in contrast with the United States, where governments must provide a priori justification for intervening in any field involving the general interest (Perret, 2003). It would thus appear that in France, the performance of evaluations is less the outcome of a legislative requirement than is the case elsewhere (Jacob, 2005, p. 65). It remains to be seen whether this opinion will change as the result of enshrining evaluation in the Constitution in 2008 and making evaluation a major responsibility of Parliament, with the support of the *Cour des comptes* (Court of Audit). To this end, the National Assembly created a committee for the evaluation of control of public policies in July 2009. This committee may conduct evaluation programs and give its opinion concerning impact assessments accompanying draft legislation; in addition, it constitutes a forum for discussion across all the institutions making up the National Assembly.

In short, evaluation has, in France, been integrated into the “techniques and approaches making up the public management

² “The purpose of the evaluation of a public policy is to inquire into whether the legal, administrative or financial means implemented [indeed] serve to produce the effects expected from this policy and to achieve the objectives that have been assigned to it” (RF, 1990, our translation).

palette, more particularly in relation to knowledge of the conditions of public policy management” (Thoening, 2005). Aside from this culture emphasizing process evaluation and influenced by the sociology of knowledge, French evaluation practices stand out for the quality of sectoral expertise (especially in the corps of inspectors and ministry and agency research departments), serious reflection concerning evaluation methods and ethics, as well as an emphasis, ahead of most other countries, on the importance of confronting points of view at the time of evaluation (Barbier, 2010; Perret, 2003). From this perspective, the three-fold logic set out by the SFE in its Charter in 2006 appears to summarize (to a good extent at least) the evolution of evaluation in France. According to the SFE, the purpose of evaluation is “produce knowledge about public actions, respecting their effects, with the dual objective of enabling citizens to appreciate the value of these actions and assisting decision-makers in improving them” (SFE, 2006, our translation). Although the SFE does not claim to represent all the types of actors who perform evaluations in France, its Charter enjoys growing recognition, is used by university master’s programs in evaluation, and serves as a framework of reference for meta-evaluation research (Barbier, 2009, p. 31).

2.1.2. The regional framework of public health policies

During the last decade, the context surrounding the implementation of public health policies has been shaped by numerous influences, and in particular by the gradual recognition of the regional level as being relevant for the carrying out of health policies. Act 2004-806 pertaining to public health policy, adopted in August 2004, has helped to consolidate this level by laying the groundwork for the implementation of regional public health plans and groups. The *plan régional de santé publique* (PRSP), a plan built on the basis of a regional diagnostics performed by both institutions and professionals, is, as its name implies, the instrument of regional public health policy. Its overarching objective is to bring together, within the same framework and in the service of shared goals, health actions initiated by a number of institutions and across a range of decision-making levels. It must, on the one hand, create a roadmap for national priorities (e.g., the national cancer program or the national health and nutrition program) and, on the other hand, develop health programs and actions specifically designed to meet the needs of each region. Each plan is implemented by the recently created structure called the *groupement régional de santé publique* (GRSP), or regional public health group, which brings together the decentralized services of the central government, the regional hospitalization agency, the regional union of health insurance funds, and those local communities interested in participating. However, the Act prescribes that the central government remains the main driver of the GRSP, reserving a majority of votes on the board of directors to it. The Act also prescribes that these plans shall be required to undergo evaluation but does not, however, define the framework or methods to be used, leaving it up to regional actors – chief of which the regional representatives of the central government – to make choices in this regard. The French ministry of health has not developed a set of standards, properly speaking, but in 2009 it distributed a guide aimed at aiding regional actors to build their own evaluation (RF, 2008b).

2.2. The evaluation of the Provence Alpes Côte d’Azur region’ PRSP: a “first” in France

The Provence Alpes Côte d’Azur region

Covering 31,397 km², Provence Alpes Côte d’Azur (PACA) is a region located in southeastern France, presenting strong contrasts between a broad littoral fringe and Alpine mountain areas.

The region’s population (4.8 million inhabitants in 2007) and economic activities are concentrated in an urban area extending the length of the Rhône valley and the Mediterranean Sea. The region is divided into 6 administrative *départements*. In addition, respecting the offer of care and services, the region is divided into 9 “health territories” and 25 “community (services) territories.”

According to health indicators, the region fits with the national average. However, it is home to a higher proportion of persons suffering from chronic diseases owing to the higher demographic weight of its elderly inhabitants. “Medical density” is thus higher in this region than elsewhere on account of the attractive climatic conditions.

Source: Institut National d’Etudes Economiques et Statistiques (National Institute of Statistics and Economic Studies) <http://www.insee.fr/fr/regions/provence>.

The PACA region’s PRSP, established for a five-year period (2005-09), covers 23 programs targeted at the prevention of diseases presenting high regional prevalence (e.g., HIV/AIDS, hepatitides, and chronic diseases), the prevention of at-risk behaviours (e.g., addictions), or the delivery of care to specific populations (e.g., accessibility for elderly or economically disadvantaged people). The plan sets out more than 150 operational goals, and its implementation has mobilized more than 39 theme-based project leaders and 90 local correspondents, who work within a framework of institutional and professional cultures that can occasionally differ considerably from one another.

List of PACA PRSP programs

- | | |
|-----|--|
| P01 | Taking environmental action |
| P02 | Taking environmental action in the workplace |
| P03 | Reducing road risk |
| P04 | Reducing addictive behaviours and substance abuse |
| P05 | Promoting a balanced diet and exercise |
| P06 | Improving the prevention, early diagnosis and detection of cancer |
| P07 | Improving the detection of dental disease within the framework of oral hygiene promotion |
| P08 | Improving the detection of health problems among at-risk populations through medical exams |
| P09 | Facilitating access to adapted contraception in order to prevent unwanted pregnancies |
| P10 | Preventing everyday accidents |
| P11 | Promoting the appropriate use of medication |
| P12 | Improving vaccine coverage |
| P13 | Improving the detection, prevention and management of psychological suffering and suicide |
| P14 | Improving the prevention, detection and monitoring of HIV/AIDS/STIs |
| P15 | Improving the prevention, detection and overall management of hepatitis B and C |
| P16 | Program of health education for schools and outreach |
| P17 | Supporting parents in early child care |
| P18 | Preventing ageing-related risks |
| P19 | Implementing provisions relating to the Regional program of access to prevention and care |

P20	Organizing the handling of alerts and the management of emergency health situations
P21	Rare diseases
P22	Chronic diseases
P23	Female genital mutilations

This network of actors is a source of richness owing to the diversity of expertise brought into play. It is also a source of managerial complexity inasmuch as the director of the GRSP, in his capacity as the representative of the central administration, exercises hierarchical authority over only approximately 15% of them. Implementation of the plan has generated new methods and procedures for bringing consistency to work and highlighted the gap between the announced ambitions and the means made available to achieve these ends. Indeed, GRSPs constitute flexible structures, legally speaking, but are also weak and financially vulnerable structures in terms of their allocated resources and their political weight.

Two years after the plan first began to be implemented, the officer in charge of the regional public health policy decided to launch an evaluation process; in his view, evaluation, even if conducted early, would be a means of supporting change, re-framing (where required) the strategies or certain objectives of the plan, and legitimating the GRSP by showcasing its first outcomes. This evaluation, the first to be conducted in France in connection with a PRSP, accorded with the regional culture, in which evaluation is conceived of as a process involving the management and regulation of public actions and is clearly oriented toward a formative function.

3. Methodology

3.1. The conceptual framework

As was noted above, the charter of the SFE (2006) has identified three logics that are pursued by evaluation. Although these three logics complement one another, in our view they also seem to be mutually exclusive. They will serve here as a conceptual framework and subsequently help to guide the analytical approach.

3.1.1. The managerial logic

Public management “focuses on the organizational machinery for achieving policy goals” (Shafritz, 2004, p. 242). It can be defined as the art of carrying out public policies that produce as close a fit as possible between goals and resources (human, financial, organizational) in a dynamic sectoral or cross-sectoral environment.

According to this first type of logic, evaluation should be useful in terms of decision-making and the carrying out of public action. The instrumental character of the finality being pursued is what radically distinguishes this logic from pure research, thus generating not only the associated risks (i.e., the instrumentalization of knowledge) but also the potential benefits (i.e., the production of knowledge that is useful to public actors and society).

The expectations surrounding the evaluation of plan implementation vary. Thus, in the case of managers who draft statements of policy, the implementation of a public action represents the distance separating the policy statement from the public policy (Edwards & Sharkansky, 1981). From this perspective, the evaluation of policy implementation stands out as a mode of control over public action (Gibert, 2003) that should offer a basis for both measuring this distance and proposing ways of reducing it. For those whose daily work consists in translating the policy

statement into practice, “implementation may be viewed as a process of interaction between the setting of goals and actions geared to achieving them” (Pressman & Wildavsky, 1973, p. xv). In accordance with this point of view, “implementation appears as a stage on which the actors, who remain neither neutral nor passive, take action” (Radin, 1977, in Meny & Thoenig, 1989, p. 243). With this in mind, the evaluation of implementation should have the capacity to identify the difficulties encountered and, at the same time, propose solutions and engage a revision of policy goals so that the latter more fully correspond to the realities encountered in the field. These two expectations, which in principle stand in contradiction with each other, indeed represent one of the aspects of the principle of pluralism, put forward by the SFE.

3.1.2. The democratic logic

While the participatory logic of evaluation has enjoyed greater recognition in the United States of late (Perret, 2003), its advocates in France asserted themselves in the early 1990s in response to the partisans of a managerial approach to evaluation (Jacob & Varone, 2004, p. 138, citing Viveret). In France, public debate is recognized as a necessary phase in the decision-making process surrounding public policy in general (Callon, Lascoumes, and Barthe, 2001; Lascoumes, 2005; Phillips & Orsini, 2002; Torjman, 2009) and public health policy in particular (Abelson et al., 2002; Ghadi & Polton, 2000; Zmirou-Navier, 2006). There, various forms of public debate have been instituted, including regulated organizations for the inclusion of the “average citizen,” citizens’ juries or symposia, or, more recently, “Estates general” or forums at which all French citizens may engage in debate over major healthcare issues.

Respecting evaluation in particular, in consideration of how such assessment concerns public action, the democratic ethos requires the mobilization of all stakeholders in the formulation process. The goal is not to achieve a truth of an exclusively scientific nature but instead to develop a statement to enhance public debate. For Perret (2001), the point is not so much “to construct scientific facts as to [build] reasonably defensible arguments. For this reason, evaluation should combine scientific knowledge of various types with practical knowledge framed in terms of another form of validity.”

3.1.3. The scientific logic

According to this third logic, evaluation means, first and foremost, gaining scientific understanding of the production mechanisms and results of public action, and measuring the specific effects thereof with the mobilization of the research methods and techniques of data gathering and analysis associated with a range of disciplines, and with social science in particular. Accordingly, scientific debate represents a process of knowledge production and sharing that conforms to a set of rules specific to the field of science, particularly in relation to methods.

The scientific debate occurring in the field of evaluation presents some resemblances with the larger debate in the world of science, in which different ways of “doing science” occasionally clash with one another. Thus, for evaluators of a more positivist tendency, the science of evaluation is founded on three pillars (Dubois & Marceau, 2005, pp. 12–13): first the independence between the object of study and the evaluator that enables him or her to assign an objective value to the object using standards of merit (Scriven, 1980); second the models used as a basis of the evaluation and that guide the evaluator in his or her choice of dependent and independent variables; and third the improvement of the well-being of society or the community through measurements of the efficiency and effectiveness of public action.

An evaluator of the constructivist stripe recognizes ontological and epistemological bases that stand quite at odds with the previous approach (Dubois & Marceau, 2005, p. 18). To begin with,

evaluation is unable to provide an objective account of reality since, as Guba and Lincoln (1989) have pointed out, this reality is multifaceted, in keeping with the view that each actor holds of it. The same authors also note that evaluation represents more the outcomes of a process of negotiation between the various actors involved than it does a judgment based on “objective” evidence on the part of the evaluator. Finally, the evaluator strives to develop an understanding of the program in accordance with the environment and context, and will thus opt for methodologies that are open-ended and loosely structured.

Evaluation approaches do not, however, neatly fit according to the positivist-constructivist dichotomy. There are indeed other approaches that can be drawn on. Realistic evaluation, in particular, attempts to achieve a sociological understanding of the real using a realist methodology that accords a more extensive role to theory and that requires that the evaluator refrain from considering evaluation as an exact science (Pawson & Tilley, 1997, pp. xii–xiii).

In short, members of the SFE subscribe to the organization’s charter, which states that their evaluations should “take into account, in a reasoned fashion, the various opposing interests and gather the diversity of relevant views concerning the evaluated action, whether these views originate among actors, experts or any other person concerned” (SFE, 2006, our translation).

3.2. The analytical framework

In relation to these three logics, the evaluation conducted in the PACA region brings forward several questions. Did the evaluation produce results that were useful for carrying out changes? In what ways did it create a basis for opening up the results of public action to public debate? Did the protocol that was implemented satisfy the desire to leverage knowledge for research purposes?

At this point, and in relation to each of the three logics, the evaluation process will be analyzed, with a focus on: the evaluative survey (i.e., the nature of the concerns expressed), the evaluation organization (e.g., type of organization, *modus operandi*, and relationships between actors), the conduct of the evaluation (e.g., sequence of phases, activities conducted, etc.), the evaluation strategy adopted (e.g., methodology and posture of evaluation team), and the outcomes of the evaluation (e.g., decisions, effects on actors, and appropriation of results).

4. Description and analysis of the evaluation process used with the PACA region’s PRSP

4.1. A list of questions focused primarily on managerial and technical concerns

As set out in the terms of reference, the list of 32 questions was framed according to three main themes covering, respectively: (1) the design and structure of the PACA regional health plan (e.g., diagnostics, plan architecture, and underlying logic of actions); (2) the conditions surrounding the conduct of the survey, not only in terms of procedures, processes, facilitation/leadership, follow-up and evaluation but also in terms of communication and the formalization of partnerships; and (3) the impacts of this new tool on institutional actors (e.g., degree of involvement in the implementation of the PRSP and partnerships), front-line actors (perceptions, attitudes, and modified practices), and health policy. In view of the formative nature of this evaluation, the effects of plan implementation on program beneficiaries and the regional population as a whole were not covered. All told, the terms of reference referred to 32 questions organized under nine different headings distributed unevenly between the three main themes, with 6 questions being associated with Theme 1 (design/structure), 12 with Theme 2 (survey conduct) and 14 with Theme

3 (impacts on actors) (Table 1). Furthermore, this distribution shows that these items are expressed according to registers that are primarily technical (e.g., quality of initial diagnostics, identification of an underlying logic of action, information and follow-up system) and processual (conditions surrounding plan development and implementation). Indeed, 17 questions can be associated with implementation processes (managerial logic), 4 with the conditions of consultation (democratic logic) and 11 with cognitive contributions (scientific logic).

4.2. An evaluation organization receptive to the confrontation of multiple points of view

The evaluation organization included: a self-contained, independent evaluation panel, assisted by a technical unit led by the GRSP (*Groupe régional de santé publique*)’s official representative, and an external provider for the actual conduct of the evaluation.

The evaluation panel received a mandate from the GRSP setting out its missions – namely, selecting the evaluation questions, validating the terms of reference, selecting an evaluation provider, validating the report, proposing its recommendations for improving plan implementation, and laying the groundwork for the next generation of PRSPs. As constituted by the GRSP, this panel numbered 17 members: four representatives from each of four stakeholders (decision-makers, health promotion/prevention/education project proponents (*promoteurs d’actions*) and health professionals, independent elderly care monitors (*personnes qualifiées*), and users) along with a chair hailing from an institution recognized for its independence and expertise and who resided outside the region.

This work organization made it possible to establish and maintain connections between, on the one hand, the places of decision (decision-maker), execution (health professionals, health project proponents) and analysis (evaluation provider, experts) regarding the health policy, and, on the other hand, the GRSP official representative, who provided interface between these three levels.

4.3. The conduct of evaluation, in accordance with the constraints of the plan implementation schedule

Evaluation took place in three phases over a nine-month period: launch of the evaluation process (July to October 2007), follow-up on provider’s work (November 2007 to May 2008), followed by validation of this research and the production of recommendations by the evaluation panel (June 2008).

The panel convened three times during the first phase, with meetings being dedicated to identifying the evaluation questions, validating the terms of reference and selecting the evaluation provider. The members of the panel provided considerable input during this first phase, thanks in large part to the facilitation methods used. Thus, at the first meeting, more than 40 evaluation questions emerged. The selection of the evaluation provider constituted an important phase. The winning team was chosen primarily on account of its knowledge of the context (management-mindedness), pluridisciplinarity (blend of knowledge), capacity for dialogue and proximity to actors in the field (strong interest in re-launching debate).

Subsequently, the evaluation panel held three work sessions with the selected evaluation provider, during which time they were able to flesh out the evaluation survey, validate the methodology (particularly respecting the selection of four territories for observation), and discuss the initial results and recommendations. At the final session, time was set aside for the panel to develop its own recommendations.

The conduct of this evaluation was determined by the constraints deriving from the plan implementation schedule, as

Table 1
Breakdown of evaluative questions according to the three logics.

Evaluative survey, as covered under 3 themes and 9 headings (H)	Managerial	Democratic	Scientific
Theme 1 – Design/structure of PRSP (6)			
R1. Diagnostics and priorities			
- Quality and homogeneous of the diagnostic method used			×
- Method of prioritization and relevance of priorities			×
R2. Internal and external consistency of structure and delineation			
- Interrelationships of national and regional themes			×
- Interrelationships between programs and other policies and local public health plans			×
R3. Logic of action			
- Do the action strategies of various programs appear explicitly in the objectives?			×
- Is this logic action homogeneous throughout programs?			×
Theme 2 – Implementation of PRSP (12)			
R4. Implementation and decision-making procedure			
- Launch process (call for projects)	×		
- Decision-making process (selection and prioritization)	×		
- Efficacy of mobilization of means	×		
R5. Facilitation of programs and territories			
- Facilitation and management of programs (cross-sectoral aspect)	×		
- Facilitation and management of territories (territorial inequalities)	×		
R6. Quality of discussion/debate processes			
- Communication tools	×		
- Method of ongoing input gathering	×	×	
- Method of mutual consultation			
R7. Follow-up			
- Qualitative assessment			×
- Relevance and reliability of indicators	×		×
- Quality of information system			×
- Use for decision-making purposes			
Theme 3 – Impacts on actors and partnerships (14)			
R8. On front-line actors			
- Appropriation, shared values	×	×	×
- Transparency, acceptability of plan	×		
- Impact on professional practices	×	×	
- Dissemination of culture of results	×		
- Sustainability of actions			
- Development of innovations and experiments	×		
- Identification of deadweight			
- Satisfaction/perceptions of actors			
R9. On institutional partners and public health policy			
- Real involvement of partners	×		
- Mobilization of budget allocations	×	×	×
- Effects on the development of local policies	×		
- Emergence of new forms of cooperation	×		
- Impact on the conditions of user participation			
- Improved understanding of effective preventive strategies			
Total (32)	17	4	11

the evaluation's recommendations were expected prior to pursuing the plan in the year following. The GRSP's official representative not only saw to compliance with timelines but also played a facilitating role in terms of ensuring access to data, holding meetings, and framing the evaluation within the context of action.

While, on the whole, the involvement of panel members was widespread at the onset of the process, disengagement gradually subsequently set in, with panel members harbouring questions over the pay-off of their involvement in the second phase of the work. Only three members of the group (out of 17) attended the final – and yet decisive – meeting dedicated to the presentation of results.

4.4. A flexible evaluation method based on a pluralistic approach

The method adopted by the evaluation provider was designed according to an iterative mode in which work analyzing documents alternated with field investigations so as to adapt to needs as much as possible. In particular, in-depth observations were conducted in four territories with a view to assessing the conditions of PRSP implementation at the local level and the way in which programs meshed with territorial dynamics; as well, these observations were designed to gather the opinions and statements of front-line actors concerning the PRSP as a policy instrument.

Respecting the evaluation of the PRSP, the position adopted was one of attentiveness and observation in as close proximity as possible to the local level of execution. It is true that there is a decidedly comprehensive aspect to the conditions surrounding the roll-out of activities, the restraining factors, and the constraints giving rise to the dysfunctions ultimately identified. That being said, the posture expected of evaluators was not one of involving themselves in closely supporting front-line actors so as to induce the latter to play a more active role in the plan's workings and procedures or to elicit their support for the evaluation organization. Even less so was it a question of negotiating the terms of the evaluation with these actors.

The evaluation team strived to respect the distance required to acquire or maintain the most neutral posture possible while preserving enough proximity in order to observe and grasp the reality experienced by the actors as well as the relevance of the latter's analyses for incorporation into proposals.

4.5. Outcomes of the evaluation: a debate over proposals coupled with the adoption of measures

Approximately one month prior to the date prescribed for remitting the final report, the director of the GRSP, with the

consent of the evaluation panel, took the initiative of organizing a seminar dedicated to the communication of results for the purpose of launching a debate over the evaluation's conclusions and enabling actors to express their views and put forward some proposals. This seminar brought together the main actors involved in the plan – i.e., more than 80 people – and gave rise to considerable discussion and debate that pinpointed – without, however, alleviating – the constraints stemming from the plan. All the same, numerous constructive proposals were put forward, thus affording a basis on which to build and quickly implement a plan of action.

5. Is it possible to conciliate the three logics?

Table 2 recapitulates the results of the analysis of the evaluation process in relation to the selected criteria (e.g., nature of the evaluative survey, organization and conduct of the evaluation, modus operandi, and outcomes) for each of the logics examined (i.e., managerial, democratic, scientific).

The foremost finality of this evaluation was clearly managerial in nature. With the implementation of regional plans, France's public health act conveyed the political intent to consolidate the regional level as well as to bring consistency to a whole set of actions and programs developed by numerous actors at various decision-making levels. The evaluative survey surrounding the PACA region's PRSP faithfully reflects the preoccupations of the director of the GRSP and his concern with assessing the capacity of this new instrument to address the challenges stemming from this legislation – namely, to rationalize public action in the health sector and to mobilize the support of the community of actors concerned, partners from other government sectors, local communities, and operators. The evaluation work was framed by a timeline designed to allow the results to be used in decision-making concerning the renewal of actions in the year following.

Finally, a further rationale for the order lay in managing the network of PRSP actors, which had become cumbersome as the result of an uneasy alliance between several different professional cultures. Focusing primarily on processes, this evaluation was designed to be a strategy for infusing actors with new dynamism. It provided actors at all levels with an opportunity to voice not only their perceptions, or claims, but also to put forward constructed analyses of potential use in the pursuit of the plan but that had not yet found a way through to the decision-making level.

The dual need to restore dialogue with the actors, on the one hand, and to strengthen the legitimacy of a fragile institution (i.e., the GRSP) by developing an appreciation of its work, on the other hand, triggered an openness to democratic debate, in anticipation of the consultation process prescribed by regulations within the framework of regional health conferences. This impetus is clearly perceptible from the preoccupations pertaining to the perceptions and relations occurring between the different protagonists, the mobilization of a pluralistic panel, the choice of a qualitative evaluation having an explanatory aim, and, above all, the priority accorded to transparency through the sharing the results of the evaluation and the launching of debate over its findings.

However, given the way that control of the evaluation alternated between the evaluation panel and the evaluation provider's team, this process must be viewed as having remained at considerable remove from what is termed a “democratic” evaluation – that is, a form of evaluation aimed at enhancing the potential for citizens' debate and action in the field of policymaking (Viveret, 1989, our translation). Instead, it fits more with a pluralistic approach – the preferred approach in France, moreover (Lascoumes & Setbon, 1996, p. 10). This approach openly challenges that vision giving primacy to methodological perfection and the monopoly of experts. As well, it responds to criteria that distinguish it from so-called scientific research. According to Rossi and Wright (1994), the basic function of the pluralistic approach is

Table 2
Components of the evaluation process in relation to the three logics.

	Managerial	Democratic	Scientific
Nature/object of evaluative questions	Conditions governing implementation and facilitation in relation to this new instrument Effective adherence of partners Knowledge of the evaluation apparatus' mode of operations	Identification of dysfunctions in order to restore dialogue with actors	Analysis of object of PRSP (construction features and ability to achieve consistency)
Organization of the evaluation	Organization facilitating interrelationships between different actors and networks (decision-making, action, expertise) GRSP official representative ensures compliance with terms of order Legitimacy of team	Pluralistic panel, owing to representatives from various categories of actors (including project proponents) Independent body	Chairing of panel entrusted to recognized evaluation expert Pluridisciplinary team originating in recognized institutions
Management	Facilitation of access to information sources Facilitation of debate Adhering to schedule Priority of elaborating useful proposals		Autonomy of evaluators for performing their work Compliance with methodological choices Control of the evaluation process by the evaluators
Methodology	Gradual adjustment of methodology to the evaluation order	Comprehensive and qualitative approach	Production of knowledge concerning the conditions of Plan implementation and functioning Appropriate distance for approaching reality while yet maintaining neutrality Qualitative methodology
Communication of results and management of outcomes	Reintegration of actors' viewpoints at process end Constructive proposals Decisions made in relation to recommendations	Presentation of results at regional health conference – but following decisions by Board of Directors. In-depth presentation of results, recommendations, and decisions at local meetings	Observations and conclusions explored in greater depth during seminar

to provide feedback to program stakeholders. Its value derives from its capacity to validate and explain the internal causality of a program. Under this approach, the evaluator maintains control over the evaluative process in situations in which he or she “must be sure to understand and take into account all elements relating to the program [and in which he or she] suspects the existence of certain conflicts of values between different program stakeholders” (Ridde, 2006, p. 5). This, in effect, appeared to be the case respecting the evaluation of the PACA region’s PRSP.

Finally, for want of rigorous impact measures that would have mobilized currently non-existent observation and data systems, the proponents of the evaluation were well aware that in order to generate public debate, organizational change and legitimacy, the knowledge produced via the evaluation had to rest on an irreproachable methodology, as the credibility of the entire process depended on it. A number of additional aspects – i.e., the considerable number of evaluation questions bearing on technical aspects, the quest for a specific, legitimate expertise, the respect shown for the independence of the evaluation team concerning its methodological choices, the production of in-depth knowledge about the conditions underlying plan implementation – all attest to the priority accorded to the production of well-argued knowledge and, rigorously elaborated judgment.

Is there a way of striking a balance between these three logics? Is tension not bound to arise between the approaches to which they refer? All three logics bring different actors into play, are framed by different temporalities and are governed by potentially contradictory approaches. For example, the managerial type of evaluation is closely linked to action and the institutional environment in which evaluation is carried out; because the objective is to produce knowledge that can be used to transform action, the timeline of evaluation has to mesh with that of action. The democratic logic tends to make evaluation into a space for debate in which a judgment is collectively negotiated on the basis of a multiplicity of points of view that, being voiced by a broad range of stakeholders, convey values and preferences that may not be all that homogeneous. It is in the long run that this type of evaluation fills its function of “collective learning process” (Perret, 2003), enabling all the stakeholders to appropriate the results not only of the evaluation but also of the action thus evaluated. In contrast, the post-positivist scientific logic brings evaluation into line with traditional research, with which it shares the requirement of creating a distance between the evaluator and the object of evaluation, or, a deduction-based framework of reasoning serving to produce a solid line of argument for decision-making purposes.

In the tradition developed in France, evaluation is an exercise that strives to reconcile these three logics either from a constructivist perspective – which holds that “[as] evaluation cannot provide an objective account of reality, an evaluator continually mediates the values, attitudes, beliefs and meanings that each person assigns to the object of study and strives to develop an understanding of the program based on the environment and the context” (Dubois & Marceau, 2005, p. 17, our translation) – or from a realistic perspective – whose objective is to function “as a decree forbidding evaluators from hiding behind those secret, scientific languages in delivering their verdicts (...) There is no universal ‘logic of evaluation’ no absolute ‘science of valuing’ (...)” (Pawson & Tilley, 1997, p. xiii). However, the sponsoring authority (client) remains the actor exerting the greatest influence on the evaluation process, as it is he or she who has control over the launching and closing phases of the evaluation. It is he or she who is ultimately responsible for the choice of evaluation questions, the creation (or not) of a structure for debate, and the type of control that he or she will exert over this structure. It is also he or she who establishes the operational

framework surrounding the evaluation (providing for more or less leeway in the timeframe), determines the amount of financial resources to be allocated, authorizes communication about the evaluation, and accepts or rejects to open the evaluation findings to debate. Thus, an evaluator who proposes a managerial, scientific or democratic type of evaluation may be confronted with a sponsor who does not share his or her vision and who may impose a type of evaluation that is more in keeping with sponsor preferences. That being said, even where the sponsor is animated by primarily managerial concerns (for example), this does not rule out integrating other dimensions whose importance becomes increasingly manifest over the course of the evaluation process, as was the case in the PACA region. As it so happened, following discussions with the evaluation team, who voiced their observations to him, the sponsor came to view an organized debate over the conclusions and recommendations (to be held at the end of the evaluation process) as a useful avenue for defusing a latent conflict with actors in the field.

6. The lessons of this evaluation

The lessons of this evaluation can be analyzed from three different angles.

For planners, the evaluation of the PRSP, according to the manner in which it was conducted, clearly provided a basis for reorienting certain managerial processes – by, for example, prompting the GRSP to make its priorities better known, relaunching the “territorialization” project (e.g., updating local diagnostics, appointing and training of territorial reference persons), seeking avenues for administrative simplification (i.e., reducing red tape), and fostering a more constructive dialogue with the local association sector. This evaluation also made it possible to raise questions over the logic underlying the construction of the PRSP; such reflection is useful from the perspective of the development of future regional health projects that will cover the entire area of responsibility (prevention, care, medical social service) assumed by the new regional health agencies created in 2009. Therefore there are grounds for considering that the evaluation thus performed will constitute a valuable reference for regional actors.

For the evaluation panel, its position of interface is a source of questions. Specifically, the panel produced a report that not only summarized its assessment of findings but also prioritized the recommendations of the external evaluation provider; so doing, it played a mediating role between the external provider and the main decision-maker, who was offended by a report that he deemed to be exceedingly harsh. Should the evaluation’s severe observations been tempered? Should greater effort have been dedicated to achieving balance between the evaluation questions? Ultimately, did the evaluation manage to maintain a balanced position between the various actors involved?

For the evaluators, confronted with a three-fold constraint (i.e., the requirement of producing an evaluation that have an irreproachable methodology and was, shared and useful), their participation in the last phase (which originally had not been planned) is also a source of questions. The problem, as it were, was to render the conclusions with the greatest objectivity possible while, at the same time, striving to avoid exacerbating tensions between front-line actors and decision-makers. The presence of these evaluators at the side of the decision-makers, coupled with the nuanced tone of the findings, prompted some front-line actors to raise doubts over the team’s neutrality. Did the involvement of the evaluators serve to facilitate the consultation process or, on the contrary, to disqualify their impartiality?

The three logics pursued in this formative evaluation of the PACA region’s PRSP offer a realistic picture of the efforts currently

being undertaken in the field of evaluation and planning with a view to reconciling the respective viewpoints and contributions of managers, scientists and the general public.

In practical terms, these efforts are apparent in a number of different ways: the quest for representativeness in the evaluation organization, the scope of the evaluation survey, the adaptation of evaluation methods to realities in the field, and an approach to the conduct of the evaluation in a manner conducive to dialogue between these three stakeholders, whose input is critical to any planning and evaluation process. In a recent article in this review, Gore and Wells (2009) argue “that the framing of many program evaluations tends to eschew any rigorous exploration of the governance aspects of program design and implementation.” By dedicating some effort to analyzing the role of various stakeholders in the governance of regional public health plan management, the PACA region’s evaluation of the PRSP met this need, at least in part.

Moreover, as Stame (2006, p. 7) has pointed out, evaluation should be a means through which governance is improved. The implicit object of the evaluation was, in fact, the governance of the PRSP and its associated modes of regulation, compromise-building, empowerment, and dissemination of a shared culture. From this perspective, program evaluation is an exercise whose value derives as much from the process proper, coupled with secondary effects (e.g., creation of a space for debate and mediation, construction of a collective reflection process), as from the quality of evaluation findings, as formally stated in a report.

However, Gore and Wells (2009) note that “despite this welcome interest in issues of governance, authors such as Stame and Hanberger, as well as Pollitt seem to accept any given set of governance arrangements as simply the contemporary form of societal coordination.” The evaluation of the PACA region’s PRSP, sociologically situated as it may be, and nevertheless affords some useful lessons to the community of evaluators and program managers owing to how it is founded on three logics whose validity is widely recognized – namely, science, democracy, and usefulness for government.

Acknowledgements

The authors wish to thank all the participants who have contributed to the evaluation for their time and willingness to share their experiences, as well as all relevant stakeholders for their involvement in the process. The authors would also like to thank the anonymous reviewers for their comments and suggestions on previous drafts of this article.

References

- Abelson, J., Forest, P.-G., Eyles, J., Smith, P., Martin, E., & Gauvin, F. P. (2002). Obtaining public input for health-systems decision-making: Past experiences and future prospects. *Canadian Public Administration Journal*, 45(1), 69–97.
- Barbier, J. C. (2009). La charte de la Société française de l'évaluation: des principes pour les participants à l'évaluation des politiques et programmes pblics. *Actualités et dossier en santé publique*, 69, 30–31.
- Barbier, J. C. (2010). Eléments pour une sociologie de l'évaluation des politiques publiques en France. *Revue française des affaires sociales*, 1–2, 27–49.
- Boussaguet, L., Jacquot, S., & Ravinet, P. (2004). *Dictionnaire des politiques publiques*. Paris: Presses de la fondation nationale des sciences politiques.
- Callon, M., Lascoumes, P., & Barthe, Y. (2001). *Agir dans un monde incertain. Essai sur la démocratie technique*. Paris: Le Seuil.
- CSE [Conseil scientifique de l'évaluation]. (1996). *Petit guide de l'évaluation de politiques publiques*. Paris: La Documentation française.
- Dubois, N., & Marceau, R. (2005). Un état des lieux théoriques de l'évaluation: une discipline à la remorque d'une révolution scientifique qui n'en finit pas. *Revue canadienne d'évaluation de programme*, 20(1), 1–36.
- Edwards, G. C., & Sharkansky, I. (1981). *Les politiques publiques: élaboration et mise en oeuvre*. Paris: Les éditions d'organisation. (pp. 227–251) (Originally published 1978 in English as *Policy predicament: Making and implementing public policy* (336 pp.). New York: W.H. Freeman & Co. Ltd.).
- Fontaine, C., & Monnier, E. (2002). Evaluation in France. In J.-E. Furubo, R. C. Rist, & R. Sandahl (Eds.), *The evaluation atlas: Seventeen country profiles* (pp. 63–75). New Brunswick, NJ: Transaction Publishers.
- Furubo, J.-E., Rist, R. C., & Sandahl, R. (Eds.). (2002). *International atlas of evaluation*. New Brunswick, NJ: Transaction Publishers.
- Ghadi, V., & Polton, D. (2000). Le marché ou le débat comme instruments de la démocratie. *Revue Française des Affaires Sociales* 2(April–June).
- Gibert, P. (2003). L'évaluation de politique: contrôle externe de la gestion publique? *Revue française de gestion*, 29(147), 259–274.
- Gore, T., & Wells, P. (2009). Governance and evaluation: The case of EU regional policy horizontal priorities. *Evaluation and Program Planning*, 32(2), 158–167.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Jabot, F. (2004). L'évaluation des programmes régionaux de santé. *ADSP, la revue du Haut Comité de santé publique*, 46, 48–55.
- Jacob, S. (2005). *Institutionnaliser l'évaluation des politiques publiques*. Bruxelles: P.I.E. - Peter Lang. Collection Cité européenne no. 35.
- Jacob, S., & Varone, F. (2004). Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970–2003). *Politiques et management public*, 22(2), 135–152.
- Lascoumes, P. (2005). *Expertise et action publique. Problèmes politiques et sociaux, Vol. 912*. Paris: La Documentation Française.
- Lascoumes, P., & Setbon, M. (1996). *L'évaluation pluraliste des politiques publiques Enjeux, pratiques, produits*. Paris: GAPP – CNRS Commissariat Général du Plan.
- Meny, Y., & Thoenig, J.-C. (1989). *Politiques publiques*. Paris: Presses universitaires de France.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage Publications.
- Perret, B. (2001). *L'évaluation des politiques publiques*. Paris: La découverte.
- Perret, B. (2003). L'évaluation des politiques publiques. *Informations sociales*, 110, 14–25.
- Phillips, S. D., & Orsini, M. (2002). *Mapping the links: citizen involvement in policy processes. Canadian Policy Research Networks discussion paper no. F 21*. Ottawa: CPRND.
- Pressman, J. L., & Wildavsky, A. (1973). *Implementation: How great expectations in Washington are dashed in Oakland*. Los Angeles: University of California Press.
- RF [République française]. (1990). *Décret no 98-82 du 22 janvier 1990 relatif à l'évaluation des politiques publiques*. France: Assemblée Nationale.
- RF [République française]. (1996). *Loi no 96-517 du 14 juin 1996 tendant à élargir les pouvoirs d'information du Parlement et à créer un Office parlementaire d'évaluation des politiques publiques* (p. 8911). France: Journal officiel.
- RF [République française]. (1998). *Décret no 98-1048 du 18 novembre 1998 relatif à l'évaluation des politiques publiques*. France: Assemblée Nationale.
- RF [République française]. (2001). *Loi organique no 2001-692 du 1^{er} août 2001 relative aux lois de finances*. France: Assemblée Nationale.
- RF [République française]. (2004). *Loi no 2004-806 relative à la politique de santé publique du 9 août 2004*. France: Assemblée Nationale.
- RF [République française]. (2006). *Centre d'Évaluation, de Documentation et d'Innovation Pédagogiques (CEDIP)* [en ligne]. http://www.cedip.developpement-durable.gouv.fr/article.php3?id_article=146. Last accessed on: February 17, 2010.
- RF [République française]. (2007). *Lettre de mission à l'instance d'évaluation*. DRASS de la région Paca, August 31, 2007
- RF [République française]. (2008a). *Décret n°2008-663 du 4 juillet 2008 relatif à l'évaluation des politiques publiques*. France: Assemblée Nationale.
- RF [République française]. (2008b). *Évaluation régionale des plans régionaux de santé publique. Quelques repères pour les acteurs*. France: Ministère de la santé et des sports.
- Ridde, V. (2006). Suggestions d'améliorations d'un cadre conceptuel de l'évaluation participative. *Revue canadienne d'évaluation de programme*, 21(2), 1–23.
- Rossi, P., & Wright, S. (1994). L'évaluation: Appréciation de la théorie, de la pratique et des aspects politiques. *Revue Politiques et Management Public*, 12(4), 93–135.
- Scriven, M. (1980). *The logic of evaluation*. Inverness, CA: Edgepress.
- SFE [Société française de l'évaluation]. (2006). *Charte de l'évaluation des politiques et des programmes publics – version actualisée le 20 juin 2006*. Paris.
- Shafritz, J. M. (2004). *The dictionary of public policy and administration*. Westview: Boulder, CO.
- Stame, N. (2006). Governance, democracy and evaluation. *Evaluation*, 12(1), 7–16.
- Thoenig, J.-C. (2005). L'évaluation: un cycle de vie à la française. In F. Lacasse & P. E. Verrier (Eds.), *Trente ans de réforme de l'Etat. Expériences françaises et étrangères. Stratégies et bilans* (pp. 117–127). Paris: Dunod.
- Torjman, S. (2009). *Community roles in policy*. Ottawa: Caledon Institute of Social Policy. p. 25.
- Tourmen, C. (2009). Evaluators' decision making: The relationship between theory, practice, and experience. *American Journal of Evaluation*, 30(1), 7–30 (Available for viewing at Sage Journals Online).
- Viveret, P. (1989). *L'évaluation des politiques et des actions publiques*. Paris: La documentation française.
- Zmirou-Navier, D. (2006). De la démocratie en expertise. Le cas des risques sanitaires environnementaux. *Santé publique*, 3, 483–500.

Françoise Jabot is a Medical Doctor and Public Health Expert and also holds a degree in Anthropology. She is currently a Senior Lecturer at the Department of human/social and health behaviour science at France's École des hautes études en santé publique (EHESP School of Public Health), where she provides training on health and evaluation policies and programs. She is also conducting research in both these fields, with special focuses on the structuring of health policies at the regional level, the territorialization

process, and health policy evaluation in France. As a member of the Société Française de l'Évaluation (French Evaluation Society), she is currently leading a taskforce on evaluation in the health sector.

Jean Turgeon, Ph.D., is full professor at the Ecole d'administration publique (ENAP, Québec City, Canada). He is co-director of the Groupe d'étude sur les politiques publiques et la santé (GEPPS) and the Centre de recherche et d'expertise en évaluation (CREXE), both at ENAP. His research is focused primarily on the organization and management of the health and social sector, the analysis of policies and programs in

Quebec, Canada and France, as well as the analysis of government actions having a potential impact on population health and welfare.

Léopold Carbonnel holds degrees in Sociology and Political Science. For eight years now, he has performed duties in the regional units of France's ministry of health and social affairs relating to the auditing and evaluation of health and social cohesion policies. In addition, he is an evaluation instructor and has contributed to the development of various guides on evaluation methodology. He is a member of the Société Française de l'Évaluation (French Evaluation Society).

Croisement des finalités d'évaluation et enjeux pluriels

Outre les tensions entre les trois types de finalités soutenues par la charte de la SFE, cette expérience illustre l'intrication entre ces finalités et montre que ces dernières, d'une part, entretiennent des rapports de dépendance mutuelle et, d'autre part, reflètent plusieurs types d'enjeux portés par des acteurs situés à des niveaux différents : national/législatif, régional/institutionnel, individuel/méthodologique.

Savoir si l'institution régionale a réussi le pari assigné au dispositif - restaurer de la cohérence et rallier les partenaires des collectivités - est une préoccupation cognitive qui intéresse autant le niveau national que régional.

En revanche, c'est grâce à l'approche adoptée - la recherche de dialogue avec les acteurs de terrain - que les objectifs managériaux seront atteints, les résultats de l'évaluation ayant permis de réviser les procédures de mise en œuvre du plan.

La dimension scientifique, située *a priori* sur le registre de la connaissance, soutient les objectifs managériaux, en donnant à l'évaluation la crédibilité nécessaire amenuisant la contestation des constats. Elle sert également de caution aux mesures à venir et de démonstration de l'intérêt de l'évaluation comme outil de décision, dès lors qu'elle est conforme aux normes de bonnes pratiques.

Grâce à la collusion des trois logiques, cette évaluation donne lieu à trois formes d'utilisation :

- instrumentale, en réponse à la finalité managériale : cette évaluation a permis de réajuster le PRSP ;
- conceptuelle, en réponse à la finalité scientifique : cette évaluation a construit de la connaissance sur l'action conduite et un argumentaire sur les dysfonctionnements du dispositif ;
- légitimative, en réponse à la finalité démocratique : cette évaluation a permis de remobiliser les acteurs et de restaurer une légitimité à une institution fragilisée.

Cette expérience, parce que c'était la première, a inspiré le mandat d'évaluation de plusieurs régions. Pour autant, le point suivant montre que ces applications prendront des formes différentes.

3. La prise en compte des résultats des évaluations dans un contexte réformateur : analyse dans neuf régions (publication)

Cet article analyse l'influence des évaluations des PRSP sur les politiques de santé publique au niveau régional dans la période qui a précédé et suivi la mise en place des agences régionales de santé.

Il s'agit d'apprécier l'utilisation dans ses dimensions instrumentale (mesures prises à l'issue de l'évaluation, traduction des recommandations) et conceptuelle (évolution des représentations, conceptions, approches, intégration de nouvelles connaissances, modification des pratiques des acteurs).

Par ailleurs, l'influence de cette dynamique sur le niveau national pour nourrir les réflexions dans le cadre de la préparation de la loi 2009 est également questionnée.

L'évaluation des plans de santé publique: un levier de changement pour les politiques régionales de santé en France?

Francoise Jabot¹ et Annie-Claude Marchand²

Résumé : Entre 2007 et 2010, toutes les régions françaises ont réalisé l'évaluation de leur plan de santé publique, au moment même où était engagée une réforme de la gouvernance du système de santé régional. L'objectif de cet article est d'analyser l'influence de cette dynamique d'évaluation sur les politiques de santé aux niveaux régional et national. L'analyse approfondie dans neuf régions montre des retombées à court terme, inégales selon les régions, en fonction notamment, du calendrier de réalisation et du processus d'évaluation mis en place. En revanche, les apprentissages qu'elles ont générés, ont renforcé l'expertise locale utile pour le nouvel exercice de planification. Ces travaux régionaux n'ont pas fait l'objet d'une exploitation au niveau national mais on ne peut exclure l'hypothèse qu'ils aient irrigué la réflexion au sein des groupes de travail œuvrant à la préparation de la réforme. (*Global Health Promotion*, 2014; 21 Supp. 1: 64–69).

Mots clés : évaluation, utilisation, influence, changement, planification, politiques de santé, programmes locaux et régionaux de santé

Introduction

En France, l'évaluation est un phénomène relativement récent et ne figure pas systématiquement dans le cycle des projets, programmes ou politiques de santé. Pourtant, avec le développement de la programmation en santé, les travaux se sont multipliés, en réponse à des obligations réglementaires ou des initiatives de pouvoirs publics. De plus en plus d'acteurs se sont lancés dans l'exercice; les administrations l'ont introduit dans leurs services, les législateurs dans les textes. Produire de la connaissance pour éclairer la décision, améliorer la gestion publique, rendre compte des résultats, engager un débat avec les citoyens : les motivations sont plurielles et répondent aux fonctions classiquement conférées à l'évaluation (1–3). Pour autant, l'évaluation est-elle un levier de changement pour les politiques régionales de santé ?

Le développement de l'évaluation a suivi le processus de structuration des politiques régionales de santé entamé depuis une vingtaine d'années. Dès

1996, les régions ont disposé d'un cadre pour débattre des questions de santé et promouvoir des programmes répondant aux priorités régionales. En 2004, la loi relative à la politique de santé (4) a consolidé cet échelon en le dotant des outils nécessaires à la conception et au pilotage des politiques, notamment, de plans régionaux de santé publique (Prsp¹) destinés à donner plus de cohérence aux différentes interventions de prévention/promotion de la santé. Enfin, un pas de plus a été franchi en 2009 avec la réforme du système de santé réunissant les trois domaines de la santé (prévention, soins, médicosocial) au sein d'agences régionales de santé (ARS) chargées d'élaborer une politique de santé globale incarnée dans le projet régional de santé (PRS) (5). Attendu que la loi leur en faisait obligation, toutes les régions ont procédé, entre 2007 et 2010, à l'évaluation de leur politique de santé publique, à travers leur Prsp. Ce mouvement à l'échelle nationale est un fait inédit dans l'histoire

1. Ecole des hautes études en santé publique, affiliée au laboratoire EA 4360, APEMAC, Université de Nancy 1, France.
2. Agence régionale de santé de Champagne Ardenne, France.

Correspondance à : Françoise Jabot, Ecole des hautes études en santé publique, Université de Nancy 1, Avenue du Professeur Léon Bernard, CS 74312 - 35043 Rennes cedex, France. Email : francoise.jabot@ehesp.fr

Global Health Promotion 1757-9759; 2014; Vol 21 Supp. 1: 64–69; 510535 Copyright © The Author(s) 2014, Reprints and permissions: <http://www.sagepub.co.uk/journalsPermissions.nav> DOI: 10.1177/1757975913510535 <http://ghp.sagepub.com>

Encadré 1. Analyse du changement.

-
- 1) Cibles du changement
 - Structure de l'instrument de la politique
 - Priorités : nature, méthode adoptée
 - Architecture du programme : approche, mode de construction
 - Gouvernance de la politique régionale
 - Modalités de mise en œuvre : lancement et gestion de l'appel à projet
 - Modalités de sélection : prescription/adaptation, continuité/innovation
 - Relations avec les opérateurs : communication, contrôle/accompagnement
 - Acteurs
 - Représentations, conceptions, approches adoptées
 - Pratiques : intégration de nouvelles connaissances, routinisation / adaptation
 - 2) Temporalité
 - Court terme : opérationnalisation des recommandations
 - Moyen terme : prise en compte des recommandations dans le plan stratégique
 - Long terme : culture des acteurs
 - 3) Interactions
 - Horizontales : dynamiques inter régionales entre responsables de la planification
 - Verticales : entre échelons national, régional, local
-

des politiques de santé et, d'autant plus particulier, qu'à cette période, le chantier de la réforme des ARS esquissait le cadrage de la nouvelle planification régionale.

L'objectif de cet article est d'analyser, dans ce contexte de rénovation de la gouvernance du système de santé régional, l'influence de cette dynamique d'évaluation sur les politiques de santé aux niveaux régional et national. Sachant que pour les trois-quarts des régions, les évaluations se sont déroulées après la parution de la loi instituant les ARS, nous nous sommes interrogées sur l'exploitation de ces travaux pour la révision de la politique au regard des dispositions impulsées par le niveau national. En contrepoint, nous avons tenté d'apprécier dans quelle mesure la connaissance produite par la somme des évaluations régionales avait été capitalisée au niveau national et retraduite dans les nouvelles orientations.

Comment apprécier le changement ?

L'évaluation comme moteur de changement est au cœur des préoccupations de la communauté des évaluateurs et la problématique de l'utilisation de l'évaluation a été largement explorée (6–9). Les travaux ont mis en exergue les multiples dimensions du concept d'utilisation. En effet, celui-ci ne peut être

seulement appréhendé à travers l'exploitation des résultats pour la prise de décision ; il convient également de considérer l'influence exercée sur les représentations, les schémas conceptuels, les pratiques et les compétences des individus et des organisations (10,11). Le processus d'évaluation lui-même, l'engagement des parties prenantes et le rôle des évaluateurs sont des déterminants du changement.

Inspirées par les travaux de Mark et Henry (11), nous avons construit un cadre théorique qui puisse rendre compte de l'influence de l'évaluation, tant sous l'angle de l'utilisation des travaux que de ses effets sur les individus et organisations. Notre analyse intègre trois dimensions : le potentiel de changement, la temporalité (court, moyen, long terme) dans lequel ce dernier s'inscrit, les dynamiques suscitées par l'évaluation (Encadré 1). Le changement a été étudié sur trois cibles : la structure de l'instrument de planification (le plan a-t-il été révisé ?) ; sa gouvernance (les modalités de mise en œuvre, d'animation et de pilotage du plan ont-elles été renouvelées ?) ; la culture des acteurs (les pratiques ont-elles été transformées ?). S'agissant des dynamiques, nous nous sommes intéressées aux interactions entre les responsables des Prsp des différentes régions et entre les échelons territoriaux. Enfin, nous avons confronté les préconisations des rapports d'évaluation aux orientations nationales afin

Encadré 2. Méthodologie de l'étude.
Echantillon

Sélection de régions parmi les 16 pour lesquelles un rapport d'évaluation était disponible

Critères de sélection

- Diversité des régions (caractéristiques sociodémographiques, nombre de départements, état de santé, temporalité de l'évaluation),
- Disponibilité des informations (documents, interlocuteurs)
- Implication des auteurs dans les travaux d'évaluation

Régions retenues (9)

- Aquitaine, Bretagne, Champagne Ardenne, Corse, Franche Comté, Ile de France, Provence Alpes Côte d'Azur, Réunion, Rhône Alpes

Recueil des données

- Documents : (régionaux) rapports d'évaluation, mesures post évaluation, supports de communication, documents d'étape dans la préparation du PRS, plan stratégique régional de santé; (nationaux) guides pour la construction du PRS
- Entretiens : acteurs clés impliqués dans la planification pré et post ARS des échelons régional (14) et infrarégional (5); responsables du pilotage national de la réforme des ARS (3)

Analyse des données

- Identification des mesures prises / recommandations des évaluations (court terme)
 - Traduction des recommandations dans les plans stratégiques
 - Confrontation des préconisations des évaluations aux orientations nationales influence des évaluations régionales sur la politique nationale
-

d'apprécier l'influence des apprentissages tirés du Prsp versus le cadrage national pour la construction du PRS. Le travail s'appuie, pour l'analyse du changement au niveau régional, sur les données recueillies dans neuf régions. Pour le second axe de questionnement (capitalisation des évaluations régionales au niveau national), nous avons collecté tous les rapports des évaluations régionales étudiés dans un précédent travail (12), les guides du Ministère chargé de la santé et conduit des entretiens avec les responsables successifs du pilotage de la réforme (Encadré 2).

Quelles retombées pour l'évaluation dans les régions ?

Les neuf régions de cette étude constituent un panel illustrant la diversité des territoires. Deux régions ont achevé leurs travaux en 2008, quatre à partir du second semestre 2009 et trois au premier semestre 2010, au moment même où étaient installées les ARS.ⁱⁱ

Compte tenu de l'existence d'un cadrage commun, l'analyse des neuf évaluations faisait état de constats partagés : des plans peu lisibles, du fait d'un nombre

élevé d'interventions et d'un déficit de priorisation, mais consensuels et générateurs de passerelles avec d'autres secteurs (éducation, environnement, travail) ; une gestion administrative lourde, au détriment de l'accompagnement des opérateurs, mais une amélioration de la qualification des professionnels ; des articulations insuffisantes entre prévention et soins et avec les autres politiques publiques ; des difficultés inhérentes au déséquilibre entre objectifs affichés et ressources allouées, mais des mesures d'ajustement local. Si la structuration de la politique de santé était jugée en progression, elle n'en demeurerait pas moins perfectible. Les recommandations étaient également convergentes (Tableau 1).

Les suites immédiates de l'évaluation

Malgré un consensus sur les défauts inhérents à leur architecture, les Prsp n'ont pas été révisés, vu la lourdeur des procédures à mettre en œuvre et la perspective de la réforme. Cependant, des mesures correctrices ont été adoptées : simplification des directives, appel à projets commun à plusieurs programmes, recentrage sur quelques priorités,

Tableau 1. Recommandations des 16 évaluations des Prsp et axes de la réforme des ARS.

<i>Recommandations des évaluations</i>	<i>Axes de la réforme</i>
Structure du Prsp (prévention/promotion)	Structure du PRS (prévention/soins/médicosocial)
– adopter une approche globale structurante	– 3 objectifs généraux
– recentrer sur quelques priorités	– 5 domaines prioritaires
– créer de la transversalité	– approche par les parcours de vie des personnes
Gouvernance	Gouvernance
– améliorer la concertation et la communication	– multiplication des instances de concertation
– clarifier les responsabilités entre les niveaux	– rôle du niveau départemental précisé
– réformer la mise en œuvre	– incitation à la commande publique
– développer la territorialisation	– guidance pour la mise en œuvre territoriale
– renforcer l'évaluation	– évaluation inscrite dans les textes
Acteurs	Acteurs
– renforcer la formation des professionnels sur la conduite de projet et l'évaluation	– accompagnement des opérateurs prévu
– soutenir les initiatives locales	– ateliers d'échanges de pratiques pour les responsables de la planification, production de documents

contrats de financements pluriannuels, formation des opérateurs. Parce qu'elle donnait l'occasion aux différents acteurs d'explicitier les enjeux et les difficultés, l'évaluation a parfois permis de désamorcer des conflits et de restaurer le dialogue entre institutionnels, responsables de programmes et opérateurs. En revanche, les effets d'apprentissage favorisés par les échanges réguliers au cours de l'évaluation sont essentiellement restés limités aux seuls participants des comités d'évaluation, ces derniers n'assurant que rarement le relais des informations dans leur réseau.

La comparaison des situations régionales révèle que les effets observés dépendent étroitement du processus d'évaluation, notamment de l'organisation en place, du degré de participation des acteurs et de la communication instaurée. Le calendrier a joué un rôle déterminant, la prise en compte des travaux étant d'autant moins envisageable que la fin de l'évaluation coïncidait avec la création des agences.

Les leçons tirées des évaluations pour le nouvel exercice de planification

Les principes majeurs de la réforme figuraient déjà dans les recommandations des rapports, y compris des plus anciens.

Adopter une approche globale plus structurante. Le modèle du PRS, instrument unique de la politique, résout structurellement le déficit de transversalité et de priorisation. Basé sur un modèle global de santé, il vise, en cohérence avec le parcours de vie des personnes, le développement des coopérations transversales entre des secteurs historiquement cloisonnés (promotion/prévention, soins, prises en charge médico-sociales) (13). Le plan stratégique, qui en est le socle, devrait générer cette transversalité dans la mesure où il définit « les priorités de santé de la région et les objectifs qui en découlent » (14).

Conscients des défauts du Prsp que l'évaluation avait pointés, les acteurs régionaux ont privilégié, au moment de la construction du plan stratégique, des approches plus globales (par déterminants, milieux ou populations, en référence à la promotion de la santé), mieux à même de contribuer à la réduction des inégalités de santé, de faire le lien avec les autres politiques publiques et de garantir une cohérence d'ensemble. Ainsi, l'approche fondée sur les parcours de vie suggérée par le niveau national, recueillera l'adhésion des acteurs et sera adoptée pour dessiner l'ossature du PRS.

Faire participer tous les acteurs à la construction de la politique. La concertation sur l'avancement du

PRS est encadrée par la loi qui a prévu le recueil des avis des instances à différentes étapes. Instruites des situations antérieures, et soucieuses de dépasser le caractère formel de l'exercice, certaines régions se sont efforcées de renouveler les modes de consultation et les inscrire dans un processus de coproduction au service d'un état des lieux véritablement partagé par les acteurs (forums de discussion, relecture des diagnostics par les professionnels).

Piloter la politique au plus près des territoires. Préoccupation partagée à tous les niveaux, la territorialisation de l'action publique était présente dans le questionnement évaluatif de la majorité des régions. Ainsi, des expériences originales en matière d'animation territoriale ont été documentées et diffusées lors d'échanges interrégionaux ou par l'intermédiaire d'évaluateurs œuvrant dans plusieurs régions. De son côté, la réforme a fourni des outils pour une mise en œuvre de proximité pour laquelle une réflexion était déjà bien engagée. Le redécoupage des territoires, exercice obligé pour les ARS, a été l'occasion de faire des espaces de concertation ou d'animation plus conformes aux caractéristiques locales que les évaluations avaient fait apparaître. Pour alléger les procédures et surtout, évoluer d'une logique d'offre à une logique de commande, des appels à projets ciblés en faveur de territoires ou de publics prioritaires avaient été lancés dans les suites des évaluations et ouvraient la voie à plus de démarches communautaires, d'expérimentations et d'actions innovantes que le cadre précédent avait rarement permis de soutenir.

Enfin, la dynamique créée lors l'élaboration d'un guide pour l'évaluation des Prsp (15) n'est pas étrangère à l'investissement des régions dans leur propre évaluation. Le débat national pour la validation du guide, l'accomplissement du même exercice par l'ensemble des régions et les échanges de pratiques qu'il a générés, sont autant de facteurs explicatifs de la qualité des contributions régionales à la réflexion pour les évaluations des futurs PRS.

L'évolution des cadres de référence des acteurs

La dynamique engendrée par le processus d'évaluation a suscité une remise en question des

schémas traditionnels et des pratiques de routine, l'adoption de nouveaux modèles (de la culture du contrôle à celle de l'évaluation) et l'évolution des rapports avec les opérateurs locaux. La nature des relations avec l'échelon national est aussi interrogée, entre pilotage par directives et accompagnement basé sur des échanges de pratiques à des fins de capitalisation des savoirs et de production collective. L'approbation d'une formule plus équilibrée dans le jeu des relations, en rupture avec le modèle pyramidal classique, ne gomme pas la culture normative bien ancrée dans les pratiques ; les acteurs s'interrogent sur la valeur prescriptive de documents dès lors qu'ils ne sont pas officiellement validés. La frontière entre contrôle et évaluation conserve aux yeux de quelques-uns une certaine porosité.

Les dynamiques entre les travaux produits à différents niveaux

Compte tenu de la réforme et de son cortège de prescriptions, l'évaluation des Prsp était-elle un exercice superflu ? À court terme, les évaluations ont eu des retombées inégales selon les régions. Les premières évaluations, plus volontiers participatives, s'inscrivaient dans une démarche de réajustement, alors que les plus tardives, motivées par une mise en conformité avec la loi, perdaient en pertinence du fait de la parution des directives nationales. Réussir la réforme implique de relever un double défi : surmonter les difficultés méthodologiques de la planification du PRS, structure hautement complexe, et garantir une mise en œuvre homogène de la réforme sur le territoire national. De multiples documents ainsi que des ateliers de travail encadrent la démarche et visent le partage de repères communs afin de faire progresser les régions au même rythme. Dans un contexte de profonds bouleversements (fusion de personnels de cultures différentes, restructuration des services, changements de locaux) et dans un calendrier contraignant, les éléments de continuité entre les deux périodes ont facilité l'exercice. En particulier, la présence de professionnels porteurs de la mémoire des travaux antérieurs et capables de mobiliser les savoirs acquis, a pu jouer un rôle clé. En effet, les expériences de planification accumulées ces dernières années sont autant de strates à la construction d'une expertise, que les récentes évaluations sont venues renforcer.

La réalisation des évaluations régionales et la préparation de la réforme ont constitué deux démarches parallèles, chacune poursuivant ses propres objectifs. Les travaux conduits en région n'ont pas fait l'objet d'une exploitation au niveau national. La préparation de la loi s'est appuyée sur une série de rapports d'experts et de groupes de travail, sur des débats régionaux, mais n'a pas tiré directement avantage des constats issus des évaluations régionales, en cours d'exécution pour la plupart d'entre elles. Pour autant, on ne peut exclure l'hypothèse que ces travaux aient irrigué la réflexion nationale, notamment grâce à la présence dans les groupes nationaux de personnes instruites des situations régionales. Enfin, les regroupements pour la construction du PRS ont été un lieu de brassage des savoirs et d'apprentissage, certains acteurs jouant le rôle de passeurs.

Des conditions propices au changement

Parce qu'elle est par essence, un processus de mise (ou remise) en question, qu'elle suscite la controverse, l'évaluation porte en elle les germes du changement. Un certain nombre d'éléments discutés dans ce travail en montrent la trace. Toutefois, ce changement ne s'accomplira pleinement que si quelques conditions préalables sont respectées : l'engagement des décideurs, la mise en place d'un cadre ouvert à l'ensemble des acteurs concernés, favorisant le dialogue et la transparence, ainsi qu'un calendrier de l'évaluation accordé avec celui de la décision.

Remerciements

Les auteurs tiennent à remercier particulièrement Bernard Basset, Laurent Chambaud, Stella Chopinet, Stéphanie Farge, Christine Ferron, Alain Lopez, Béatrice Luminet, Germain Madeline, Danièle Marquis, Claude Michaud, Maxime Pouvelle, Nicolas Prisse, Martine Valadie-Jeannel et Jean-Louis Wyart, sans qui ce travail n'aurait pu être réalisé.

Financement

Cette recherche n'a reçu aucun financement particulier des secteurs public, privé, ou non-lucratif.

Notes

- i. En raison de la similitude des acronymes, nous avons, pour faciliter la lecture, opté pour des minuscules pour

- l'outil de planification précédant la réforme (Prsp) et en majuscules pour l'outil promu par la réforme (PRS).
- ii. Les ARS ont été officiellement créées le 1^{er} avril 2010.

Références

1. Conseil scientifique de l'évaluation. Petit guide de l'évaluation de politiques publiques. Paris: La Documentation française; 1996.
2. Weiss CH. Evaluation: Methods for Studying Programs and Policies. 2nd ed. Upper Saddle River: Prentice Hall; 1998.
3. Hanberger A. The real functions of evaluation and response systems. *Evaluation*. 2011; 17(4): 327–349.
4. Loi n°2004–806 du 9 août 2004 relative à la politique de santé publique. *Journal officiel de la République française*, Paris, France.
5. Loi n°2009–879 du 21 juillet 2009 portant l'hôpital et relative aux patients, à la santé et aux territoires (1) *Journal officiel de la République française*, Paris, France.
6. Cousins JB, Leithwood KA. Current Empirical Research on Evaluation Utilization. *Review of Educational Research*. 1986; 56(3): 331–364.
7. Shulha LM, Cousins JB. Evaluation use: Theory research and practice since 1986. *Evaluation Practice*. 1997; 18: 195–208.
8. Leviton LC. Evaluation use: Advances, challenges and applications. *American Journal of Evaluation*. 2003; 24(4): 525–535.
9. Contandriopoulos D, Brousselle A. Evaluation models and evaluation use. *Evaluation*. 2012; 18(1): 61–77.
10. Kirkhart KE. Reconceptualizing evaluation use: an integrated theory of influence. In Caracelli VJ and Preskill H, eds. *The Expanding Scope of Evaluation Use. New Directions of Evaluation*. San Francisco: Jossey-Bass; 2000; 88.
11. Mark MM, Henry GT. The mechanisms and outcomes of evaluation influence. *Evaluation*. 2004; 10(1): 35–57.
12. Jabot F, Marchand AC. Regard sur la performance du système de santé régional en France. In : *L'amélioration continue de la performance des systèmes et services de santé : un enjeu incontournable, sous la direction de J.C. Suárez-Herrera, A.P. Contandriopoulos, Z. Hartz et F. Cupertino*. Brasil: Presses de l'Université de Montréal & Conselho Nacional de Secretários de Saúde do Brasil. A paraître.
13. Ministère de la santé et des sports, Ministère du travail, de la solidarité et de la fonction publique. *Elaboration du PRS: éléments méthodologiques*. France; 2010.
14. Ministère de la santé et des sports, Ministère du travail de la solidarité et de la fonction publique. *Plan Stratégique Régional de Santé: guide méthodologique*. France; octobre 2010.
15. Ministère de la Santé, de la Jeunesse et des Sports et de la Vie associative, direction générale de la santé. *Évaluation des plans régionaux de santé publique : quelques repères pour les acteurs*. France; 2008.

Enseignements et retombées des évaluations

L'étude présentée s'intéresse à l'utilisation à court/moyen terme de ces évaluations. Elle confirme, pour une utilisation instrumentale, l'importance de deux facteurs : l'engagement des décideurs ; l'adéquation entre le calendrier de l'évaluation et celui de la décision. Elle démontre par ailleurs, le rôle majeur du contexte de l'évaluation.

En accord avec les constats de Mark (2011), elle montre qu'un usage conceptuel peut se diffuser à d'autres sites que celui dans lequel l'évaluation se déroule. Si la diffusion des connaissances construites au fil de la mise en œuvre successive des évaluations en région ne fait aucun doute, leur exploitation au niveau national est moins certaine, même si un apport conceptuel ne peut être exclu.

Ce travail soulève plusieurs questions qui seront approfondies dans la section qui suit : la pertinence et l'impact d'évaluations de politiques dans un contexte de changement majeur affectant ces politiques ; l'ambiguïté du niveau national dans l'accompagnement des prescriptions législatives en matière d'évaluation ; l'articulation de travaux conduits à différents échelons et dans des perspectives pas strictement convergentes.

4. L'évaluation, un système d'action complexe: analyse des répercussions de l'évaluation dans cinq régions (article soumis)

Cette partie est dévolue à l'exploration de la nature complexe de l'évaluation. En effet, l'évaluation est bien plus qu'une intervention compliquée, elle constitue en elle-même, un système d'action complexe.

L'évaluation, une intervention compliquée

L'évaluation est une intervention compliquée parce qu'elle est découpée en une série d'activités à réaliser à différents temps du processus, par les principaux acteurs de l'évaluation.

L'élaboration de la commande, qui relève de la responsabilité du commanditaire, nécessite la mise à jour des enjeux et objectifs de l'évaluation ; c'est à cette occasion que s'expriment des discordances entre les représentations des acteurs relatives à la politique visée par l'évaluation.

Le temps dévolu aux travaux de collecte et d'analyse est de la responsabilité de l'évaluateur ; l'adoption d'un référentiel commun et l'aptitude de l'évaluateur à ajuster ses travaux et cadres de référence aux besoins de son commanditaire sont les enjeux cruciaux de cette étape.

Enfin, la restitution des résultats accompagnée ou non d'un processus de diffusion et de mise en débat représente l'étape finale sous la responsabilité du commanditaire.

Les frontières sont poreuses entre chacune de ces phases marquées par l'alternance des responsabilités.

L'évaluation, un système d'action complexe

Cependant, l'évaluation ne se limite pas à une série de tâches et travaux. Elle met en relation et fait interagir une diversité d'acteurs, porteurs d'intérêts, de représentations, de valeurs, de logiques et d'attentes variés, diversité potentiellement génératrice de désaccords voire de conflits. Par ailleurs, parce qu'elle s'inscrit dans la durée, l'évaluation est soumise aux aléas de l'environnement dans lequel elle se déroule. La survenue d'évènements sociaux, politiques,

économiques imprévus, l'évolution de la trajectoire ou de l'engagement des acteurs responsables, sont des facteurs qui fragilisent l'évolution de l'évaluation et rendent son devenir imprévisible.

Pour comprendre le phénomène de l'utilisation de l'évaluation, il est nécessaire d'appréhender l'évaluation en tant que système, dans une double perspective :

- holistique, c'est à dire comme un tout indissociable incluant les composantes de l'évaluation, les acteurs impliqués, l'environnement dans lequel elle survient et les effets qu'elle génère ;
- dynamique, en analysant les interactions entre le processus d'évaluation, les acteurs et le contexte, tout en tenant compte des contraintes liées aux situations paradoxales, aux divergences, à l'incertitude, à l'échelle de temps.

Nous avons appliqué le modèle théorique que nous avons construit pour analyser cinq évaluations de PRSP en tenant compte de cette double perspective.

Contributions and limits of evaluation in reviewing health policies: in-depth analysis in five French regions

Françoise Jabot^{1,2*}, Jean Turgeon³, François Alla^{1,4}

¹ EA 4360 Apemac, Faculté de médecine, Université de Lorraine, France

² EHESP Rennes, Sorbonne Paris Cité, France

³ Ecole nationale d'administration publique, Canada

⁴ Inserm, CIC-EC, Centre hospitalier universitaire, France

*Corresponding author

Francoise.Jabot@ehesp.fr

Département Sciences humaines, sociales et des comportements de santé, Ecole des hautes études en santé publique, Avenue du Pr Léon Bernard, 35043 Rennes, France

Abstract

This article is the second in a series of two papers. The first set out to apprehend evaluation as a complex intervention and presented an appropriate conceptual framework using, as explaining concepts, the context, the characteristics of individuals likely to make or cause use, and the evaluation itself, in order to analyse evaluation use in the decision-making process. This article presents an application of this framework to the evaluation of regional health policies in France. The research strategy combines qualitative research based on five case studies with a multicriteria analysis. Due to a major reform on health system, contextual factors are the most influent on evaluation use. Notwithstanding this, the motivation of high-ranking decision-makers, the major role of users and the emergence of a leader are critical points. Evaluation itself is less influent. Despite some limits, multicriteria approach provides a complementary interpretation and enriches qualitative analysis.

Key words : context, evaluation process, decision-makers, evaluation use, health policy, multicriteria analysis, users

The abundant literature on evaluation use bears witness to the interest of researchers in understanding the paradox of evaluation. If use is the *raison d'être* of evaluation, can an evaluation that is not used be described as such? (Hojlund, 2014). It is difficult to predict, much less guarantee the outcome of an evaluation. Indeed, the use of evaluation is subject to many different factors (Cousins and Leithwood, 1986; Johnson et al., 2009) and the long, entangled paths taken (Mark and Henry, 2004) do not always successfully lead to actual use.

This article is the second in a series of two papers. The first explored the issue of evaluation use and then set out to apprehend evaluation itself as a complex intervention (Shiell et al., 2008). It also presented an appropriate conceptual framework for this perspective in order to analyse evaluation use in the decision-making process. This article presents an application of this framework to the evaluation of regional health policies in France.

The regionalisation of health policies in France

Since the 1990s, a process of health system reform has been engaged to develop health policies in local territories (Castaing, 2012). The level of the administrative region has gradually become established as the locus for developing and conducting health policy. There have been a number of reforms during the course of this shift. Since 1996³⁰, French regions³¹ have been allowed to identify their own public health priorities and have had a framework for regulation of the healthcare offering. The decentralisation laws transferred a large number of powers, and in doing so have opened up the way for local government to take responsibility for the issue of health. The gradually increasing involvement of local government has altered the relationships between the State (still 'lead partner' in terms of health) and its institutional partners, establishing more cross-cutting governance of public action (Jabot and Loncle-Moriceau, 2010). A further step was taken with the adoption of a Public Health Policy Act in 2004³². Faced with fragmented institutional powers, multiple actors and a plethora of interventions (Chambaud, 2008), the law gave new instruments to the regional level – regional public health plan (*plan régional de santé publique*, PRSP), aimed at developing more coherent public health policy. This public health plan brought together all health prevention and promotion interventions. Regional public health group (*groupement régional*

³⁰ Order no 96-345 dated April 24, 1996 on the medical control of healthcare expenditure

³¹ France is currently divided into twenty-seven administrative regions (with between 300,000 and 12,000,000 inhabitants)

³² Law no. 2004-806 dated August 9, 2004 on public health policy

de santé publique, GRSP) was body bringing together the institutions engaged in public health under the joint leadership of the State and state health insurance representatives, with the remit of implementing and monitoring these plans. Five years later, a new law continued this drawing together, with the establishment of regional health agency (*agence régionale de santé*, ARS): this was to bring the fields of prevention, treatment and medico-social care together under a single authority for each region.³³ This new law enshrined a ‘long-term administrative strategy’ (Pierru, 2011) which had been widely accepted since the beginning of the decade.

The PRSPs were intended to be for five years, ending in 2008 or 2009. The act establishing them also provided for them to be evaluated before coming to an end. To assist regional stakeholders with these evaluations, France’s health minister, aided by regional representatives, drafted a guide and organised a national debate on the scope of these evaluations with a view to renewal of the PRSPs. All the regions in France proceeded with this exercise. Those that began late did so with the imminent reform looming; the GRSPs were to be discontinued and PRSPs replaced by regional health projects (*projet régional de santé*, PRS): new instruments for the overall health policy of a region. By October 2009, the ‘forerunners’ (the future heads of ARSs) had been appointed in the regions to organise the merger of staff originating from a range of statuses and cultures (Pierru, 2011), with the creation of ARSs scheduled for April 1, 2010. The upshot of all this was that the reorganisation of the health system raised the question of the *raison d’être* and outcomes of these evaluations.

This study relates to the analysis of five regional PRSP evaluations in France with a view to appraising those evaluations use in the review of regional health policy, the factors that influenced it, and the interactions between these factors. The notion of ‘use’ is here understood in the broad sense of the term, to include instrumental dimensions (concrete decisions), conceptual dimensions (progress in knowledge and representations, changes in practice) and symbolic and legitimising dimensions (justification of policy actions) (Weiss, 1998).

After briefly listing the components of the conceptual framework, we shall describe the method used to test the hypotheses, the forms of use identified, the factors they have been

³³ Law no. 2009-879 dated July 21, 2009 on hospitals, patients, health and territories

influenced by, and the dynamics between these factors, before turning to a discussion of the contributions and limits of this framework.

Method

The conceptual framework for analysis of the decision-making process

The suggested conceptual framework apprehends evaluation as a whole, and from a dynamic perspective: as an event that interacts with the components of the environment it is a part of (Jabot et al). Based on the literature, three categories of determinants of use have been selected to construct the model (figure 1): the context, which facilitates or impedes change; the characteristics of the individuals likely to make or cause use (users); and the evaluation, insofar as it produces knowledge that is useful, credible, and can be appropriated. These three categories are broken down into a number of variables. The context-based variables relate to the intervention, the institutional and organisational environment of the evaluation, and the contingent factors interfering with the decision. The user-related variables concern users' personal characteristics and their ability to influence and interact with other stakeholders. The variables relating to the evaluation activity itself concern the mandate, the evaluation process and the data produced. The evaluation, dealt with as a system, constantly involves many different interactions between the three groups of determinants. The socio-political and organisational context within the institution shapes the attributes of the evaluation, and models the representations, interests and expectations of users – and by extension, the practices of evaluators. In return, the evaluation alters the context and characteristics of decision-makers, who may in turn influence changes to the system. We have assumed that interactions may be envisaged between the three categories of determinants and that use or non-use of the evaluation may also have a retroactive effect on the determinants.

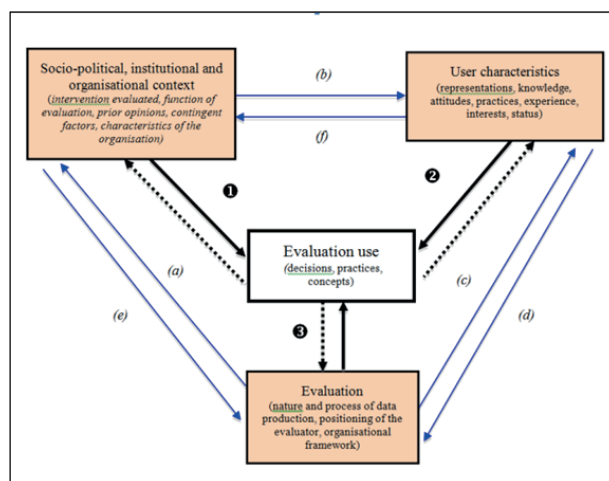


Figure 1. Conceptual framework

The research strategy combines qualitative research, based on multiple case studies with several levels of analysis (Yin, 2009), with a multicriteria analysis.

Qualitative research based on multiple case studies

The case study method is a research strategy that is appropriate for the purposes of exploring a phenomenon in its context and analysing the interactions of this phenomenon with a number of elements relevant to the research (Yin, 2009).

The cases here are the five regional evaluations. The cases were selected according to two criteria: the diversity of the regions (their size and their geographic, administrative and socio-economic characteristics), and the capacity for in-depth documentation of each case. For each of them, the authors were involved in the evaluation work, with different roles and at different times (before and after the ARSs were set up). These roles were: evaluator (one region), member of the monitoring committee (one region), support worker for the preparation of the regional health project evaluation (three regions). This special ‘observer-participant’ status (Malinowski, 1989) provided detailed knowledge of the context and facilitated understanding of the key issues.

Individual case study. Each evaluation constituted a unique case, the progress of which was studied along with the evaluation outcome, the factors influencing use, and their interactions, as well as the mechanisms underlying the decisions and the transformations observed.

A number of sources of information were used to document each case: collection of documents relating to the evaluation of PRSPs and regional health project planning,

individual interviews with potential users (managing directors of GRSPs and ARSs, heads of departments³⁴, PRSP and/or evaluation managers, and representatives of consultative forums). A total of fifty individual interviews were conducted (between 7 and 14 for each case). An interview guide was produced in order to explore the variables in our model. The interviews were recorded, transcribed, categorised and entered into a table based on the questions used: post-evaluation changes, the reasons given, evaluation perceptions (contributions, how the evaluation was perceived within the institution, and predictive factors as to the success of a given evaluation). Information collected at work meetings in three regions and during two training sessions on evaluation with all the individuals in charge of the PRS evaluation, comprised an additional source. An analysis grid was drawn up to apprehend the changes caused by the evaluation, explore the hypotheses for each category of determinants, and identify the mechanisms operating at each of the levels.

The evaluation use was assessed by : a) in the short term, the decisions that altered the implementation of the PRSP (review of strategies, procedures, relations with project bearers); b) in the medium term (following the establishment of ARSs), the drawing up and governance of the PRS; c) in the longer term (4 to 5 years after the evaluation), the extent of progress in the knowledge, representations, opinions and attitudes of users, transforming their practice.

The data was analysed using the model. A monograph was written for each region, offering an interpretation of the conditions influencing the evaluation use.

Comparative case study. A cross-cutting analysis of the five cases was conducted to compare the changes introduced by the evaluation and the influences of the three chosen categories of determinant. This comparison made it possible both to highlight the factors influencing use, on the basis of the accumulated data and the recurring elements and to select the most discriminating variables.

Multicriteria analysis. To deepen understanding of the interactions between the variables, a multicriteria analysis was adopted. Our use of ‘Multicriteria Decision Analysis’ (MCDA) methods involved a process of triangulation to retrospectively analyse the role of the variables in evaluation use in the different regions, as well as the relationships between the variables themselves. The results of this analysis were compared with the results of the case studies analysis.

³⁴ French regions do not all count the same number of ‘departments’: it varies between one and eight.

MCDAs are regularly used for decision support (Belton and Stewart, 2002) and to analyse complex situations involving information that is complex and conflicting in nature (Roué-LeGall et al., 2005). They also make it possible to consider a set of criteria of different natures – both quantitative and qualitative – within a single tool. There are a number of decision-related aspects: choice (selecting the best option); ranking (ranking actions from the best to the least good); and classification (classifying potential actions in predefined categories). Our approach falls into the latter type, since it involves differentiating evaluation situations on the basis of the potential for the evaluation to be used. We used Visual PROMETHEE³⁵ software, which is based on the PROMETHEE and GAIA methods developed by Mareschal et al. (1986). The PROMETHEE approach (Preference Ranking Organisation METHod for Enrichment Evaluations) proceeds by comparing pairs of actions: this results in a ranking of ‘actions’ according to the level of stakeholder preference, based on predetermined ‘criteria’. The ‘preference’ function represents the difference in performance between two actions for any given criterion, and may be applied to each. Each criterion may be weighted. The levels of preference are combined according to the weight of the different criteria to form a multicriteria preference ranking. In addition, the GAIA analysis (Geometrical Analysis for Interactive Aid) makes it possible to view the main characteristics of the decision-making problem, by showing conflicts and synergies between criteria.

In our study, the ‘actions’ are the ‘cases’ and the ‘criteria’ are the ‘variables’ that explain the evaluation use. The study did not seek to predict use but to observe as a whole how the variables explain the result observed in the regions, which use (or did not use) the evaluation work.

The method involved, firstly, filling in the table available in Visual PROMETHEE using empirical data. Each case was ‘assessed using a list of twenty variables. Each of them was scored in two ways: either on a five-point scale (very poor, poor, average, good, very good), or on a binary basis (yes/no). For the ‘preference’ aspect, the variable was qualified according to its ability to influence use. For example, we worked on the assumption that the more cohesion there was between the institutional partners engaged in regional policy (C1), the more likely it was that the evaluation would be used. Table 1 summarises the variables chosen, the scoring method descriptors, and the scoring system used. For qualitative data, the usual preference function was chosen; this means that there is no indifference threshold and that the highest score is the best. Initially, the same weight was assigned to all the variables.

³⁵ <http://www.promethee-gaia.net/indexF.html>

	Variables	Descriptors	Score
C1	Cohesion	Importance ascribed to the PRSP and partner consensus	5 pts
C2	Propensity to evaluate	Culture and dynamic of evaluation	5 pts
C3	External pressure	Influence of the consultative body (CRS)	3 pts
C4	Continuity of tier 1 decision-makers (directors)	None in ARS, 1 out of 2, both present	3 pts
C5	Continuity of tier 1 decision-makers (managers)	Dispersal, partial regrouping, most, head of department	5 pts
C6	Presence of a leader	Individual serving as interface, displaying strong leadership	Y/N
D1	Motivation of PRSP for evaluation	Interest and engagement	5 pts
D2	Favourable view of evaluation	Representations of usefulness, positive experiences	5 pts
D3	Agreement with findings	Congruence with representations	Y/N
M1	Motivation for PRSP evaluation	Interest and engagement, involvement	5 pts
M2	Favourable view of evaluation	Representations of usefulness, positive experiences	5 pts
M3	Agreement with finding	Congruence with representations	Y/N
M4	Collective competency	Presence of trained individuals and/or individuals with expertise	5 pts
E1	Envisaged goals	Explicit goals, use envisaged	Y/N
E2	Status of the evaluation committee chair	Internal or external with respect to the mandating institution	Y/N
E3	Working framework	Participatory, discussion at all stages, debate about findings	5 pts
E4	Consensus on evaluative questions	Development with stakeholders	5 pts
E5	Credibility of the evaluation	Credibility of the findings and of the evaluator	Y/N
E6	Timeliness of the evaluation	Prior to or after 2008	Y/N
E7	Dissemination of conclusions	None, report placed online, other media, debate	5 pts

Table 1. Matrix variables

PROMETHEE-GAIA delivers overall and relative analysis for each case. The PROMETHEE approach generates a ranking of cases, from the most preferred to the least preferred; the most preferred being the case deemed to be potentially the most inclined to use the evaluation. It supplies the profile of each case, showing the performance of each variable. The calculation of a stability interval indicates the limits within which the weight of the variable may be altered without any impact on the overall ranking. The GAIA approach allows the variables that work the most in favour of preference, the positioning of cases with respect to these variables, and similarities between cases to be identified.

Results

After a brief presentation of the five evaluations, the results of the case study analysis and MCDA analysis are set out below.

Summary presentation of the five evaluation cases

Case 1 (c1). The evaluation was launched to meet the regulatory requirement. In this small region, evaluation is not a routine activity. Resources were nonetheless dedicated to it: training the person in charge, official appointment of the evaluation committee members, and

recruitment of an external service provider. The committee met a number of times to validate the methodological choices; the evaluator, who was geographically some distance away, was rarely present. The report was submitted in November 2009 and presented to the Regional Health Conference (*Conférence régionale de santé, CRS*), a body bringing together all stakeholders that had to be consulted for any decision affecting the PRSP. The conclusions had little immediate impact, as the managers were taken up with preparations for the reform. However, its use was clear in the new planning of health prevention and promotion actions (*prévention/promotion de la santé, PPS*).

Case 2 (c2). The evaluation (which was the first of all regional evaluations with respect to PRSPs) was launched mid-term, to readjust the scheme. Evaluation culture is very pronounced in this region (one of the four largest regions in France), with plenty of events (trainings, seminars, meetings) and activities pursued within the institution, and was seen as a priority in the GRSP; there was also a professional in charge of deployment, as well as training of operational staff, and evaluations of interventions). A considerable budget was allocated to the evaluation; this allowed a team of researchers from two universities to be recruited. A pluralist committee, chaired by a recognised figure in this field, actively contributed to the procedure. Following the report (June 2008), which was deemed to be somewhat harsh by the mandating body, a regional meeting was organised in order to discuss the findings and recommendations. An action plan was drawn up on the basis of the evaluation recommendations and executed the following year. The findings of the evaluation fed into the subsequent planning and contributed to strengthening collective competency.

Case 3 (c3). This medium-sized region is characterised by a close institutional partnership for regional health policy development. Although it is not a predominant activity, the fact that evaluation is practiced is evidenced by evaluations of programmes and schemes, and more occasionally evaluations within departments. The evaluation was launched in the year prior to the reform, in order to comply with the law. The CRS wished to be actively involved, and lobbied for a working group to be set up that would be open to partners and independent of the mandating institution. The evaluator recruited conducted a participatory-based procedure that encouraged points of view to be exchanged. However, knowledge was not shared very much beyond this group. The evaluation report was submitted (June 2009) just prior to the appointment of the ‘forerunner’ and presented to the CRS at the end of the mandate. There

were no direct decisions as a result. Nevertheless, the lessons learned from the evaluation led to strategic thinking about the structuring of planning for PPS.

Case 4 (c4). Evaluation is a marginal activity in this small region. The evaluation was launched after two years of implementation, due to legal requirements as well as a desire to assess the deployment of the plan in local territories. Based on the experience of the first PRSP evaluation (c2) and the recommendations of the guide, it was conducted with the help of an external service provider, supported by a multi-institutional monitoring committee. A report was submitted well before the creation of the ARSs (December 2008), but not widely circulated. An action plan was drawn up but few measures were applied, except for those aimed at communicating about the actions in progress. Given the amount of time devoted to the process, the resulting benefits were deemed to be minimal by the main stakeholders involved. The evaluation was not used to support the new planning.

Case 5 (c5). The evaluation was launched late, at a time when the main decision-maker was engaged in reorganisation, made more difficult by this being an island territory. Evaluation was not performed routinely there and competencies were rare. To minimise the financial outlay, it was entrusted to the GRSP, which used a team of external providers. The latter conducted individual interviews and organised a group meeting to present the findings. The report was submitted in June 2009 during the ‘forerunner’ period. The proposals were not acted upon and the report was not circulated widely apart from a formal presentation to the CRS. The lessons learned were not used extensively, with national guidelines exerting the most influence in the establishment of the subsequent plan.

Influence of the evaluations on health policies

The ways in which the evaluations were used have been appraised on the basis of changes brought about in health policies, in the short term in the PRSP, in the medium term in the PRS, and in the long term in the knowledge and attitudes of stakeholders.

Changes in management independently of the reform. The evaluations completed in 2008 resulted in an action plan. In one case (c2), this was carried out almost in full, with a range of measures relating to implementation (refocusing actions on priority territories, simplifying

procedures, local governance, exchanges of practices between operators, institutional coordination, communication campaigns); in another (c4), it was carried out partially (internal and external communication actions, review of funding assignment procedures, reinforcing monitoring). The recommendations were not followed up on in c5, since the dissolution of the GRSP had been announced by the time the report was submitted. The evaluations which were completed in June and November 2009 respectively were not followed up on directly in regions c1/c3.

Changes in practice

Redesigning the structure of the PRS. Within a limited timeframe, the stakeholders attempted to organise wider (c1/c2), more participatory (c1) consultation, better coordinated with thinking at the level of the *département* (c2). Although mention was never made of the PRSP with regard to PRS strategic orientations, references are to be found in the specific planning document for the PPS (c1/c2/c3) and provide grounds for its structure; the problems noted previously led those drawing up the document to adopt a more population-based, cross-cutting approach.

An alternative approach to governance of action deployment. During the transitional year, the implementation of actions continued unchanged. The PPS stakeholders then consolidated the approaches that the evaluation recognised as being promising: territory-based approaches (c1/c2/c3); support for innovative actions (c2/c4); new contractualisation procedures with operators (c2/c4).

Anticipating the evaluation. The stakeholders in the five regions, engaged in strategic thinking about the evaluation of PRSs, listed three concerns in the light of previous experience: the establishment of stringent reporting, which had been lacking for PRSPs; the inclusion of questioning of the outcome, in response to the legitimate expectations of stakeholders; and earlier and more regular communication with consultative bodies.

Influence on representations

The individuals involved in the evaluation (38/50) said that they had gained knowledge that had transformed their practice and could be used in other settings: structuring

plans/programmes and organisation of departments; running project procedure; and for consultation issues with stakeholders. More specifically, with respect to the evaluation, they said that they had a better understanding of the evaluation activity and what was at stake, e.g. the link between evaluation and programming; the distinction between evaluation and control; the political and social nature of evaluation; the importance of the quality of the mandate; the organisation of a common framework of reference and work; engaging in advocacy with the director; the limits of process evaluation; the role of communication; the identification of relays to help spread change.

The determinants of decision-making

Before dealing with this point, it is essential to recall the elements of the national context forming the background to the implementation and evaluation of regional health policies.

The national context

Little recognition of evaluation within institutions. Evaluation within regional health administration formally began in 1999, with the creation of departments responsible for inspection, control and evaluation. However, these departments have only carried out evaluation very occasionally; more usually, the latter falls within the remit of the departments responsible for the interventions in question. The evaluation of programmes is usually entrusted to outside bodies. The practice of evaluation exists, but it is fragmented, poorly identified, undervalued, and uneven in terms of references and coordination. Competencies are rare and the related culture is present to very different degrees in different regions.

The onset of a reform. The recommendations by regional evaluations converged on three points: reducing partitioning between prevention, healthcare, and social care; coordinating policies locally in territories; and improving citizen participation (Jabot and Marchand, 2014). As things turned out, the provisions for the creation of ARSs went some way to addressing these recommendations. The new planning instrument for health policies explicitly defined the orientation of health policy, thus restoring the cross-cutting aspect which was otherwise lacking. In the future agencies, there were to be dedicated departments to provide leadership for territories. Moreover, there was an increase in the number of consultative bodies, in order

to facilitate better consideration of partners' points of view. The changes envisaged as a result of the evaluation, however, had to await the publication of national instructions.

The regional context

Cohesion between partners (C1, table 1). Whether high (c2/c3) or lower (c1/c4/c5), the degree of cohesion between institutional partners when supporting the PRSP was not very significant, due to the institutional reorganisation.

The region's propensity to evaluate (C2) accounts for the difference in outcomes between c2 and c4. In the first case, evaluation was a priority activity that formed a routine part of the operations of the GRSP and was designed to be an instrument of regulation and coordination. The evaluation tradition in this region has been emphasised above; in the other cases (c1/c4/c5), evaluation was replaced by more of a 'review' exercise.

External pressure (C3). The law gave responsibility to the body bringing together all stakeholders, but did not specify the breakdown of roles between this body and the GRSP. The CRSs got involved in evaluation mainly to lay claim to their rightful position, influence how the evaluation was conducted (c2/c3) or even impose themselves as joint coordinator (c3); in one case (c1), evaluation offered an opportunity to lay claim to greater recognition in health-related decision-making. In two cases (c1/c4), their expectations in terms of evaluative questioning were not taken into account. Their desire to target health outcome came up against methodological constraints, as the implementation timeframe did not allow for this type of analysis. The presentation of the report could perhaps have been an opportunity for proposals on their part for new orientations, but the reform specified new bodies. Indeed, the evaluation coincided with the end of their mandate. In spite of genuine investment, the group work (c3) could not be completed, given that the bodies in question were being reformed at the time.

Staff continuity (C4-C5). The directors of the future agencies, appointed by the Council of Ministers, all came from regions other than the ones in which they were to head up an agency. It was their responsibility to produce an organisational chart and form the team that was to be in charge, under the responsibility of the deputy director, of defining the strategic orientations

of health policy. The future of GRSP directors, ‘strategic decision-makers’ (D) was variable. A number of possibilities for representatives of the State were observed. They became directors; deputy managing director of the agency (c3/c4/c5) but not necessarily with responsibility for coordinating health policy (except in c4); of another regional organisation (c2); in another region (c1). The same was true of their co-directors; the latter joined the ARS (c2/c5) or took over as director of another institution (c1/c2/c3). The same process was apparent within the teams in charge of managing GRSPs (operational decision-makers), with some joining the department dedicated to the PPS within the ARS³⁶, others joining other ARS departments, others again moving to other regions or to other careers. Depending on the individuals’ choices and the organisational structure of the agency (some kept a dedicated department for PPS, others (c4/c5) did not), the former teams were split up or merged. Depending on the circumstances, continuity with the previous period – a favourable factor for lessons from the evaluation being taken into account – was sometimes maintained, sometimes not; one major factor was whether or not responsibility for the PPS department was assigned to an actor previously involved in the evaluation.

The presence of a leader (C6) enabled changes to be introduced in the short, medium and long terms (c1/c2), with advocacy actions carried out within the institution to organise a debate resulting in recommendations (c2). In both cases, this involved stakeholders engaged in the working groups and networks relating to the evaluation.

The users/decision-makers

The users/decision-makers of the evaluation were individuals with decision-making power or influence on the decision-maker with regard to the intervention being evaluated: ‘strategic’ decision-makers (D) (managing director and heads of department); ‘operational’ decision-makers or managers (M); decision preparers (those in charge of PRSPs/PRS/evaluation); and stakeholders (institutional partners, CRS) with the power to intervene at the level of the decision-maker.

The motivations of decision-makers (D1/M1). Contrasting situations can be observed in the two categories of decision-makers: tensions for the ‘strategic’ decision-makers between their commitment to public health policy and their motivation to evaluate it (c3); and between

³⁶ These departments have taken over the responsibilities formerly entrusted to GRSPs.

‘strategic’ and ‘operational’ decision-makers in their motivation to evaluate (c1/c5). In two regions, disagreements can be identified between the stakeholders concerning the evaluation questions (c1) or how the evaluation was conducted (c3). In the other regions, there was a consensus between the stakeholders in favour of the evaluation, this being either strong (c2) or merely moderate (c4). The motivation to evaluate and to use the evaluation depended on a number of factors: the existence of a positive experience of evaluation; the presence of a recognised figure in the evaluation; the latter’s place in the organisation (enough leeway to use the results); and whether or not they intended to remain in the future agency. Generally speaking, ‘strategic’ directors had little experience to bring to bear with respect to evaluations that had an impact on decision-making.

Opinions of evaluation (D2/M2). The differences in how evaluation was viewed correlate less with the rank of the decision-maker than with their experience in evaluation and policy management. Experienced high-ranking directors expected to know about the outcome (which had not been forthcoming from previous evaluations), the findings expected concerning both the efficacy and the effectiveness of public action. They considered evaluation to be a management, communication and accountability instrument for public action. Decision-makers who had become involved more recently in the management of the health system sought to gain better knowledge of the arrangements for the governance of public action. For ‘operational’ decision-makers, the essential function of evaluation was to question and ascribe meaning to actions, thereby providing a fresh impulse to the network of professionals and external partners. All of them complained that this activity was not sufficiently developed, mentioning a number of hindrances: a growing number of national mandates; instability in terms of orientation and organisation; the difficulty of gathering field data; a lack of resources and competencies; and difficulty in identifying a clear leader given the segmentation of activities.

Agreement with findings (D3/M3). This was achieved more frequently among ‘operational’ decision-makers than with ‘strategic’ decision-makers.

Competencies (M4). The situations were diverse in terms of competencies within the institution: present (c2); strengthened through the expertise of the evaluation committee (c1/c3); being learned (c1/c5); or low (c4).

The evaluation process

Motivation (E1). Regulatory requirement was the main motivation, except for c2 where the evaluation formed a management tool. For the regions that conducted it at a later stage, the purpose of the evaluation – in the light of an impending, substantial review of the planning instrument – was re-envisaged as a ‘deed of transfer’ to the successors.

Work setting, committee leader, consensus on issues (E2/E3/E4). The national guide and initial evaluation were points of reference for all regions. Engaged in a dynamic of inter-regional discussions, the actors took on board certain principles, such as the formation of a pluralist committee or the preference for outsourced evaluation, or opted for similar approaches (local observations) – all the more so due to the fact that a number of evaluations were conducted by the same provider. The work was conducted by an external operator (except c5) who benefited from greater recognition when they were from outside the region. The evaluator was sometimes positioned as leading the procedure (c2/c3), sometimes as a more remote expert (c1/c4). In one case (c5), the evaluator belonged to the institution, and so delegated data-gathering to partners whilst they themselves focused on analysing this information.

The credibility of results (E5). The questions related mainly to management, resulting in a shift from evaluation of the plan to evaluation of the institution. The observations were unanimous concerning the problems with the scheme and were difficult to accept. In one case (c2), although felt to be ‘highly critical’, the judgement was not questioned because ‘it was based on facts, was well-argued, comprehensible and backed up’, while in another case (c4) it was challenged on the grounds that it was ‘non-scientific and intuitive’, being based mainly on operators’ points of view. Elsewhere, the credibility of the evaluation was adversely affected by the close ties of the evaluator to the institution (c4) or the fact that they were actually part of it (c5).

Timeliness of the evaluation (E6). Timing was a determining factor: in cases where the agency was to be set up imminently, it was impossible to implement concrete measures (c1/c3/c5).

Communication (E7). With the exception of c2, there was a lack of communication in the evaluations, both during the process and at the end, with the report to the CRS being quite a

formal exercise. The reports were not circulated (c1/c3), or else circulation involved nothing more than them being placed online (c4/c5).

Although all five regions were subject to the same rules, case analysis reveals individual characteristics for each situation, affected by the context rooted in local history, experience in the implementation of health policies, the tradition of evaluation, the actors present, and the way in which the evaluation itself was conducted. Consequently, there is a clear difference in the outcomes of the evaluations. Case 2 demonstrates the use of the many different dimensions of evaluation (instrumental, conceptual, processive and for legitimisation), aided by the fact that it had a great many assets. Conversely, cases c4/c5 had an abundance of difficulties, the lack of motivation of the decision-makers being a major determining factor. Cases c1/c3 had a few assets but were dominated by considerations relating to the reform.

Analysis of the variables using PROMETHEE GAIA

The PROMETHEE GAIA analysis provides a complementary interpretation, including both a ranking of the cases and an analysis of the influence of the variables resulting in this ranking.

Case ranking. The ranking (diagram 2) is a result of the overall analysis (comparison in pairs for each variable) and is as follows: case 2/case 3/case 1/case 4/case 5. The highest-ranking case corresponds to the one deemed to be the most favourable to evaluation use.

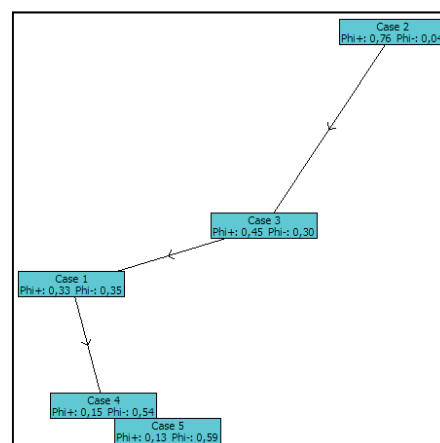


Figure 2. PROMETHEE Graph

Case profiles. Figure 3 details the case profiles. For each case, it provides a graphic representation of the performance of each variable. A colour code, in addition to a letter designating the category of variable, makes it easy to distinguish the most favourable category for each case and compare the cases with each other.

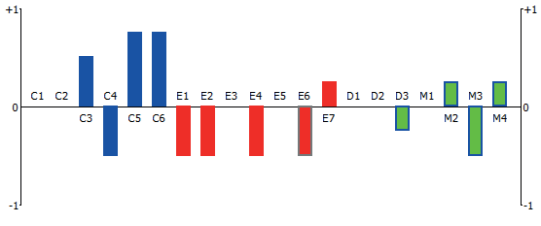
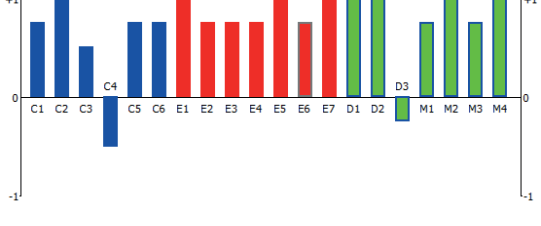
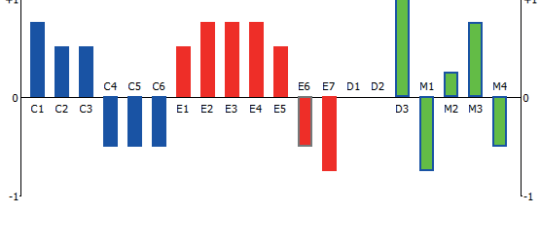
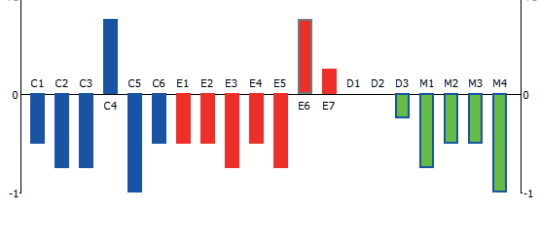
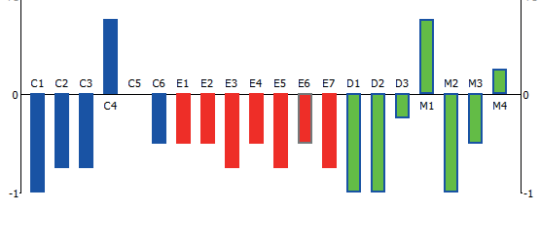
<p>Case 1</p> 	<p>Factors conducive to use</p> <ul style="list-style-type: none"> - pressure from CRS - GRSP professionals regroup in ARS - presence of a leader <p>Unfavourable factors</p> <ul style="list-style-type: none"> - change in leadership on transition to ARS - weak motivation for evaluation - decision-makers and managers disagree with the findings
<p>Case 2</p> 	<p>Factors conducive to use</p> <ul style="list-style-type: none"> - all except two <p>Unfavourable factors</p> <ul style="list-style-type: none"> - directors leave when ARS transition occurs - decision-makers disagree with the findings
<p>Case 3</p> 	<p>Factors conducive to use</p> <ul style="list-style-type: none"> - institutional cohesion - pressure from CRS - quality of process evaluation - decision-makers and managers agree with the findings <p>Unfavourable factors</p> <ul style="list-style-type: none"> - change in directors and managers when ARS transition occurs - late submission of the evaluation - not widely disseminated
<p>Case 4</p> 	<p>Factors conducive to use</p> <ul style="list-style-type: none"> - directors stay when ARS transition occurs - report submitted at an opportune time <p>Unfavourable factors</p> <ul style="list-style-type: none"> - unfavourable context - little credibility ascribed to the evaluation - unfavourable user opinions
<p>Case 5</p> 	<p>Factors conducive to use</p> <ul style="list-style-type: none"> - pressure from CRS - users motivated about the evaluation - continuity of directors <p>Unfavourable factors</p> <ul style="list-style-type: none"> - general context - little credibility ascribed to the evaluation - lack of interest on the part of decision-makers
<p>Letters designate categories as follows: C/Context, E/Evaluation, D/'Strategic' decision-makers, M/Managers for 'operational' decision-makers</p>	

Figure 3. Case profiles (PROMETHEE GAIA)

Case positioning. The GAIA plan (figure 4a) shows the axis of preference (thick red line on the bottom left): this is clearly directed towards c2. It positions the cases in one of four squares, those on the right (c2/c3) being in the ‘preference zone’ (stick direction), those on the left being in the opposite zone (c4/c5). The proximity of these two cases shows they are similar in terms of their potential for being used. Case 1 is positioned on the dividing line. Thus c2/c3 are deemed to be potentially more inclined to use the evaluation.

Diagram 4 (b) shows the breakdown of variables, represented by axis. Those expressing the same preference extend in similar directions, while those that are opposed extend in opposite directions. The longer the axis of the variable, the more discriminating it is.

Conflicting variables. The GAIA plan shows the conflict between variable C4 (the continuity of ‘strategic’ decision-makers between the two organisations) and the other variables. Whereas the continuity provided by ‘operational’ decision-makers is a favourable factor for evaluation, the opposite is observed with respect to high-ranking decision-makers. This variable is the reason for the ranking of c4/c5. The visualisation of this conflict is one way of expressing the tension between continuity and the spirit of the reform.

Proximate variables. The plan reveals distinct variables that contributed to the ranking (axis close together and extending in similar directions); since this is a relative analysis, they are not necessarily the variables with the best performance in each case. Although c2 has the highest number of favourable variables, it stands out from the others as regards variables C2, C3, D2, M2, and E5. The fact that the decision-makers accept the results of the evaluation is a characteristic of c3, but this is insufficient in the absence of other factors (timeliness, motivation, continuity). As has just been emphasised, c1 has assets in the form of action by ‘operational’ decision-makers (who are motivated and provide continuity), the presence of a leader capable of working internally to introduce the lessons learned from the evaluation, and collective competency in evaluation. The commitment of the decision-makers is however lacking.

Sensitive variables. An analysis of the stability intervals revealed eight ‘sensitive’ variables, i.e. variables capable, with higher weighting, of altering the ranking. In order of sensitivity, these were: motivation of ‘operational’ decision-makers; collective competency in evaluation;

continuity of ‘operational’ decision-makers; emergence of a leader; circulation of the evaluation; timeliness; continuity of ‘strategic’ decision-makers (which had the opposite effect); acceptance of the results by the decision-makers.

a. Cases position



b. Variables breakdown

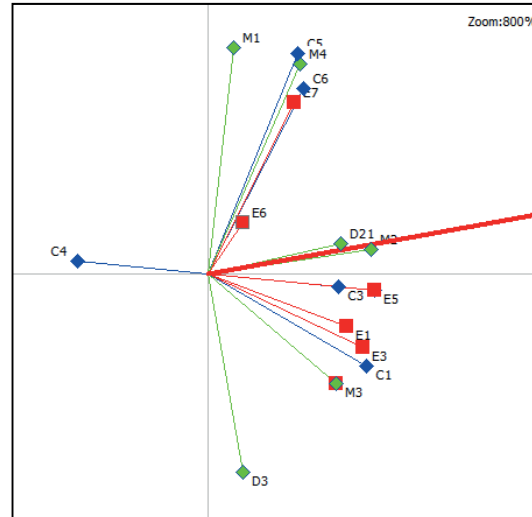


Figure 4. GAIA Plan

Weighting of variables. The model ranks c3 in second position. However, the comparative and qualitative study shows that there was barely more use than for c1. This is due to the fact that the ranking was carried out based on a set of variables with equal weight. Case 3 has eleven favourable variables, but most of them do not have any determining influence on use. The qualitative analysis of the case studies showed that certain variables (C6/leader, E5/credibility, E6/timeliness, D1/motivation of decision-makers) played a more important role. Equal weighting between all the variables (5%) does not take this into account. If higher weighting is assigned to the latter variables (11% to E5 and 13% to the others), c1 comes out ahead of c3.

A case-by-case analysis (GAIA plan) shows the most important variables for each case. A comparison of the cases identifies the following ‘sensitive’ variables: the presence of a leader (context); timeliness; dissemination (evaluation); motivation; continuity and competency (‘operational’ decision-makers); agreement with the results (‘strategic’ decision-makers). These variables do not constitute predictive factors as far as use is concerned; they

explain and provide a better understanding of the relative role of the variables by comparing the cases. They may be interpreted as sources of leverage that may be brought to bear to increase the chances of the evaluation being used.

Discussion: evaluation as a system of interactions

A context that limits users' room for manoeuvre

In the five cases presented, the context forms the background for the evaluation: the onset of a reform aimed at simplifying coordination of the health system and the inclusion of different components of health. The creation of ARSs, deemed to be an 'institutional innovation' (Pierru, 2011), and involving the merger of staff from different institutions with different statuses and cultures, was aimed at establishing more efficient management of the health system. This cultural and organisational revolution marked a complete break with the previous, partitioned organisation and its lack of coherence (Rolland and Pierru, 2013) – thereby effectively banishing references to past experience.

'There was the desire to establish new teams and a new institution and gather teams to build something new together, something that would favour cohesion, rather than producing something based solely on one of the components that had been brought together (...) there were issues of repositioning, power and people' (a GRSP director).

The evaluation of a system from an era consigned to the past is of little use in the present, even as a 'deed of transfer', especially if it focuses on the meanderings of policy management. The overhaul of the institution altered the position and attitude of the actors. For 'strategic' decision-makers, joining the agency amounted to agreeing to a disruptive change, in which case using the evaluation to build the future was all but meaningless. 'Operational' decision-makers, however, lost both their status and their room for manoeuvre. Their perception of the reform and the degree of recognition they received as a result served as grounds for their choice – either to prolong their action in the new organisation, or to take up another position in order to avoid being 'demoted'. Continuity between the two eras depended on their choice, as well as that of the new directors, who tended to opt either for a stance leaning towards new beginnings or one favouring continuity.

At the interface between the context of the evaluation and that of the institution, the users adopted a position whereby they either complied with or distanced themselves from the values and norms of the institution, depending on their experience/expertise in evaluation, status, personal project and the room for manoeuvre they felt they had. The dissemination of the national guide made a significant contribution to harmonising evaluative thinking, with the regions adopting the suggested evaluative questions. On the other hand, irrespective of the lessons learned from the procedure, the transformation of structures meant that it was not really possible for the lessons learned to have an effect within the institution.

Lastly, although it is difficult to prove formally in this instance given the limited number of cases, the size of the region could play a role, inasmuch as in smaller institutions, the smaller number of professionals are both less available and less specialised in the missions to be performed, and consequently less likely to take part in outside activities or events, which generate new ideas and changes.

Twofold credibility favouring acceptance of evaluation by users

The credibility of evaluation depends on a process of interpretation (Hurteau et al., 2010), combining the capacity of the evaluation to co-build reflection and provide new knowledge (documentation of public action in the territory), the quality of argumentation, the variety of the data, and the diversity of sources of information. The relationship between users and evaluators contributes to this; when closely associated with the construction of the procedure (c2), users understood the conclusions better, accepted them and applied them. In addition, the recognition given to the provider and the presence of an outside expert to lead the procedure (c2/c3) affected how much value was ascribed to the results. However, other conditions need to be met, too: although deemed to be of good quality, in some cases (c3) evaluation did not have any effects due the late timing and a change in the policy intermediary.

The emergence of a leader, building bridges between the institution, actors and evaluation

The presence of an individual (c1/c2) who is familiar with the workings of the institution, able to identify the opportunities and mechanisms liable to introduce change, and who can act as a link between the world of the institution and that of the evaluation, is a determining factor.

Acting as an interface between the evaluator and the decision-maker, this person undeniably plays a role (c2) in helping the decision-maker to accept the findings and encouraging them to organise reconciliatory discussions with the stakeholders. Trust and mutual recognition between the protagonists (decision-maker, evaluation manager, evaluators, working group leader) were all mechanisms that underpinned the success of this experience.

Appropriateness of the evaluation to the context of the institution

In addition to attributes (quality, timeliness, credibility) that meet users' expectations, evaluation must benefit from recognition and formal organisation if it is to lead to change. Cases c2 and c4, which have comparable credibility, can be distinguished on the basis of the regional and institution setting of the function.

Today, faced with the task of preparing the PRS, decision-makers are moving towards managerial forms of evaluation, which they wish to be in control of. Conducting 'documented management', involving teams, and bringing together stakeholders are the common concerns of the majority of regions, taking the view that work carried out internally is likely to be better suited to achieving these goals than evaluations that are more detailed from a methodological point of view. Awareness of such issues is gradually tending to lead to a re-evaluation of the expert evaluation model which was prevalent until recently.

Conclusion

The following main points emerge from this study: the dominant influence of the national context, which confined and reduced the autonomy of decision-makers and users; the role of the motivation of 'strategic' decision-makers, which mirrored their commitment to the system being evaluated and their opinion of the evaluation; the major role of 'operational' decision-makers, a force for lessons to be learned and in the creation of innovation; the significance of the emergence of a leader who could be an avenue for change at the most strategic level. The evaluation activity itself acted in some sense as a means of adjusting to the other components of the evaluation system.

Saunders (2012) establishes a distinction between *use* and *usability*, defined by this author as the ability to increase the potential for use of the evaluation on the basis of the design and

outputs of evaluation. However, the two are closely linked, and the engagement of decision-makers remains critical. The role played by the leader figure is the component which can bring the two together.

The overwhelming influence of organisational context on use has been highlighted by Hojlund (2014). He considers that evaluation is subject to a twofold influence: external pressure from the immediate environment, due to the particular constraints of the field in question, and the internal propensity for evaluation. By combining these two factors, he identified four models for the adoption of evaluation, in which different forms of use are distinguished: coercive, mimetic, normative and voluntary. The five cases in our study can be presented according to this classification. Case 2 falls within the mimetic model, with high propensity, high external pressure, and the presence of a leader able to encourage adoption of the evaluation; the evaluation has been useful both for management (instrumental use) and conflict resolution, thanks to dialogue with the stakeholders (legitimising use). Cases c1/c4/c5 are closer to the coercive adoption model, with a low propensity for evaluation and regulatory obligation; the separation between evaluation and other activities is not very conducive to use. Case 3, half way between the preceding cases, could fall within the normative adoption model, with the provision of external expertise, but with a lack of institutionalisation that impedes use.

Lastly, Ledermann's works (2012) make a stimulating contribution, shedding light on the links between the context, mechanisms and actors (top triangle in the diagram) and inviting us to engage in complementary interpretation of the regional cases.

Contributions and limits of the model

Certain limits of the model must be emphasised. The first is due to the singular context in which these evaluations took place; the advent of a reform is an exceptional event which may immediately be viewed as an obstacle to use. However, this situation illustrates the actual state of affairs in France, characterised by successive re-organisations of the health system (Tabuteau, 2013). The second is due to the low number of cases. This limit is less significant given the previous point, namely the weight of the national context, which evens out the constraints and setting of these evaluations. Our participation in other work in other regions,

not included in this study, confirms this hypothesis. It would however be worth testing the PROMETHEE GAIA approach on more cases and other examples. Notwithstanding this, the approach is worthwhile in that it provides a complementary interpretation in addition to the case study analysis. It also provides a comparative judgement, highlights the most illustrative variables in each case and makes it possible to determine the relative importance of each variable in the situation observed.

The evaluation of complex interventions calls for methodological innovations. Taking evaluation as one such complex intervention, we have experimented with a systemic approach and a diverse range of methods in order to assess its impact on regional health policies in France. Further research is needed in order to explore the MCDA method further; this appears to us to be a promising avenue, complementing qualitative approaches.

Acknowledgments

We would like to thank all the people, decision-makers, practitioners and all relevant stakeholders who have contributed to this work by giving some of their time to share their opinions and experiences on health policies evaluation. We would especially like to thank Anne Roué Le-Gall for her support in initiating and using Multicriteria method.

The authors declare they have no competing interests.

References

- Belton V and Stewart T (2002) *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer Science & Business Media.
- Castaing C (2012) La territorialisation des politiques de santé. <http://www.leh.fr/edition/>, text, Available from: <http://www.leh.fr/edition/hors-collection/la-territorialisation-des-politiques-de-sante-9782848743691.html> (accessed 9 July 2014).
- Chambaud L (2008) Entre continuité et rupture : la création des agences régionales de santé. *Actualités et dossier en santé publique*, 33, 7-10.
- Cousins JB and Leithwood KA (1986) Current empirical research on evaluation utilization. *Review of educational research*, 56(3), 331–364.

- Hojlund S (2014) Evaluation use in the organizational context - changing focus to improve theory. *Evaluation*, 20(1), 26–43.
- Hurteau M, Valois P and Bossiroy A (2010) Jugement crédible en évaluation de programme: définition et conditions requises. *Canadian Journal of Program Evaluation*, 25(2), 83-101.
- Jabot F and Loncle-Moriceau P (2010) Evolution des politiques locales de santé publique en France : quel partage des pouvoirs entre Etat et collectivités locales ? *Etat des savoirs en santé publique*, 372–382.
- Jabot F and Marchand A-C (2014) L'évaluation des plans de santé publique: un levier de changement pour les politiques régionales de santé en France ? *Global Health Promotion*, 21(1 suppl), 64–69.
- Johnson K, Greenesid LO, Toal SA, et al. (2009) Research on Evaluation Use A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
- Ledermann S (2012) Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159–178.
- Mareschal B, Brans JP and Vincke P (1986) How to select and how to rank projects: the Prométhée method. *European Journal of Operational Research*, Available from: <http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/9307> (accessed 13 July 2014).
- Mark MM and Henry GT (2004) The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10(1), 35–57.
- Pierru F (2011) L'administration de la santé en fusion : la difficile naissance des Agences régionales de santé. *Pouvoirs Locaux*, 75.
- Rolland C and Pierru F (2013) Les Agences Régionales de Santé deux ans après : une autonomie de façade. *Santé Publique*, Vol. 25(4), 411–419.
- Roué-LeGall A, Lucotte M, Carreau J, et al. (2005) Development of an Ecosystem Sensitivity Model Regarding Mercury Levels in Fish Using a Preference Modeling Methodology: Application to the Canadian Boreal System. *Environmental Science & Technology*, 39(24), 9412–9423.

- Saunders M (2012) The use and usability of evaluation outputs: A social practice approach. *Evaluation*, 18(4), 421–436.
- Shiell A, Hawe P and Gold L (2008) Complex interventions or complex systems? Implications for health economic evaluation. *BMJ*, 336(7656), 1281–1283.
- Tabuteau D (2013) 1983-2013 : les évolutions de la politique de santé. *Journal de gestion et d'économie médicales*, Vol. 31(1), 53–67.
- Weiss CH (1998) Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, 19(1), 21–33.
- Yin RK (2009) *Case Study Research: Design and Methods*. SAGE.

Chapitre 6 – Synthèse et discussion

S'attarder sur le paradoxe de l'évaluation ne relève pas d'une vaine intention de vouloir l'élucider, tant la réalité est complexe, mais plutôt d'une entreprise de recherche des raisons expliquant une certaine « vitalité » (Perret, 2009a) et pérennité de l'évaluation en dépit d'une influence contingente.

Au terme de ce travail, trois constats peuvent être avancés. Premièrement, des usages pas nécessairement délibérés se dissimulent derrière une apparente inutilité de l'exercice évaluatif, la prise de décision formelle n'étant pas la seule forme d'utilisation (1). Deuxièmement, si le devenir des travaux d'évaluation est en soi imprévisible, il existe un corpus de connaissances sur les conditions requises pour une exploitation de ces travaux ; ces conditions sont plurielles et interviennent à plusieurs niveaux (2). Enfin, l'évaluation elle-même peut être un levier de son utilisation (3).

1. Les usages délibérés et non délibérés de l'évaluation

Pour Dobrow (2004), le contexte et les données probantes (*evidence*) sont deux paramètres fondamentaux de la prise de décision. Pour situer l'équilibre entre ces deux paramètres dans les politiques basées sur les preuves, il propose de considérer deux axes : l'axe des preuves et l'axe du contexte.

La médecine fondée sur les preuves (*evidence-based medicine*) exige un haut niveau de preuves et tend à neutraliser le contexte. A l'opposé, la décision politique traditionnelle accorde plus de poids aux éléments contextuels (analyse des besoins, conséquences économiques, préférences des citoyens) et peu à la nature des données. La politique fondée sur les preuves, *evidence-based policy (EBP)*, est située à mi-chemin, dans la recherche d'un compromis entre les deux, considérant que mieux on comprend le contexte, mieux on est en mesure d'utiliser les données probantes.

L'évaluation n'est pas la recherche et, contrairement à l'EBP, il ne s'agit pas d'étayer les politiques sur les meilleures données disponibles pour prendre des décisions à l'échelle d'une population sur un sujet donné, mais de s'appuyer sur les données d'évaluation d'une politique pour aider les décideurs à prendre des décisions avisées concernant cette politique. La

problématique peut être formulée en d'autres termes, ainsi que le suggère Ledermann (2012). L'évaluation peut-elle informer les décisions (*evaluation-informed decisions*) en santé ? Cette question interroge la construction et le périmètre de la décision, les acteurs en position de décision ou d'utilisation, la nature des connaissances produites, les formes d'utilisation de ces connaissances, les critères permettant d'apprécier le bien-fondé ou le succès de l'utilisation ainsi que les conditions nécessaires à l'appropriation des connaissances pour informer la décision.

Par ailleurs, la prise de décision ne traduit qu'un aspect de l'utilisation et ne saurait résumer *l'influence* de l'évaluation dans ses multiples manifestations qu'il nous faut alors retracer.

1.1. Le défi de la traçabilité de l'utilisation

Les travaux de recherche ont majoritairement documenté l'utilisation de l'évaluation dans sa dimension instrumentale (Johnson, 2009), plus facilement objectivable et imputable à l'évaluation. Les autres formes d'utilisation sont plus difficiles à cerner.

La connaissance dérivée de l'évaluation, contrairement à la connaissance dérivée de la recherche, ne concerne pas seulement des idées mais aussi des savoirs issus de situations expérimentées, ce que traduit la distinction entre utilisation conceptuelle et utilisation processuelle. La première fait référence aux apprentissages générés par la production de connaissances nouvelles sur l'intervention évaluée tandis que la seconde relève de la participation au processus d'évaluation. Cette distinction, enrichissante pour la compréhension des mécanismes en jeu dans l'utilisation de l'évaluation, est cependant moins pertinente sur le plan pratique car il est difficile de faire la part entre ce qui est attribuable à l'une ou l'autre de ces deux sources d'apprentissages, les acteurs participant à l'évaluation étant généralement en situation d'exploiter les productions de l'évaluation.

L'influence de l'évaluation, quelle que soit sa source, est difficile à tracer pour un ensemble de raisons. Tout d'abord, précisément du fait de sa temporalité. Comme elle s'inscrit dans le moyen et long terme, les acteurs présents au moment de l'évaluation sont moins facilement identifiables, eu égard à l'évolution de leur trajectoire professionnelle. Par ailleurs, les apprentissages issus de l'évaluation se diluent dans le temps et dans les acquis de multiples expériences conduites au cours de l'exercice professionnel. Lors des entretiens

conduits avec les acteurs ayant participé cinq ans auparavant à l'évaluation des PRSP, nous avons constaté que ces derniers ne faisaient pas toujours très clairement la distinction entre l'évaluation et la mise en œuvre du plan lui-même, faisant des allers et retours constants entre les deux. Par exemple, la participation aux concertations organisées avec les partenaires extérieurs était rapportée comme une situation facilitant la compréhension des enjeux de chacun et enrichissante pour la pratique, mêlant tour à tour les échanges vécus, tant au sein du comité de suivi de l'évaluation que dans le cadre des conférences de santé. Il est également apparu que les personnes qui évoquaient le plus les bénéfices de l'évaluation, n'étaient pas toujours celles qui avaient été le plus étroitement ou durablement associées à l'évaluation, mais qui l'avaient été à quelques moments clés. Enfin, leur discours peut être modifié au fil du temps, les personnes incorporant dans leurs conceptions et dans leurs pratiques, des éléments devenus signifiants du fait de changement de situation professionnelle.

L'utilisation de cette connaissance se traduit à un niveau plus global, dans la mise en œuvre de la politique. Pour DeGroff (2009), la mise en œuvre d'une politique, et d'autant plus qu'elle est intersectorielle, varie selon le mode de gouvernance entre les partenaires, le contexte sociopolitique et l'influence du *New Public Management*. De fait, l'appréciation portée sur l'utilisation dépend de ces trois paramètres. L'utilisation instrumentale relève plus d'une logique descendante (*top down*) ; la fidélité aux orientations préconisées étant le critère de succès de l'utilisation. A contrario, dans une logique plutôt ascendante (*bottom up*), l'utilisation relève de mécanismes d'adaptation à partir de l'appropriation par les acteurs ; les formes d'innovations étant le critère de succès.

A défaut de mesures affectant directement une intervention (le PRSP dans notre étude), d'autres formes d'usage, notamment la capacité de transposition - si ce n'est d'innovation - des savoirs acquis dans d'autres contextes, témoignent de l'utilité de l'évaluation.

1.2. Le périmètre de la cible : les utilisateurs

Prenant acte que l'utilisation n'est pas limitée à un ensemble de décisions discrètes, le périmètre de la cible de l'évaluation, à savoir les personnes en mesure de faire usage de l'évaluation, s'agrandit, des décideurs formels aux utilisateurs. Identifier la bonne cible, les vrais utilisateurs, et les moyens de l'atteindre, est un enjeu majeur de la démarche

d'évaluation. Le terme utilisateur créé une ambiguïté selon qu'il désigne *a minima*, toute personne tirant bénéfice d'une évaluation, sans présumer de son lien avec l'intervention évaluée, ou toute personne apte à modifier l'intervention évaluée et/ou le cadre de sa mise en œuvre. Les utilisateurs ne sont pas *stricto sensu* seulement des décideurs. A contrario, les décideurs sont-ils les utilisateurs ? Les décisions formelles sont-elles les actes qui traduisent le mieux les changements et conduisent-elles toujours au changement ?

Patton (2010) définit les vrais utilisateurs comme « les innovateurs et autres personnes qui travaillent à apporter du changement¹ ». Dans cette acception, il s'agit moins d'identifier des catégories que de considérer la capacité d'influence des personnes, au regard de leur pouvoir de décision, formel (décideurs/directeurs) ou d'influence (conseillers, préparateurs de la décision), leur capacité à orienter et gouverner la mise en œuvre (planificateurs et gestionnaires des programmes) ou à modifier les pratiques (professionnels qui opèrent au plus près des bénéficiaires). La capacité d'influence de chaque catégorie dépend de la logique (*bottom up* ou *top down*) et du mode de gouvernance de la politique, plus ou moins formalisé par des normes et procédures, et des marges de manœuvre disponibles pour chacune d'elles. En effet, nous avons observé, analysant la performance du système de santé régional (article 1 du chapitre 5), les tensions entre des logiques contradictoires portées entre les acteurs des niveaux régional et territorial, les premiers soutenant la déclinaison des orientations nationales (*top down*), les seconds défendant les innovations territoriales (*bottom up*). Face à des procédures relativement peu flexibles (révision du plan par exemple) ou des situations figées, les acteurs, fonction de la représentation de leur rôle, de leur conformité à la norme et de leur potentiel créatif, ont (ou non) développé des stratégies pour assurer la couverture du territoire en termes de structures ou la survie de projets.

Ces éléments montrent combien il est difficile de répondre de façon binaire à la question de l'utilisation (Blake and Ottoson, 2009), tant les formes d'exploitation sont multiples, diffuses et à distance de l'évaluation. De fait, l'emploi du terme *usage* semble davantage approprié pour évoquer les retombées, difficilement tangibles et délibérément ou non recherchées, d'une évaluation.

¹ « *the primary intended users are innovators and others working to bring about major change* » (Patton, 2010, p.14).

2. Les conditions du changement

La dynamique de changement s'appréhende dans la globalité des interactions entre le contexte, les acteurs et l'évaluation. Nous retenons des travaux sur l'évaluation des PRSP, l'importance du contexte, de l'engagement et de la motivation des décideurs/utilisateurs et de la qualité du processus d'évaluation sur l'utilisation de l'évaluation, même si ce critère n'était pas le plus prégnant.

Cette section fait un focus sur les conditions du changement inhérentes au contexte institutionnel et sur les mécanismes par lesquels les acteurs facilitent ou entravent le changement.

2.1. Le terrain de l'évaluation : atouts et contraintes

La prise en compte de l'environnement de l'évaluation s'est progressivement accrue ces dernières années (Conner, 2012 ; Fitzpatrick, 2012 ; Greene, 2005, 2009 ; Rog, 2012) avec un intérêt beaucoup plus marqué sur le rôle du contexte institutionnel (Appelton-Dyers et al., 2012 ; Hanberger, 2011 ; Holjundt, 2014 ; Ledermann, 2012 ; Pattyn, 2014).

2.1.1. Les travaux récents

Deux travaux récents retiennent notre attention afin de confronter leurs conclusions aux résultats de nos analyses de l'évaluation des PRSP (articles 1 et 4 du chapitre 5) : ceux de Pattyn (2014) qui étudie l'inclination des organisations à évaluer au regard de leurs caractéristiques propres et ceux de Cousins and Bourgeois (2014) qui mettent en relation capacité à faire et capacité à évaluer.

La propension à évaluer

A la lumière des apports d'Holjund (2014), nous avons retenu dans notre modèle théorique, la propension à évaluer d'une institution parmi les variables pertinentes pour analyser le contexte de l'organisation (article 1 du chapitre 3). Par la suite, nous avons tenté de mettre en lien, les formes d'utilisation et les modèles d'adoption de l'évaluation (coercitif, mimétique, normatif et volontaire) dans les cinq régions étudiées (article 4 du chapitre 5).

Les travaux de Pattyn (2014) apportent un éclairage additionnel sur les conditions reliées à l'activité d'évaluation. En effet, cet auteur a étudié les motifs pour lesquels les organisations du secteur public en Flandres évaluent ou n'évaluent pas leur politique, à partir de l'analyse des interactions entre les acteurs et le contexte institutionnel dans lequel ils exercent. Ce travail lui a permis de distinguer les conditions pertinentes pour expliquer l'activité ou la non-activité d'évaluation de celles qui sont sans pouvoir explicatif. Certaines relèvent de la culture, d'autres de la capacité d'évaluation (De Peuter and Pattyn, 2008).

Nous retenons de ses conclusions les points suivants.

- Le fait d'avoir réalisé une évaluation crée une « appétence » pour cette pratique ; les organisations ayant déjà évalué renouvellent l'exercice ;
- Les organisations qui pensent que les résultats (outcome) sont faciles à mesurer sont enclines à initier des évaluations alors que la situation inverse est observée pour les réalisations (outputs) : si elles sont faciles à mesurer, cela ne justifie pas une évaluation ;
- L'inclination à évaluer dépend de la formulation des résultats (SMART²) et des attitudes vis-à-vis de l'évaluation ;
- Les organisations qui évaluent, ont un socle minimum de compétences, au moins pour externaliser l'évaluation ;
- Si la présence d'une unité d'évaluation dans l'organisation ne semble pas essentielle, elle joue cependant un rôle parce qu'elle facilite la mobilisation des ressources. Inversement, elle peut exercer une influence négative pour les organisations qui ont été récemment responsabilisées dans ce domaine et ne souhaitent pas voir leur activité restreinte au sein d'un seul service, l'ancrage institutionnel de la fonction étant un élément favorisant ;
- L'absence d'activité d'évaluation est expliquée par l'absence de demande de la part du management organisationnel ;
- L'existence d'un système d'informations développé, la participation des équipes aux formations et aux réseaux, la contrainte réglementaire, la stabilité organisationnelle et l'autonomie de l'organisation ne sont pas des conditions validées dans son étude.

La relation entre capacité à utiliser et capacité à faire

Les travaux de Cousins and Bourgeois (2014) complètent les apports de Pattyn. Ils soulignent la corrélation entre la capacité à utiliser l'évaluation et la capacité à faire l'évaluation.

² Spécifique, mesurable, atteignable, réaliste, temporellement défini.

La capacité à faire est largement déterminée par les caractéristiques et les stratégies de l'organisation : présence de structures, bien positionnées, disposant de ressources et d'une équipe dédiée ; existence d'une expertise interne capable, non seulement de promouvoir l'évaluation, mais aussi de développer des procédures et outils adaptés, de communiquer en interne et de répondre aux besoins de l'institution, y compris sur l'exigence de rendre compte ; culture d'apprentissage effective favorisant l'implication des équipes et permettant de faire de l'évaluation, une base de réflexion stratégique ; engagement des décideurs de niveau stratégique.

Les auteurs insistent sur le caractère mixte de la formation à assurer dans l'organisation, théorique et expérientielle (*learning by doing*), l'importance de créer des partenariats avec des organismes pouvant accompagner les acteurs ainsi que le rôle indiscutable de l'expérimentation comme préalable à l'apprentissage (Cousins et al., 2014).

2.1.2. Nos résultats face aux conclusions de ces deux recherches

La contrainte réglementaire n'a eu qu'une influence modérée sur l'évaluation : efficace pour initier la démarche d'évaluation mais insuffisante dans ses modalités d'application. Les régions ne se sont pas engagées avec le même degré de motivation, la date de lancement étant un indicateur de l'intérêt supposé à l'évaluation ; seulement un quart des régions avait lancé leurs travaux bien avant le terme du plan tandis qu'un autre quart l'avait fait alors que le passage en ARS était imminent. Les premières, plus entraînées à l'évaluation, et aussi plus « appétentes », engageaient des travaux en s'efforçant de satisfaire aux règles de bonnes pratiques et en affectant les ressources nécessaires, tandis que les secondes, moins familières avec l'exercice, exécutaient la clause, dans le cadre d'un exercice interne plus proche d'un bilan et plus facile à exécuter qu'une véritable évaluation.

L'expérimentation de l'évaluation, tant à titre individuel que collectif, fonctionne bien comme un stimulus pour initier la pratique et entraîner l'adhésion des acteurs ou institutions.

L'existence d'une structure formelle, la MRIICE, n'a pas été un appui pour promouvoir, soutenir ou exploiter l'évaluation, faute d'avoir construit une compétence dans ce domaine. La confusion longtemps entretenue entre évaluation et contrôle a discrédité ces services vis-à-vis de l'évaluation et, par extension, écarté les professionnels de la conduite de cette activité pour le PRSP. Quand un professionnel d'une MRIICE y a contribué, c'est en raison de son

intérêt personnel et de sa familiarité de l'évaluation. Cousins et al. (2014) rapportent que dans les organisations étudiées, l'association des fonctions d'évaluation et d'audit dans la même unité est préjudiciable au développement de l'évaluation ; il en est de même lorsque la fonction de reddition de comptes prime sur celle d'apprentissage.

Le rôle du niveau national a eu un effet d'incitation, par le biais d'une enquête sur l'état d'avancement des travaux en région, vécue parfois comme une injonction, et un effet d'entraînement en créant une dynamique d'accompagnement. Cependant, la préparation de la loi HPST a stoppé cet élan. Faute du maintien de l'intérêt envers les démarches régionales et de capitalisation des travaux entrepris, l'évaluation s'est vue reléguée au rang des multiples tâches administratives. Si l'utilisation des données valorise les données (Cousins et al., 2014), a contrario, la désaffection dévalorise l'exercice.

Enfin, *l'inscription des équipes dans les réseaux professionnels* nous paraît exercer une influence favorable, en motivant les acteurs et en important de nouvelles pratiques. Dans les deux régions de notre étude, les promoteurs de l'évaluation en étaient des membres actifs. Nous avons pu constater dans nos études sur l'activité des services régionaux en matière d'évaluation (Carbonnel et al., 2009 ; Jabot, 2005) que les plus actifs avaient généralement des professionnels présents dans des groupes de travail, nationaux ou interrégionaux, suivaient ou intervenaient davantage dans des formations en évaluation. Ce point mériterait d'être validé et approfondi dans des recherches futures.

Chaque situation d'évaluation survient sur un terrain plus ou moins favorable sur lequel les acteurs jouent un rôle majeur. Les approches qui explorent les interactions entre les deux apportent une contribution à la compréhension des mécanismes explicatifs des changements observés (Henry and Mark, 2003 ; Ledermann, 2012 ; Mark and Henry, 2004).

2.2. Les dispositions et comportements des acteurs envers l'évaluation

Les dispositions des acteurs évoquent des orientations individuelles relevant des registres, aussi bien cognitif qu'affectif ou motivationnel (intentions, inclination, opinions) qui influencent leurs attitudes et comportements envers l'évaluation. Elles sont à relier aux mécanismes générateurs de changement identifiés par Henry and Mark (2003) pour expliquer l'influence de l'évaluation.

En effet, ces auteurs, analysent l'influence de l'évaluation à trois niveaux : sur les représentations, conceptions et comportements des individus eux-mêmes, sur le collectif à partir des échanges entre les personnes et sur l'organisation. Ils ont identifié six mécanismes intervenant au niveau individuel (changement d'attitude, importance prépondérante accordée à un sujet, formation d'une opinion, accroche à un concept, acquisition de savoir-faire, changement de comportement)³, cinq au niveau interpersonnel (persuasion, justification, agent de changement, influence de l'opinion minoritaire, normes sociales)⁴ et quatre au niveau organisationnel (mise à l'agenda, politique orientée par l'apprentissage, changement de politique, diffusion)⁵.

Dans notre étude (article 4 du chapitre 5), nous avons relevé dans nos entretiens, que les décideurs stratégiques, situés à un rang élevé dans la hiérarchie, citaient peu d'expériences marquantes d'évaluation leur ayant forgé une conviction sur les apports positifs de l'évaluation. Dans le cas contraire, nous avons constaté que leur intérêt à l'évaluation a été « accroché » (*priming*) par un évènement particulier ou un échange avec une personne. Le plus souvent, ils ont expérimenté l'évaluation comme une tâche réglementaire, destinée à rendre compte de l'action menée auprès du niveau national ou des partenaires régionaux, de « vérifier le bon usage des fonds publics », ou minimalement de « faire un état des lieux des avancées ». Les décideurs justifiaient l'absence d'évaluation par le flou des objectifs et l'incapacité à définir des indicateurs pertinents et le manque de savoir-faire, faute d'objectifs suffisamment précis ; constat qui rejoint celui de Pattyn (2014). Difficile, exigeante en temps et ressources, l'évaluation est par ailleurs concurrencée par des activités jugées hautement prioritaires (inspections/contrôles des établissements de santé) et les commandes successives du niveau national. Exercice imposé, elle reste une activité de second plan.

Les décideurs opérationnels témoignaient davantage d'expériences d'évaluation jugées enrichissantes ou utiles, auxquelles ils avaient contribué de façon active à une ou plusieurs étapes. Sans négliger la dimension réglementaire, ils voient dans l'évaluation, l'occasion d'accumuler de la connaissance et de la rétrocéder aux partenaires. Ils identifient précisément les obstacles à la promotion de l'évaluation dans leur institution (faible motivation du décideur, surcharge de travail, déficit de données et de système d'informations, insuffisance

³ Attitude change, salience, elaboration, priming, skill acquisition, behavioral change.

⁴ Justification, persuasion, change agent, social norms, minority-opinion influence.

⁵ Agenda setting, policy-oriented learning, policy change, diffusion.

de compétences, changement fréquent des orientations) et déploient, pour certains d'entre eux, un ensemble de stratégies pour convaincre ou faire pression sur leur directeur (*persuasion*) et pour mobiliser leurs collègues et collaborateurs. A cet égard, nonobstant le débat sur le risque de conflits d'intérêt, un consensus semble se dessiner en faveur d'évaluations internes pour entraîner (dans la double acception du terme) les équipes et écarter le risque d'être « dépossédé » de l'évaluation. Cette perception de l'évaluateur externe qui « confisque » l'évaluation, relativement récente dans les discours, traduit la volonté de (ré)investir cette activité .

L'identification des ressorts qui interviennent pour susciter l'intérêt et l'adhésion des individus et du collectif, s'accompagne d'une analyse des atouts et des contraintes du contexte. Les raisons pour lesquelles les décideurs de haut rang n'intègrent pas l'évaluation dans la panoplie des outils de management et de décision sont à rechercher du côté de leur expérience propre mais aussi des contingences extérieures qui restreignent leur marge de manœuvre. A l'articulation entre le contexte institutionnel et le dispositif d'évaluation, et à l'interface des acteurs plus spécifiquement chargés du pilotage de l'évaluation et des décideurs stratégiques, les décideurs opérationnels constituent une cible privilégiée pour activer les leviers et devenir des vecteurs du changement.

2.3. Les interactions entre le contexte et les acteurs

Les travaux de Ledermann (2012) apportent une contribution intéressante sur les interactions entre les conditions contextuelles, les mécanismes activant le changement et les perceptions des acteurs. Son modèle d'analyse dérive de celui de Valovirta (2002) qui retient deux facteurs majeurs pour le changement : le niveau de conflit entre les parties prenantes et le degré de pression externe. En combinant les deux, il en résulte quatre mécanismes par lesquels l'évaluation est susceptible d'activer le changement : a) révélateur (*awekener*), l'évaluation révèle des problèmes non connus et suscite une prise de conscience (situation de faible conflit et faible pression) ; b) déclencheur (*trigger*), l'évaluation peut entraîner un changement facilement accepté (situation de faible conflit et forte pression) ; c) conciliateur (*conciliator*), l'évaluation facilite l'acceptation du changement (situation de conflit élevé et de faible pression) ; d) arbitre (*referee*), l'évaluation arbitre les conflits et décide de ce qui doit

être fait, faute de consensus entre les parties prenantes (situation de conflit élevé et de forte pression).

Une variable d'ajustement : la perception de la qualité de l'évaluation

Ledermann rajoute à ce cadre, une variable relative à la perception des acteurs, faisant l'hypothèse que la qualité perçue de l'évaluation, en termes de validité, valeur ajoutée et pouvoir de conviction, est plus ou moins nécessaire selon le degré de conflit entre les acteurs et la prise de conscience de la nécessité de changement. Par exemple, en cas de conflit élevé et de forte pression, dans la mesure où le changement est inéluctable, la qualité est moins requise.

L'auteur a appliqué ce cadre d'analyse à onze évaluations réalisées en Suisse afin d'identifier les raisons pour lesquelles l'évaluation conduit ou non à la décision. Ses travaux confirment les hypothèses de Valovirta : pour les situations où l'évaluation se produit dans un environnement consensuel, que le changement s'avère ou non indispensable, l'évaluation doit être de qualité et « digne de foi ». En revanche, dans les circonstances de conflit, elle conclut à une relation étroite entre degré de qualité de l'évaluation et degré de concordance des constats de l'évaluation avec les croyances des décideurs ; l'évaluation jouant le rôle de test de vérité (Weiss and Bucavalas, 1980). Si l'évaluation n'apporte pas d'éléments nouveaux mais confirme les points de vue du décideur, ce dernier approuve les résultats pour le changement, d'autant mieux que l'évaluation est de qualité. Dans le cas contraire, l'évaluation doit être de haute qualité pour amener le décideur à réviser ses positions.

Dans notre étude (article 4 du chapitre 5), nous avons effectivement constaté que la perception de la qualité avait fait basculer l'attitude d'un décideur face à des conclusions qu'il jugeait sévères, contrairement à une autre région où le caractère « non scientifique » de l'évaluation était mis en avant pour expliquer les raisons de sa faible exploitation.

Contexte et rapport à la connaissance produite par l'évaluation

Dans le cas de l'évaluation des PRSP, les travaux sont réalisés dans une période de transition précédant la mise en place d'une réforme d'envergure, qui introduit une rupture avec l'époque antérieure, ce qui constitue un frein à l'exploitation de l'évaluation, non seulement dans sa dimension instrumentale mais aussi, conceptuelle, en totale contradiction avec la dimension cognitive incontestée de l'évaluation. A plusieurs reprises, les acteurs avec qui nous nous

sommes entretenus, ont évoqué « l’amnésie collective » dont était frappée leur institution au début de la mise en place des ARS.

Types d’acteurs et leviers

Trois catégories d’acteurs institutionnels retiennent l’intérêt : les décideurs stratégiques dont il est nécessaire de saisir le modèle mental (Leviton, 2003), les décideurs opérationnels, en plus grande proximité avec l’intervention évaluée et son évaluation, qui peuvent influencer les décideurs stratégiques et les personnes responsables de la conduite de l’évaluation. Nous désignons ces dernières, les « chargés d’évaluation ».

Dans la partie suivante, nous examinons la fonction de l’évaluation comme levier de changement avec une mention particulière sur le rôle des différents acteurs impliqués.

3. Le potentiel de l’évaluation

La portée de l’évaluation proprement dite dans l’utilisation est discutée. Le travail de Ledermann (2012) ci-dessus présenté, relativise les travaux antérieurs (Fleisher and Christie, 2009 ; Johnson, 2009) qui citent la qualité comme facteur de moindre importance. Saunders (2012) qui distingue l’utilisation de l’évaluation (*evaluation use*) du potentiel d’utilisation de l’évaluation (*usability*), propose de consolider les pratiques évaluatives pour renforcer le potentiel d’utilisation, en ciblant notamment la conduite, le devis, la forme des résultats et la communication de l’évaluation. Dans tous les cas, la qualité de l’évaluation n’est pas une variable neutre, même si son influence s’équilibre avec les autres facteurs en jeu.

Nous proposons de revenir, au regard de l’observation des cinq cas (article 4 du chapitre 5) et des données de la littérature, sur trois points : la crédibilité de l’évaluation, les acteurs spécifiquement impliqués dans le dispositif d’évaluation et la conduite de l’évaluation.

3.1. La crédibilité, élément fondamental du processus d'évaluation

Considérant la relation entre qualité et utilisation de l'évaluation, les associations professionnelles d'évaluation ont élaboré des standards pour garantir la qualité des travaux. Cependant, la qualité ne suffit pas, plusieurs études ayant montré la faiblesse du lien entre la qualité scientifique de travaux d'évaluation ou de recherche et leur utilisation (Black 2001, Marmot 2004, Sheldon, 2005). La notion de crédibilité se substitue à la notion de qualité, car elle prend en compte, l'appréciation scientifique, la croyance et la confiance que les utilisateurs accordent à l'évaluation.

La crédibilité associée à l'évaluation est triple. Elle concerne les données, le processus de production et les personnes engagées dans l'activité. Nous abordons les deux premiers aspects.

3.1.1. La crédibilité en tension entre validité et acceptabilité

La notion de données probantes a été l'objet de nombreux débats (Donaldson et al. 2009). Le concept a évolué : la preuve n'est plus seulement définie à partir de la méthode de production et « ce qui compte comme données probantes » intègre la notion d'application et d'utilité pour le décideur. Les données crédibles sont valides, rigoureusement produites, objectives et argumentées et porteuses de sens, c'est à dire reliées aux connaissances et intérêts des décideurs, apportent des connaissances nouvelles et utiles pour la conduite de l'action publique.

Pour autant, l'évaluation est jugée à l'aune de la qualité de sa production (les connaissances produites) et son processus de production de connaissances (la méthode de collecte et d'analyse des données et l'argumentaire pour produire le jugement évaluatif).

Dans le contexte des politiques de santé en France que nous avons exposé précédemment, la crédibilité inclut la notion d'acceptabilité, compromis entre deux exigences contradictoires : la production de données jugées de qualité à la fois parce qu'elles documentent la réalité de l'action publique et sont produites en respectant les critères de rigueur des disciplines mobilisées ; l'intégration du point de vue des acteurs, partenaires institutionnels et bénéficiaires de l'action publique.

En effet, depuis la loi du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé, le concept de démocratie sanitaire est omniprésent dans les discours et orientations des politiques de santé. Ce concept fait référence aux droits collectifs des usagers du système de santé et aux droits individuels des personnes malades. Les usagers peuvent, par l'intermédiaire de leurs représentants, donner leur point de vue dans le débat public sur l'organisation et le fonctionnement du système de santé et, plus largement, sur la politique menée dans ce domaine. Depuis 2002, les instances de concertation se sont multipliées et ont été réformées à plusieurs reprises. Révisée par la loi de 2009, la conférence régionale de santé et de l'autonomie (CRSA) est un organisme consultatif qui concourt par ses avis à la politique régionale de la santé. Selon l'article L 1432-4 du code de la santé publique, elle peut faire toute proposition au directeur de l'ARS sur l'élaboration, la mise en œuvre et l'évaluation de la politique de santé dans la région. Bien que ses missions en matière d'évaluation soient assez peu formalisées⁶ et limitées à la production d'un avis sur le rapport d'évaluation du PRS, c'est une instance avec laquelle il faut compter. En effet, dès 2010, la présidente de l'assemblée permanente des CRS plaidait pour l'instauration d'un véritable dialogue avec les décideurs et financeurs du système de santé et une participation effective à l'évaluation du projet régional de santé, faisant de leur association à l'évaluation, un véritable enjeu de la démocratie sanitaire (Devictor, 2010). Dans le domaine médicosocial, le point de vue des usagers est érigé en principe pour l'évaluation des établissements. Aujourd'hui la conception de l'évaluation est mise en tension entre souci de validité scientifique et exigence démocratique.

Par ailleurs, la consultation s'accompagne du devoir de « rendre compte » de ce qui a été fait et donc de soumettre les résultats de l'action publique au regard des citoyens. La motivation à évaluer est ainsi doublement encadrée, par la loi et par la pression que ces instances sont en mesure d'exercer, ces dernières étant bien disposées à faire valoir leur droit. La crédibilité des travaux d'évaluation résulte donc d'un double processus, externe et interne.

3.1.2. La crédibilité de la démarche d'évaluation

La crédibilité des données est indissociable de celle de l'évaluation, incluant le mode d'interprétation des données et l'organisation des relations entre les personnes engagées.

⁶ Avis de la conférence nationale de santé du 29 novembre 2012.

Le processus de production de jugement

Hurteau et al (2012) insistent sur la notion de jugement crédible. L'évaluation est caractérisée par sa capacité à produire un jugement, l'évaluateur étant un guide dans la démarche d'élaboration de ce jugement avec les parties prenantes. Selon ces auteurs, le jugement pour être crédible, c'est-à-dire recevable, doit être juste, refléter la réalité, émaner d'un consensus et être esthétique (cohérent, congruent, lisible).

La qualité de l'argumentaire participe de la crédibilité. A l'image de l'enquête policière ou journalistique, l'évaluation doit amasser un faisceau d'indices (Hurteau et al., 2012 ; Leviton, 2003) dont l'assemblage rigoureux permet de formuler un jugement. De ce point de vue, la pratique évaluative s'analyse comme une pratique argumentative (Perret, 2009c).

L'organisation de l'évaluation

L'organisation et la conduite de la démarche d'évaluation participent de sa crédibilité. Nous avons vu dans nos travaux sur l'évaluation des PRSP, que la définition des missions des différents groupes et la répartition des responsabilités entre les personnes sont des critères de la qualité du processus ; ces caractéristiques figurent dans l'énoncé du principe de responsabilité de la charte de la SFE (2006). Cependant, bien que les standards élaborés par les sociétés professionnelles intègrent ces aspects, le fonctionnement des groupes est assez peu documenté dans la littérature (Baizerman et al., 2012). Nous souhaitons apporter un éclairage à partir de notre expérience de l'évaluation des PRSP.

Le concept français de « dispositif d'évaluation »

Les modes d'organisation ne sont pas identiques selon les pays et milieux concernés. En France, le dispositif d'évaluation désigne « le système des relations à établir dans le cadre de l'évaluation entre ses différents acteurs et participants » (Conseil scientifique de l'évaluation, 1996, p.19) ; il inclut les acteurs et différents groupes de suivi ou de pilotage.

Le terme français « dispositif » évoque un assortiment de pièces ou de mesures⁷, l'agencement de ces éléments ou le mécanisme lui-même⁸. Le dispositif d'évaluation

⁷ « Ensemble de pièces constituant un mécanisme, un appareil, une machine quelconque ; ensemble de mesures prises, de moyens mis en œuvre pour une intervention précise » (dictionnaire Larousse).

concerne à la fois les structures et les procédures. Il n'a pas son équivalent en anglais. Dans la littérature anglophone, il est fait référence à un comité consultatif (*evaluation consultation/advisory group*) qui, basé sur l'expertise de ses membres, conseille l'évaluateur sur la conduite de l'évaluation et l'utilisation des résultats (Baizerman et al., 2012).

La distinction entre comité consultatif et instance d'évaluation

Les cinq fonctions que doit remplir un comité consultatif décrites par Mattessich (2012) diffèrent peu de celles de l'instance d'évaluation décrites par le CSE (1996). Pour Mattessich, le comité doit engager les parties prenantes, garantir la crédibilité externe, concilier les points de vue politiques, promouvoir l'intégrité méthodologique et l'utilisation. Sa responsabilité de dans l'utilisation est clairement affirmée avec une influence directe sur l'utilisation, par l'accomplissement des deux premières fonctions, et indirecte, en contribuant activement aux stratégies de diffusion des résultats. La composition du groupe constitue un enjeu majeur et Mattessich, listant les catégories d'acteurs concernés, souligne les différences de positionnements et d'apports de la part des représentants du commanditaire selon la nature de leur statut au sein de l'institution.

L'instance d'évaluation française est un groupe représentant la diversité des groupes d'intérêts, indépendant du commanditaire et porteur d'une expertise. Dans l'esprit du CSE, c'est un « dispositif qui est le mieux à même de garantir une zone d'autonomie de l'évaluation entre la science et l'action ». Elle est « conçue comme un *arbitre* entre les différents points de vue et non comme un *médiateur* de différents intérêts qu'il conviendrait d'accommoder » (Conseil scientifique de l'évaluation, 1996, p.30).

Ce qui distingue les deux visions, c'est d'une part, la contribution de ces deux groupes dans l'exécution des travaux d'évaluation et, d'autre part, la nature de la personne qui pilote le groupe. Tandis que l'*evaluation consultation/advisory group* assiste l'évaluateur qui en est l'animateur, l'instance d'évaluation pilote les travaux, en exécute tout ou partie, et apporte des réponses aux questions évaluatives assorties de propositions. Cette dernière est animée par une personnalité extérieure à l'institution commanditaire, le « président de l'instance d'évaluation », gage de son indépendance.

⁸ « La manière dont sont disposés les pièces ou les organes d'un appareil ou le mécanisme lui-même ; l'ensemble de moyens disposés conformément à un plan (dictionnaire Robert).

Comme nous l'avons rapporté, dans le secteur de la santé en France, la mise en place d'un groupe d'acteurs, selon le format qui vient d'être décrit, est un principe adopté et intégré dans les pratiques d'évaluation pour assister le chargé d'évaluation d'une institution. Selon les milieux, il est désigné par plusieurs termes : comité de suivi de l'évaluation, comité d'évaluation, groupe d'évaluation ou encore, comité de pilotage, terme dévolu à une instance de décision plus formelle. Si l'appellation « instance d'évaluation » a intégré le vocabulaire, la distinction entre ce groupe et un comité de pilotage est souvent assez floue, ce que traduit le glissement sémantique, car la présidence de ce groupe est souvent confiée à un responsable de l'institution commanditaire (Jabot et al., 2010).

Des divergences de vues et de mises en œuvre existent, s'agissant du statut, de la composition du groupe de suivi de l'évaluation et du mode de fonctionnement du dispositif. Ainsi que nous venons de l'aborder, l'animateur du groupe n'est pas toujours extérieur à l'institution. Par ailleurs, toutes les parties prenantes ne sont pas toujours incluses dans le groupe ou ne figurent pas dans le processus dans son ensemble. Enfin, le moment où l'évaluation est ouverte à la concertation, varie selon les institutions, avec des conséquences sur le rôle des partenaires dans l'évaluation et par extension sur les suites de l'évaluation. Les dispositifs en place pour l'évaluation des PRSP étaient loin d'être identiques dans les régions (cf. analyse des seize régions) et expliquent pour partie l'approbation ou le désaveu des travaux par les différentes parties prenantes.

Enfin, comme nous allons le voir maintenant, la crédibilité tient aux principaux acteurs responsables de la démarche.

3.2. Trois acteurs clés du dispositif d'évaluation

Quels que soient les atouts ou les faiblesses du système d'évaluation, les acteurs jouent un rôle capital pour pallier les carences d'une organisation précaire, faire évoluer les normes, procédures et conceptions et activer les leviers du changement, d'autant plus que se crée une alliance d'intérêts entre trois personnages clés : le chargé d'évaluation, l'évaluateur et le président d'évaluation.

3.2.1. Le chargé d'évaluation, agent de changement

Nous avons constaté, s'agissant de l'évaluation des PRSP et aujourd'hui des PRS, que les personnes chargées de l'évaluation relèvent de deux catégories : décideurs opérationnels avec rang de sous-directeur ou chefs de service ; chargés de mission de niveau hiérarchique moins élevé. Ce choix dépend de l'existence d'un service dédié à l'évaluation et de son envergure, de l'importance affectée à l'évaluation et des compétences disponibles. Là encore, s'observe une mise en concurrence entre le statut des personnes et l'intérêt pour l'activité évaluative.

La conduite de l'évaluation dépend de la connaissance, de la vision et de l'expérience en évaluation, de la représentation du rôle et de la marge de manœuvre dont dispose ou pense pouvoir disposer le chargé d'évaluation.

La composition de l'instance d'évaluation, l'association de ce groupe aux prémices de l'évaluation, le processus d'élaboration du questionnement évaluatif, la nature des relations instaurées avec l'évaluateur, toutes les caractéristiques de la démarche relèvent de son initiative et de ses choix. Les étapes clés de la démarche font généralement l'objet d'une validation par le décideur ou par un comité de pilotage. Celle-ci sera d'autant plus aisée que le chargé d'évaluation bénéficie de la reconnaissance de son savoir-faire, de la confiance du décideur et d'une relative autorité au sein de l'institution. Le respect des standards de bonnes pratiques d'évaluation sont également de sa responsabilité. Il lui incombe d'impulser le dialogue, de favoriser la communication, d'entretenir une dynamique dans la durée. Grâce à son expertise propre (Cousins et al., 2014) ou renforcée par celle de l'instance d'évaluation qui l'accompagne, il fait ou non de l'évaluation, une activité utile aux attentes de la région.

Les cinq évaluations de PRSP analysées présentent cinq figures contrastées du chargé d'évaluation au regard de ces caractéristiques. Nous avons dépeint le modèle du leader/agent de changement qui, du fait de son expertise et de la reconnaissance dont il est crédité, assure la traduction entre l'univers de l'évaluation (acteurs impliqués), l'univers de l'institution (décideurs et proches collaborateurs) et celui de l'action (gestionnaires et professionnels de terrain). Trois conditions (volonté politique, compétences et légitimité) sont indispensables pour l'accomplissement de sa mission. Nous avons observé que dans les cas où l'évaluation est conçue comme une tâche réglementaire ou une tracasserie administrative, le chargé d'évaluation, même s'il dispose d'un réel savoir-faire, ne parvient pas à entraîner les équipes

dans la démarche. Par ailleurs, un chargé d'évaluation dépourvu d'expérience d'évaluation aligne sa pratique à la vision de l'évaluation ou aux routines de son institution.

Aujourd'hui, plusieurs ARS ont fait une place à l'activité d'évaluation et recruté du personnel pour la réaliser. Ces professionnels récemment formés ont acquis les connaissances théoriques et disposent des outils pour mener les travaux. Cependant, la conduite de la démarche nécessite, outre les compétences méthodologiques, un savoir construit au fil des expériences, une connaissance des enjeux associés à l'évaluation et des aptitudes relationnelles pour affronter la complexité des choix et des situations (Tourmen, 2009), ce qui plaide en faveur d'un travail d'équipe associant des professionnels aguerris au fonctionnement de l'institution pour épauler les chargés d'évaluation plus récents dans l'institution.

Cette fonction, encore récente dans les institutions, est en voie de structuration. Dans les collectivités territoriales, les chargés d'évaluation décrivent leur travail comme une activité de « transcodage » de la commande publique auprès de leurs décideurs et se définissent comme des « passeurs d'une culture de l'évaluation » au sein de leur institution (Matyjasik, 2010).

Jouant de leur pluri-appartenance, les chargés d'évaluation sont en position d'articuler les différents espaces et d'organiser la médiation entre plusieurs univers, celui de l'institution avec ses règles, ses procédures et ses contraintes, celui de la mise en œuvre des politiques et celui de l'évaluation.

3.2.2. L'évaluateur : devoir d'indépendance ou devoir d'ingérence ?

Pièce importante du système d'évaluation, l'évaluateur fait lui aussi le lien entre tous les acteurs inclus dans l'arène évaluative, articulant les enjeux institutionnels reflétés dans la commande aux enjeux des acteurs qui l'entourent et à ceux de l'exercice proprement dit. Il endosse de multiples rôles en fonction des tâches à assumer au cours des séquences de l'évaluation, ce qui est généralement une des conditions du succès de l'évaluation (Morabito, 2002) mais parfois source de conflits (Skolits et al., 2009).

Les évaluateurs doivent comprendre le contexte et les paramètres de la décision mais également, le processus de diffusion et de transfert des connaissances voire anticiper les mécanismes d'adaptation des connaissances et les innovations. Leur devis d'évaluation sera

d'autant mieux adapté qu'ils auront identifié précocement les types d'usage, les possibilités de changement et les marges de manœuvre des acteurs, ce qui n'est possible qu'au prix d'une collaboration étroite entre l'évaluateur et les représentants de l'institution commanditaire. Grob (2003) suggère un investissement de l'évaluateur afin qu'il identifie les leaders de pensée au sein de l'institution et devienne l'un d'eux. Cette position militante de l'évaluateur défendue par Grob ouvre un débat sur le rôle de l'évaluateur et les limites de son implication.

Plusieurs éléments interviennent et façonnent les relations entre l'évaluateur et son commanditaire.

Comme nous l'avons rapporté dans le chapitre sur l'évaluation dans le domaine de la santé en France, la vision normative, l'assimilation de l'évaluation aux activités d'audit et d'inspection/contrôle et leur association dans un même service, colorent la perception de l'évaluation par les acteurs institutionnels et conditionnent les relations qu'ils envisagent avec l'évaluateur. Par ailleurs, nous l'avons également évoqué, les activités de conseil se sont largement développées ces dernières années avec une contribution plus substantielle des consultants dans les réorganisations des institutions. Ainsi, les rapports instaurés et la distance recherchée avec l'évaluateur fluctuent sous l'influence de ces deux tendances. Les travaux préparatoires de l'évaluation plus proches des activités de conseil se réalisent dans une relation de proximité entre évaluateur et institution, ce qui interdit la poursuite de la collaboration pour un travail d'évaluation. Dans le cas de travaux sous contrat, une certaine distance est requise, laquelle se définit au gré des besoins, de la dynamique entre les parties prenantes dans le groupe d'évaluation, des collaborations réitérées et des communautés de pensée.

Dans les évaluations de PRSP, dans un cas, la distance géographique entre l'évaluateur et l'institution a limité la forme collaborative du travail ; dans un autre, elle n'a pas empêché des relations étroites. L'implantation régionale de l'organisme évaluateur est à double tranchant d'autant qu'il existe un rapport de tutelle, ce qui est parfois le cas quand l'évaluation est réalisée par des observatoires régionaux de santé ou des instances régionales d'éducation pour la santé. Enfin, le statut public ou privé de l'évaluateur influence les perceptions ; si une certaine méfiance semble exister vis-à-vis des consultants, les laboratoires universitaires sont

l'objet de davantage de confiance mais peuvent souffrir de l'image de la recherche déconnectée de la réalité⁹.

La posture de l'évaluateur et la distance à son objet d'étude dépendent des approches dans lesquelles il s'inscrit, en interaction avec la culture de l'institution avec laquelle il travaille. La variété des approches en évaluation créé un éventail de situations au sein desquelles l'évaluateur occupe plusieurs postures : de l'intégration au programme qu'il évalue et contribue à améliorer (Patton, 2010) à la recherche d'une distance acceptable et/ou souhaitable entre les différents protagonistes de l'évaluation. Entre devoir d'indépendance et devoir d'ingérence, l'évaluateur compose avec son éthique, ses modèles de référence et les attentes de l'institution commanditaire. Le climat relationnel entre l'évaluateur et ses interlocuteurs et le jeu des affects en termes de confiance/défiance participent des mécanismes activateurs des évolutions de modes de pensée, de transformation des repères et d'adoption de référentiels nouveaux.

3.2.3. Le président de l'instance d'évaluation

Selon le CSE, le président de l'instance d'évaluation est le garant de l'objectivité, de la neutralité, de l'expertise et du pluralisme. Bournois et Chanut (2008) se sont intéressés à cette figure emblématique du modèle de « l'évaluation à la française ». Ils soulignent la complexité du rôle, le président s'efforçant de créer des ponts entre les personnes, les idées, les temporalités, de mettre de la cohérence dans la dispersion des matériaux recueillis, des méthodes, des terrains d'exercice et de construire une vision commune. « Entrepreneur d'intelligence collective », il a besoin, pour accomplir sa mission, du crédit de légitimité, d'honnêteté et d'expertise de ses interlocuteurs.

Peu de régions ont choisi une présidence de leur comité d'évaluation de PRSP extérieure à leur institution et à la région¹⁰. Ce choix, à l'initiative des directeurs opérationnels, n'a pu toujours aboutir car il est difficile de réunir les qualités souhaitées (disponibilité, connaissance des enjeux liés à l'évaluation, expertise et appétence pour l'évaluation, extériorité à la région).

⁹ Informations issues de travaux de groupe en sessions de formation et formation/action en ARS.

¹⁰ Nous n'avons identifié que cinq régions (dont deux parmi nos cinq régions étudiées) mais cette information étant difficile à recueillir, ce chiffre ne reflète probablement pas la réalité.

Le statut du président est un élément du processus de crédibilisation de l'évaluation : ce dernier doit pouvoir partager les éléments de la culture du commanditaire sans pour autant être suspect d'une trop grande proximité, pour également bénéficier de la reconnaissance des autres partenaires.

Il joue lui aussi une fonction de médiation entre l'évaluation, hors des frontières de l'institution, et le décideur.

3.2.4. L'alliance ou la répartition des rôles

Les trois types d'acteurs sont, s'ils assument le rôle, des « intermédiaires », professionnels de la médiation, au sens de Nay et Smith (2002), capables d'intervenir sur plusieurs scènes, de se faire les relais entre les groupes, de maîtriser plusieurs cultures, de formuler une représentation commune, d'élaborer des compromis, d'inventer des solutions acceptables pour tous. De cette catégorie générique, Nay et Smith ont isolé deux figures qui se distinguent par leur activité : le « généraliste » dont l'activité consiste à construire du sens commun et le « courtier » à la recherche de solutions acceptables entre des groupes éloignés. Les premiers mènent une fonction de médiation sur le registre cognitif et les seconds sur le registre stratégique.

Les trois acteurs, du fait de leur statut, de leurs compétences et de l'espace qui leur est aménagé, jouent plus spécifiquement sur l'un ou l'autre de ces registres. Le chargé d'évaluation est le mieux placé pour introduire du changement dans son institution, « généraliste » ou « courtier » selon sa fonction et sa légitimité au sein de l'institution. L'évaluateur incontestablement « généraliste » et le président d'instance plus volontiers « courtier », tous deux, *a priori* affranchis des enjeux propres de l'institution commanditaire, ont une fonction de suppléance dans les espaces laissés vacants par leur partenaire associé, le chargé d'évaluation. Nous faisons l'hypothèse que la combinaison des rôles de ces trois acteurs constitue un facteur clé de l'utilisation de l'évaluation.

Ces différents travaux confirment que le système d'évaluation dans toutes ses composantes, y compris l'utilisation, est bien un système dynamique où les forces s'équilibrent entre les facteurs contextuels, les perceptions et attitudes des acteurs et les attributs de l'évaluation. Ils montrent par ailleurs, les interactions entre la capacité d'évaluation (exprimée grâce aux

compétences des acteurs dans la réalisation de l'évaluation) et la culture d'évaluation, véhiculée par les acteurs, qui imprègne progressivement les pratiques de l'organisation.

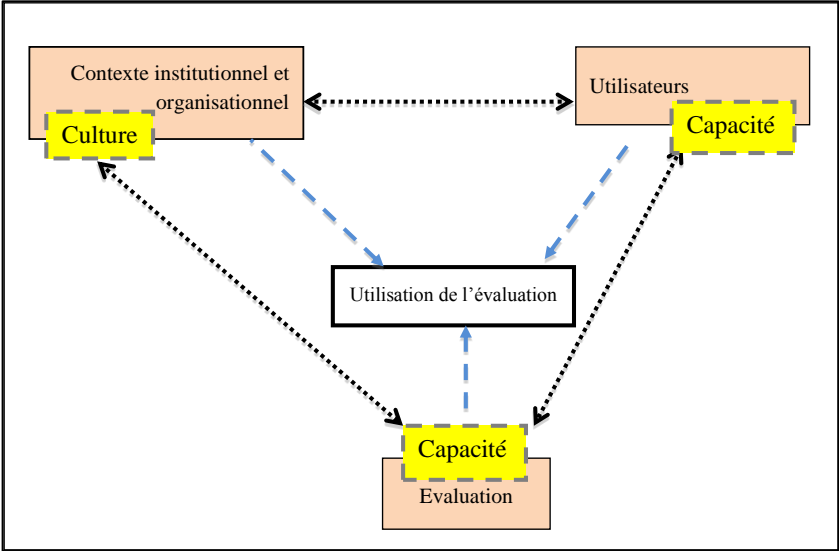


Figure 1. Interactions entre capacité et culture d'évaluation

Conclusion

L'histoire de ces dernières années n'affiche pas l'évaluation comme l'outil privilégié des politiques régionales de santé. Pour autant, les expériences répétées ont apporté leur lot de connaissances et d'apprentissages qui ont sédimenté progressivement construisant une culture d'évaluation. Les retombées de l'évaluation sont plus à chercher du côté des savoirs accumulés et des transformations des pratiques que dans des décisions formelles transformant de façon visible et radicale les politiques.

Qu'attend-on de l'évaluation aujourd'hui ?

Les projets régionaux de santé ont succédé aux plans régionaux de santé publique et représentent l'instrument de mise en œuvre de la politique des agences régionales de santé. Incités réglementairement à produire leur évaluation, les acteurs régionaux ont initié des réflexions et pour bon nombre de régions, démarré des travaux. Cette évaluation affronte des défis majeurs, aussi bien méthodologiques, en raison de la complexité de l'objet, que politiques, compte tenu des enjeux de la réforme. Opter pour une évaluation interne ou externe, intermédiaire ou finale, développer des activités de suivi plutôt que d'évaluation, poser des jalons pour la collaboration avec les instances de concertation, associer l'ensemble de l'agence dans le processus d'évaluation, sont les principaux thèmes en débat.

En parallèle à ces réflexions, des changements substantiels sont annoncés. Depuis déjà plusieurs mois, l'élaboration de la stratégie nationale de santé a ralenti le pilotage national des actions régionales, semant le doute sur le devenir des PRS. La prochaine loi de santé prévoit leur simplification. Les évolutions à venir interrogent le bien-fondé de l'évaluation, posent des défis de méthode (comment définir le périmètre de l'évaluation d'un objet en voie de redimensionnement ?), remettent en question les travaux engagés, voire démobilisent les décideurs pas toujours enclins à l'évaluation.

La fonction évaluative est en voie de structuration dans plusieurs agences, positionnée à un niveau plus stratégique, associée aux activités d'observation et statistiques et disposant de davantage de ressources et de compétences. Pour autant, elle reste toujours concurrencée, et probablement davantage dans les petites structures, par les autres missions régaliennes et prioritaires de l'institution et par le flux continu des commandes nationales. La conception de

la démarche d'évaluation a évolué, à la faveur de la diffusion d'un discours sur la démocratie, vers une pratique plus collégiale avec les autres partenaires.

Le système souffre d'un déficit d'articulations entre les évaluations conduites aux différents échelons. Les évaluations régionales des PRSP n'ont pas été exploitées par le niveau national, pas plus qu'elles n'ont été portées à connaissance des acteurs locaux. De même, les évaluations conduites sur les projets locaux sont exploitées par le niveau régional en tant que documents instruisant des demandes d'allocation budgétaire mais peu capitalisées comme somme de connaissances. Aujourd'hui, l'évaluation constitue un chantier d'envergure pour les ARS avec des interrogations sur la globalité de l'objet (le PRS) et sur ses différentes composantes (schémas, programmes, contrats) déployées à différents échelons. Les régions peinent à définir les contours de l'évaluation du PRS, identifier ou prioriser les sujets à questionner, faire des choix entre différentes formes d'exercice. En revanche, certaines opérations, plus abordables et plus ciblées seraient de nature à apporter des réponses à des problématiques transversales du PRS. Les premières démarches engagées pour l'évaluation des contrats locaux de santé dévoilent l'expression de logiques différenciées selon les acteurs, les enjeux de l'exercice pour le devenir des contrats eux-mêmes et l'équilibre des partenariats institutionnels. L'évaluation des programmes territoriaux de santé, déclinaison du PRS, est une autre alternative pour comprendre et apprécier, à une échelle plus abordable, les dynamiques à l'œuvre dans le processus de territorialisation. D'autres évaluations sur des populations, approches, thèmes ou projets ciblés pourraient, en complément, être programmées pour documenter des modes d'intervention ou des projets innovants. Les productions de ces différents travaux peuvent s'agencer comme autant de parcelles de connaissance éclairant la politique de santé globale. Les ARS ont aujourd'hui la faculté de structurer un programme d'évaluation agrégeant différents travaux.

Quelle évaluation ?

La volonté de répondre aux trois logiques, scientifique, managériale et démocratique, de l'évaluation anime les promoteurs de l'évaluation avec des difficultés récurrentes à résoudre les tensions inhérentes à la cohabitation entre ces trois logiques. La préoccupation managériale prééminente est tiraillée entre la nécessité de rendre compte et le souci de produire des données solides sur les effets de la politique.

Inspirée de la représentation de l'*Evidence-based policy* selon Dobrow (2004), l'évaluation pourrait ainsi être schématisée.

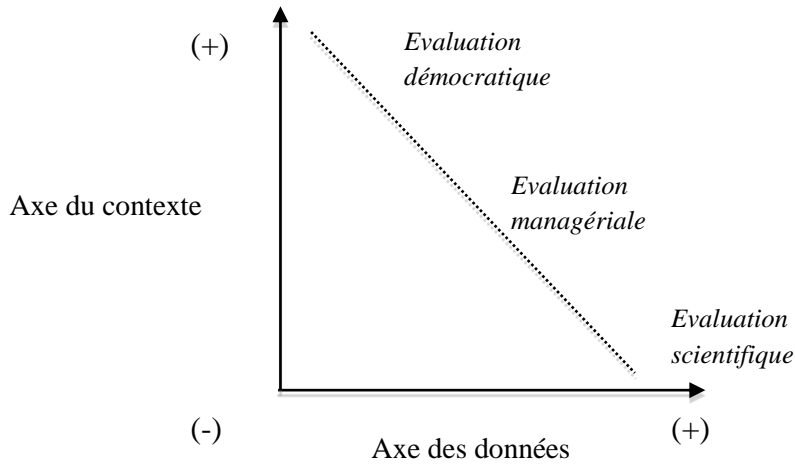


Figure 2. Axes de l'évaluation (d'après Dobrow, 2004)

L'évaluation, en tant que concept et démarche, se nomme au singulier. Sur le versant pratique, elle se conjugue au pluriel. Face à l'étendue du questionnement possible, à la pluralité des besoins et attentes dans les trois domaines (prévention/promotion, soins, médicosocial), il y a matière à conduire de multiples travaux, les uns au service du pilotage de l'action publique, les autres pour étayer la connaissance ou créer un espace de débat. Plus que dans la quête d'un idéal d'évaluation inatteignable, ce peut être dans l'association de plusieurs évaluations au service d'un projet plus global que peuvent se résoudre les tensions entre des logiques contradictoires.

La pluralité des formes d'évaluation convoque une pluralité de types d'expertise. L'évaluation nécessite la mobilisation de savoirs et savoir-faire dans différents domaines. La volonté exprimée des professionnels des ARS de s'impliquer et de se réappropriier l'évaluation incite à une mise en commun et une valorisation des formes d'expertise fragmentées mais néanmoins présentes au sein des institutions afin de faire des services d'évaluation, des lieux de construction de la capacité à faire et de diffusion de la culture d'évaluation.

La réflexion développée dans ce travail peut trouver des prolongements à d'autres terrains d'exercices, qu'il s'agisse des politiques nationales ou des politiques conduites par d'autres institutions, en particulier les collectivités territoriales.

Quelles voies de recherche ?

Les apports des travaux de recherche internationaux sur l'évaluation sont incontestablement d'une grande richesse pour comprendre les enjeux de l'évaluation. Cependant, le champ est ouvert pour développer une recherche sur les conditions de la pratique dans le contexte français.

Plusieurs axes sont envisagés.

Le premier porte sur la commande d'évaluation, étape cruciale de la démarche, dans l'objectif de rapprocher les visions des décideurs et évaluateurs. Il s'agit d'approfondir la compréhension des enjeux de l'évaluation dans le contexte institutionnel, des motivations et des besoins des décideurs et utilisateurs afin d'adapter le travail d'évaluation au plus près de leurs attentes. A cet égard, nous avons engagé un travail de comparaison du cadre des évaluations de plans régionaux de santé publique entre la France et le Québec.

Le deuxième concerne les conditions favorables au développement des capacités et de la culture d'évaluation. Nous avons isolé deux figures particulières du dispositif d'évaluation (le chargé d'évaluation et le président d'instance d'évaluation) qui jouent des rôles distincts et complémentaires. Des études pourraient être envisagées dans trois directions : a) mieux connaître les caractéristiques de ces acteurs en termes de profil, recrutement, motivation et leur cadre d'exercice ; b) analyser la dynamique entre les trois principaux protagonistes de l'évaluation (chargé d'évaluation, président d'évaluation et évaluateur) comme facteur de réussite de l'évaluation ; c) étudier le lien entre la participation des chargés d'évaluation aux groupes de travail, réseaux, formations et leur capacité d'influence dans l'institution.

Le troisième, en lien avec le précédent, concerne un champ récemment défriché en France : la question du partage et du transfert de connaissances. A l'heure où une réflexion émerge en France pour développer le transfert et le partage de connaissances au service de politiques plus efficaces et plus efficientes (Cambon et Alla, 2013), où le courtage en connaissances s'annonce comme une activité prometteuse pour créer des liens entre chercheurs et décideurs (Munerol et al, 2013), le renforcement des compétences des chargés

d'évaluation pour en faire des courtiers en connaissances dans leur domaine d'intervention, l'utilisation des connaissances produites par l'évaluation, est une piste à explorer.

Le quatrième ouvre sur un thème qui fait l'objet d'un intérêt récent en France, l'évaluation d'impact sur la santé (EIS). Méthode au service de politiques favorables à la santé, l'EIS consiste à apprécier les effets potentiels, positifs ou négatifs, d'une action publique, sur la santé de la population afin d'éclairer les décideurs politiques sur les choix futurs. Située en amont d'un projet, cette démarche n'a de sens que si elle donne lieu à une prise de décision afin de faire évoluer le projet en vue d'une atténuation des effets négatifs et d'un renforcement des effets positifs sur la santé. A bien des égards, les questions soulevées sur les conditions de réussite de cette démarche rejoignent la problématique de l'utilisation des évaluations de politiques ou programmes. Le modèle théorique que nous avons élaboré pourrait être appliqué et adapté aux travaux que nous menons actuellement dans ce domaine.

L'utilité de l'évaluation n'est guère plus démontrable que son utilisation bien qu'il semble qu'elle soit davantage perçue par les personnes ayant expérimenté une évaluation. Le développement de la recherche devrait contribuer à donner une meilleure visibilité aux apports de l'évaluation, ainsi renforcer sa crédibilité et, par extension, son utilisation.

Références bibliographiques

- Alkin MC (2004) *Evaluation roots: Tracing theorists' views and influences*. Sage Publications.
- Appleton-Dyer S, Clinton J, Carswell P, et al. (2012) Understanding Evaluation Influence Within Public Sector Partnerships: A Conceptual Model. *American Journal of Evaluation*, 33(4), 532–546.
- Baizerman ML, Fink A and VeLure Roholt R (2012) From consilium to advice: A review of the evaluation and related literature on advisory structures and processes. *New Directions for Evaluation*, 2012(136), 5–29.
- Barbier J-C (2009) La charte de la Société française de l'évaluation : des principes pour les participants à l'évaluation des politiques et des programmes publics, *Actualité et dossier en santé publique*, 69, 30-31.
- Barbier J-C (2010) Éléments pour une sociologie de l'évaluation des politiques publiques en France. *Revue française des affaires sociales*, n° 1-2(1), 25–49.
- Barbier J-C and Matyjasik N (2010) Évaluation des politiques publiques et quantification en France : des relations ambiguës et contradictoires entre disciplines. *Revue Française de Socio-Économie*, n° 5(1), 123–140.
- Barnes M, Matka E and Sullivan H (2003) Evidence, Understanding and Complexity Evaluation in Non-Linear Systems. *Evaluation*, 9(3), 265–284.
- Battesti JP, Bondaz, M, Marigeaud, M et al. (2012) *Cadrage méthodologique de l'évaluation des politiques publiques partenariales*, Inspection générale de l'administration, Inspection générale des finances, Inspection générale des affaires sociales, Paris : IGA, IGF, IGAS.
- Baslé M and Malgrange P (2007) Présentation générale: Aide à la décision publique : développements récents. *Economie & prévision*, n° 175-176(4), i–iv.
- Belton V and Stewart T (2002) *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer Science & Business Media.
- Benamouzig D and Besançon J (2007) Les agences, alternatives administratives ou nouvelles bureaucraties techniques ? *Horizons stratégiques*, 3(1), 10–24.

- Berriet-Sollicec M, Labarthe P and Laurent C (2014) Goals of evaluation and types of evidence. *Evaluation*, 20(2), 195–213.
- Black N (2001) Evidence based policy: proceed with care. *BMJ: British Medical Journal*, 323(7307), 275–279.
- Blake SC and Ottoson JM (2009) Knowledge utilization: Implications for evaluation. *New Directions for Evaluation*, 2009(124), 21–34.
- Bourdin J, André P and Placade JP (2004) *Placer l'évaluation des politiques publiques au coeur de la réforme de l'État*, Rapport d'information fait au nom de la délégation du Sénat pour la planification, Rapport n°392, 30 juin 2004.
- Bournois F and Chanut V (2009) Le Président d'instance, un entrepreneur ? In: Fouquet A et Méasson L (eds.) *L'évaluation des politiques publiques en Europe. Cultures et futurs*, Paris: L'Harmattan, pp.261-271.
- Bowen S and Zwi AB (2005) Pathways to 'evidence-informed' policy and practice: a framework for action. *PLoS medicine*, 2(7), e166.
- Brandon PR and Fukunaga LL (2014) The State of the Empirical Research Literature on Stakeholder Involvement in Program Evaluation. *American Journal of Evaluation*, 35(1), 26–44.
- Bras P-L (2009) La création des agences régionales de santé: notre système de santé sera-t-il encore mieux gouverné? *Droit social*, (11), 1126–1135.
- Cambon and Alla (2013) Transfert et partage de connaissances en santé publique : réflexions sur les composantes d'un dispositif national en France. *Santé Publique*, Vol. 25(6), 757–762.
- Carbonnel L, Jabot F and Paul S (2009) *La fonction Evaluation au sein des futures agences régionales de santé*, Note adressée au secrétariat général des ministères sociaux, non publié.
- Cases C, Grémy I and Perret B (2009) L'évaluation en santé publique en France : diversification et consolidation, *Actualité et dossier en santé publique*, 69, 18-21.
- Castaing C (2012) La territorialisation des politiques de santé. <http://www.leh.fr/edition/>, text, Available from: <http://www.leh.fr/edition/hors-collection/la-territorialisation-des-politiques-de-sante-9782848743691.html> (accessed 9 July 2014).

- Champagne F and Contandriopoulos AP (2009) Eléments d'architecture des systèmes d'évaluation de la performance des services de santé. In: Contandriopoulos AP, Harz Z, Gerbier M, Nguyen A (eds.) *Santé et citoyenneté. Les expériences du Brésil et du Québec*. Les presses universitaires de Montréal, pp.251-274.
- Chen HT (2004) The Roots of Theory-driven Evaluation: Current Views and Origins. In: Alkin MC (ed.) *Evaluation roots: Tracing theorists' views and influences*. Sage, pp.132-152.
- Chen HT (2010) The bottom-up approach to integrative validity: A new perspective for program evaluation. *Evaluation and Program Planning*, 33(3), 205–214.
- Choi BCK, Pang T, Lin V, et al. (2005) Can scientists and policy makers work together? *Journal of Epidemiology and Community Health*, 59(8), 632–637.
- Christie CA (2007) Reported Influence of Evaluation Data on Decision Makers' Actions An Empirical Examination. *American Journal of Evaluation*, 28(1), 8–25.
- Coldefy M and Lucas-Gabrielli V (2012) Le territoire, un outil d'organisation des soins et des politiques de santé ? Évolution de 2003 à 2011. *Questions d'économie de la santé*, IRDES, n°175, p. 1-8.
- Conner RF, Fitzpatrick JL and Rog DJ (2012) A first step forward: Context assessment. *New Directions for Evaluation*, 2012(135), 89–105.
- Conseil scientifique de l'évaluation (1996) *Petit guide de l'évaluation des politiques publiques*. Paris : La Documentation française.
- Contandriopoulos D and Brousselle A (2012) Evaluation models and evaluation use. *Evaluation*, 18(1), 61–77.
- Contandriopoulos D, Brousselle A and Kêdoté N (2008) Evaluating Interventions Aimed at Promoting Information Utilization in Organizations and Systems. *Healthcare Policy / Politiques de Santé*, 4(1), 89–107.
- Contandriopoulos AP, Champagne F, Sicotte C, et al. (2009) L'évaluation de la performance au service d'un pilotage décentralisé du système de santé, *Actualités et dossier en santé publique*, 69, 21-26.

- Contandriopoulos AP, Rey L, Brousselle A, et al. (2011) Evaluer une intervention complexe: enjeux conceptuels, méthodologiques et opérationnels. *Canadian Journal of Program Evaluation*, 26(3), 1-16.
- Contandriopoulos D, Lemire M, Denis JL, et al. (2010) Knowledge Exchange Processes in Organizations and Policy Arenas: A Narrative Systematic Review of the Literature. *Milbank Quarterly*, 88(4), 444–483.
- Cousins JB and Leithwood KA (1986) Current empirical research on evaluation utilization. *Review of educational research*, 56(3), 331–364.
- Cousins JB and Bourgeois I (2014) Cross-Case Analysis and Implications for Research, Theory, and Practice. *New Directions for Evaluation*, 2014(141), 101–119.
- Cousins JB, Goh SC, Elliott CJ, et al. (2014) Framing the Capacity to Do and Use Evaluation. *New Directions for Evaluation*, 2014(141), 7–23.
- Couty E (2006) La culture du résultat dans la fonction publique est-elle compatible avec les enjeux de santé ? *Les cahiers hospitaliers*, 2, 10-12.
- Craig P, Dieppe P, Macintyre S, et al. (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, 337(sep 29 1), a1655–a1655.
- Daniel C, Delpal B and Lannelongue C (2006) *Contrôle et évaluation du fonds d'aide à la qualité des soins de ville (FAQSV) et de la dotation de développement des réseaux (DDR)*. Rapport IGAS, Paris: La Documentation française.
- De Peuter B and Pattyn V (2009) Evaluation Capacity : enabler or exponent of Evaluation Culture ? In: Fouquet A et Méasson L (eds.) *L'évaluation des politiques publiques en Europe. Cultures et futurs*. Paris: L'Harmattan, pp.132-142.
- De Savigny D, Adam T, Alliance for Health Policy and Systems Research, et al. (2009) *Systems thinking for health systems strengthening*. Geneva: Alliance for Health Policy and Systems Research: World Health Organization, Available from: <http://public.eblib.com/EBLPublic/PublicView.do?ptiID=476146> (accessed 22 April 2014).
- DeGroff A and Cargo M (2009) Policy implementation: Implications for evaluation. *New Directions for Evaluation*, 2009(124), 47–60.

- Deleau M, Nioche JP, Penz P et Poinard R (1986) *Évaluer les politiques publiques, Méthodes, déontologie, organization*, Rapport du Commissariat général au plan, Paris : La Documentation française.
- Demeulemeester R and El Khoury M (2011) *Evaluabilité des plans et des programmes de santé*, Etude réalisée pour le Haut Conseil de la santé publique, non publié.
- Devictor B (2010) Les conférences régionales de santé et de l'autonomie (CRSA): qu'en sera-t-il de la démocratie sanitaire en région? *Santé Publique*, 22(1), 121–129.
- Dobrow MJ, Goel V and Upshur REG (2004) Evidence-based health policy: context and utilisation. *Social science & medicine* (1982), 58(1), 207–217.
- Donaldson SI, Christie CA and Mark MM (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Sage Publications.
- Dubois N and Marceau R (2005) Un état des lieux théoriques de l'évaluation: une discipline à la remorque d'une révolution scientifique qui n'en finit pas. *Canadian Journal of Program Evaluation*, 20(1), 1.
- Dubois N and Ouchene H (2010) *Cadre de référence pour le développement conjoint (CSSS/DSP) d'un système d'appréciation de la performance du Plan d'action régional (PAR) et des Plans d'action locaux (PAL) en santé publique du territoire montréalais*. Etude pour l'Agence de la santé et des services sociaux de Montréal, Direction de la santé publique, Québec.
- Dubois CA, Champagne F and Bilodeau H (2009) Historique de l'évaluation. In: Champagne F, Contandriopoulos AP, Brousselle A, et al. (eds.) *L'évaluation: Concepts et Méthodes*. Les Presses de l'Université de Montréal, pp.13-34.
- Duran P, Monnier E and Smith A (1995) Evaluation à la française Towards a New Relationship between Social Science and Public Action. *Evaluation*, 1(1), 45–63.
- Ellen ME, Léon G, Bouchard G, et al. (2013) What supports do health system organizations have in place to facilitate evidence-informed decision-making? A qualitative study. *Implementation science: IS*, 8, 84.
- Ecole nationale de santé publique (2006) *L'évaluation des politiques de santé : l'exemple des SROS*, Module interprofessionnel de santé publique, groupe 13, Available at: http://pmb.santenpdc.org/opac_css/doc_num.php?explnum_id=11964 (accessed 10 juillet 2014).

- Fitzpatrick JL (2012) An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, 2012(135), 7–24.
- Fitzpatrick J, Christie C and Mark MM (2009) *Evaluation in Action: Interviews With Expert Evaluators*. Sage Publications.
- Fleischer DN and Christie CA (2009) Evaluation Use: Results From a Survey of U.S. American Evaluation Association Members. *American Journal of Evaluation*, 30(2), 158–175.
- Fontaine C and Monnier E (2002) Evaluation in France. In: Furubo JE, Rist R and Sandahl (eds.) *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, pp.63-76.
- Fouquet A (2014) L'évaluation des politiques publiques : État(s) de l'art et controverses. *Revue française d'administration publique*, n° 148(4), 835–847.
- Gerhart C and Pélé G (2004) Evaluation nationale des PRAPS de première génération. *Actualités et dossier en santé publique*, 46, 56-60.
- Ginsburg A and Rhett N (2003) Building a Better Body of Evidence: New Opportunities to Strengthen Evaluation Utilization. *American Journal of Evaluation*, 24(4), 489–498.
- Green LW, Glasgow RE, Atkins D, et al. (2009) Making Evidence from Research More Relevant, Useful, and Actionable in Policy, Program Planning, and Practice. *American Journal of Preventive Medicine*, 37(6), S187–S191.
- Grob GF (2003) A Truly Useful Bat is One Found in the Hands of a Slugger. *American Journal of Evaluation*, 24(4), 499–505.
- Guba EG and Lincoln. YS (1984). *Fourth generation evaluation*, Newbury Park, Sage publications.
- Guével M-R and Pommier J (2012) Recherche par les méthodes mixtes en santé publique : enjeux et illustration. *Santé Publique*, 24(1), 23–38.
- Guilbert P, Ardwidson P, Du Roscoat ER, et al. (2009) Evaluation en prévention : les études faites à l'INPES, *Actualité et dossier en santé publique*, 69, 54-56.
- Hanberger A (2011) The real functions of evaluation and response systems. *Evaluation*, 17(4), 327–349.

- Hansen MB and Vedung E (2010) Theory-Based Stakeholder Evaluation. *American Journal of Evaluation*, 31(3), 295–313.
- Haute Autorité de santé (2004) *Réseaux de santé : Guide d'évaluation*, Available at: http://www.has-sante.fr/portail/upload/docs/application/pdf/Reseaux_2004.pdf.
- Haut Conseil de santé publique (2008) *Les politiques régionales de santé publique : Éléments de constat et préconisations dans le contexte de la loi hôpital–patients–santé–territoires*, Available at:<http://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=54>.
- Hawe P, Bond L and Butler H (2009a) Knowledge theories can inform evaluation practice: What can a complexity lens add? *New Directions for Evaluation*, 2009(124), 89–100.
- Hawe P, Shiell A and Riley T (2009b) Theorising Interventions as Events in Systems. *American Journal of Community Psychology*, 43(3-4), 267–276.
- Head BW (2010) Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society, Evidence-Based Policy*, 29(2), 77–94.
- Henry GT and Mark MM (2003) Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*, 24(3), 293–314.
- Hojlund S (2014) Evaluation use in the organizational context - changing focus to improve theory. *Evaluation*, 20(1), 26–43.
- Hunter DJ (2009) Relationship between evidence and policy: A case of evidence-based policy or policy-based evidence? *Public Health*, 123(9), 583–586.
- Hurteau M, Houle S and Guillemette F (2012) *L'évaluation de programme axée sur le jugement crédible*. PUQ.
- Hurteau M, Houle S, Marchand MP, et al. (2012) Le processus de production et de crédibilisation du jugement en évaluation. In: Hurteau M, Houle S and Guillemette F (eds) *L'évaluation de programme axée sur le jugement crédible*. PUQ, pp.77-99.
- Innvaer S, Vist G, Trommald M, et al. (2002) Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of health services research & policy*, 7(4), 239–244.
- Inpes (2010) *Outil de catégorisation des résultats des activités de promotion de la santé et de prévention*, http://www.inpes.sante.fr/outils_methodo/categorisation/guide-fr.pdf.

- Institut national de veille sanitaire (2011) *Dépistage organisé du cancer du sein : guide du format des données et définitions des indicateurs de l'évaluation du programme national*. Saint-Maurice: Institut de veille sanitaire. Available at: <http://www.invs.sante.fr/Publications-et-outils/Guides/Depistage-organise-du-cancer-du-sein-guide-du-format-des-donnees-et-definitions-des-indicateurs-de-l-evaluation-du-programme-national>.
- Israelian A (2008) *La professionnalisation de l'inspection dans le cadre des Missions régionales et interdépartementales d'inspection-contrôle-évaluation*, Mémoire de l'Ecole des hautes études en santé publique. Available at : <http://fulltext.bdsp.ehesp.fr/Ehesp/memoires/iass/2008/israelian.pdf>.
- Jabot F (2004) L'évaluation des programmes régionaux de santé, *Actualités et dossier en santé publique*, 46, 48-55.
- Jabot F (2005) *Etat de lieux de la pratique d'évaluation dans les MRIICE*. Enquête à la demande de la Mission d'animation des fonctions d'inspection, non publié.
- Jabot F (2007) *Regards croisés sur l'évaluation*. Film (DVD1) et entretiens sur l'évaluation (DVD2) Centre de ressources pédagogiques de l'ENSP.
- Jabot F and Demeulemeester R (2005) Les interactions entre les niveaux national, régional et infrarégional dans les programmes de. Santé publique. *Santé Publique*, Vol. 17(4), 597-606.
- Jabot F and Loncle-Moriceau P (2010) Evolution des politiques locales de santé publique en France : quel partage des pouvoirs entre Etat et collectivités locales ? In: Fassin D and Hauray B (eds.) *Etat des savoirs en santé publique*, pp.372–382.
- Jabot F and Marchand AC (2014) L'évaluation des plans de santé publique: un levier de changement pour les politiques régionales de santé en France? *Global Health Promotion*, 21(1 suppl), 64–69.
- Jabot F, Turgeon J and Carbonnel L (2011) The evaluation of the PACA regional public health plan: Reconciling the managerial, scientific and democratic finalities. *Evaluation and Program Planning*, 34(3), 196–205.

- Jabot F, Gerhart C, Marchand AC, et al. (2010) Dispositifs d'évaluation et appropriation des résultats. In: Mouterde F and Trosa S (eds.) *Les nouvelles frontières de l'évaluation, 1989-2009, Vingt ans d'évaluation des politiques publiques en France et demain ?* L'Harmattan, pp. 217-230.
- Jacob S (2005) Réflexions autour d'une typologie des dispositifs institutionnels d'évaluation. *Canadian Journal of Program Evaluation*, 20(2), 49.
- Jacob S and Varone F (2004) Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970-2003). *Politiques et management public*, 22(2), 135–152.
- Jansen MW, Oers HA van, Kok G, et al. (2010) Public health: disconnections between policy, practice and research. *Health Research Policy and Systems*, 8(1), 37.
- Jewell CJ and Bero LA (2008) 'Developing good taste in evidence': facilitators of and hindrances to evidence-informed health policymaking in state government. *The Milbank quarterly*, 86(2), 177–208.
- Johnson K, Greenesid LO, Toal SA, et al. (2009) Research on Evaluation Use A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
- Johnson RB (1998) Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21(1), 93–110.
- Kirkhart KE (2000) Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000(88), 5–23.
- Kundin DM (2010) A Conceptual Framework for How Evaluators Make Everyday Practice Decisions. *American Journal of Evaluation*, 31(3), 347–362.
- L'Horty Y, Duguet E and Petit P (2012) Une évaluation expérimentale d'un micro-programme social. *Revue française d'économie*, Volume XXVII(1), 107–127.
- Lacouette Fougère C and Lascoumes P (2014) L'évaluation : un marronnier de l'action gouvernementale ? *Revue française d'administration publique*, N° 148(4), 859–875.
- Lavis JN, Robertson D, Woodside JM, et al. (2003) How Can Research Organizations More Effectively Transfer Research Knowledge to Decision Makers? *Milbank Quarterly*, 81(2), 221–248.

- Lavis J, Davies H, Oxman A, et al. (2005) Towards systematic reviews that inform health care management and policy-making. *Journal of Health Services Research & Policy*, 10(suppl 1), 35–48.
- Le Bouler S (2003) Panorama de la situation en France : Les institutions, les acteurs, les promoteurs. *Informations sociales*, (110), 26-33.
- Le Bouler S (2009) L'évaluation, ça se prévoit... *Actualités et dossier en santé publique*, 69, 61-63.
- Leca J (1993) L'évaluation dans la modernisation de l'Etat. *Politiques et management public*, 11(2), 161–172.
- Ledermann S (2012) Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159–178.
- Levin-Rozalis M (2003) Evaluation and research: Differences and similarities. *Canadian Journal of Program Evaluation*, 18(2), 1–32.
- Leviton LC (2003) Evaluation Use: Advances, Challenges and Applications. *American Journal of Evaluation*, 24(4), 525–535.
- Leviton LC and Hughes EF (1981) Research On the Utilization of Evaluations A Review and Synthesis. *Evaluation review*, 5(4), 525–548.
- Lewin S, Oxman AD, Lavis JN, et al. (2009) SUPPORT Tools for evidence-informed Policymaking in health 11: Finding and using evidence about local conditions. *Health Research Policy and Systems*, 7(Suppl 1), S11.
- Ling T (2012) Evaluating complex and unfolding interventions in real time. *Evaluation*, 18(1), 79–91.
- Lomas J (2007) The in-between world of knowledge brokering. *BMJ*, 334(7585), 129–132.
- Macintyre S (2012) Evidence in the development of health policy. *Public health*, 126(3), 217–219.
- Mackenzie M, O'donnell C, Halliday E, et al. (2010) Evaluating complex interventions: one size does not fit all. *BMJ. British medical journal*, 340(7743), 401–403.
- Mareschal B, Brans JP and Vincke P (1986) How to select and how to rank projects: the Prométhée method. *European Journal of Operational Research*, Available at:

- <http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/9307> (accessed 13 July 2014).
- Mark MM (2011) Toward better research on—and thinking about—evaluation influence, especially in multisite evaluations. *New Directions for Evaluation*, 2011(129), 107–119.
- Mark MM and Henry GT (2004) The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10(1), 35–57.
- Marmot MG (2004) Evidence based policy or policy based evidence? *BMJ (Clinical research ed.)*, 328(7445), 906–907.
- Mathison S (2005) *Encyclopedia of Evaluation*. Thousand Oaks (CA) Sage Publications.
- Mattessich PW (2012) Advisory committees in contract and grant-funded evaluation projects. *New Directions for Evaluation*, 2012(136), 31–48.
- Matyjasik N (2010) *L'évaluation des politiques publiques dans une France décentralisée : institutions, marché, professionnels*. Thèse pour le Doctorat en Science politique, Université Montesquieu, Bordeaux IV/Sciences Po Bordeaux.
- Matyjasik N (2013) Des évaluations de politiques publiques et des consultants. Émergence d'un champ professionnel et segmentation des rôles. *Revue française d'administration publique*, N° 148(4), 907–921.
- Mays N, Pope C and Popay J (2005) Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research & Policy*, 10(suppl 1), 6–20.
- Migaud D (2014) Les cinq défis de l'évaluation. *Revue française d'administration publique*, N° 148(4), 849–858.
- Ministère de la Santé, de la Jeunesse et des Sports et de la Vie associative, direction générale de la santé (2007) *Situation des GRSP : bilan au 31 juillet 2007*, non publié.
- Ministère de la Santé, de la Jeunesse et des Sports et de la Vie associative, direction générale de la santé (2008) *Évaluation des plans régionaux de santé publique : quelques repères pour les acteurs*, Available at: <http://www.sante.gouv.fr/IMG/pdf/PRSP.pdf>
- Monnier E (1987) *Evaluations de l'action des pouvoirs publics*. Paris: Economica

- Morabito SM (2002) Evaluator Roles and Strategies for Expanding Evaluation Process Influence. *American Journal of Evaluation*, 23(3), 321–330.
- Muir Gray JA (2004) Evidence based policy making. *BMJ: British Medical Journal*, 329(7473), 988–989.
- Munerol L, Cambon L and Alla F (2013) Le courtage en connaissances, définition et mise en œuvre : une revue de la littérature. *Santé Publique*, Vol. 25(5), 587–597.
- Nioche J-P and Poinard R (1984) *L'Évaluation des politiques publiques*. Paris: Economica.
- Ottoson JM (2009) Knowledge-for-action theories in evaluation: Knowledge utilization, diffusion, implementation, transfer, and translation. *New Directions for Evaluation*, 2009(124), 7–20.
- Oxman AD, Lavis JN, Lewin S, et al. (2009) SUPPORT Tools for evidence-informed health Policymaking (STP) 1: What is evidence-informed policymaking? *Health Research Policy and Systems*, 7(Suppl 1), S1.
- Patton MQ (1997) *Utilization-Focused Evaluation: The New Century Text*. Sage Publications.
- Patton MQ (2010) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.
- Pattyn V (2014) Why organizations (do not) evaluate? Explaining evaluation activity through the lens of configurational comparative methods. *Evaluation*, 20(3), 348–367.
- Pawson R and Tilley N (1997) *Realistic Evaluation*. Sage Publications.
- Pawson R (2002) Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, 8(3), 340–358.
- Perret B (1997) Les enjeux épistémologiques de l'évaluation. In: Conseil scientifique de l'évaluation (ed.) *L'évaluation en développement 1996, Rapport annuel sur l'évaluation des pratiques d'évaluation des politiques publiques*. Paris: La Documentation Française, pp.283-312.
- Perret B (2003) L'évaluation des politiques publiques : problématique générale. *Informations sociales*, 110 (9),14-25.
- Perret B (2006) De l'échec de la rationalisation des choix budgétaires (RCB) à la loi organique relative aux lois de finances (LOLF). *Revue française d'administration publique*, n°117(1), 31–41.

- Perret B (2008) *L'évaluation des politiques publiques*. Paris: La Découverte.
- Perret B (2009a) L'évaluation progresse en France, mais sommes-nous mieux gouvernés ? In: Fouquet A and Méasson L (eds), *L'évaluation des politiques publiques en Europe. Cultures et futurs*. Paris: L'Harmattan, pp.163-165.
- Perret B (2009b) Acteurs, objets et méthodes de l'évaluation, *Actualité et dossier en santé publique*, 69, 27-29
- Perret B (2009c) La construction d'un jugement. In: Dagenais C and Ridde V (eds) *Approches et pratiques en évaluation de programme*. Montréal: Presses de l'Université de Montréal, pp.51-67.
- Petticrew M, Whitehead M, Macintyre S, et al. (2004) Evidence for public health policy on inequalities: 1: The reality according to policymakers. *Journal of Epidemiology and Community Health*, 58(10), 811–816.
- Pierru F (2011) L'administration de la santé en fusion : la difficile naissance des Agences régionales de santé. *Pouvoirs Locaux*, 75.
- Pineault R, Tousignant P, Roberge D, et al. (2007) Involving Decision-Makers in Producing Syntheses: The Case of the Research Collective on Primary Healthcare in Quebec. *Healthcare Policy / Politiques de Santé*, 2(4), Available at: <http://www.longwoods.com/content/18867> (accessed 25 May 2014).
- Plottu B and Plottu E (2009) Approaches to Participation in Evaluation Some Conditions for Implementation. *Evaluation*, 15(3), 343–359.
- Plsek PE and Greenhalgh T (2001) The challenge of complexity in health care. *BMJ: British Medical Journal*, 323(7313), 625–628.
- Preskill H, Zuckerman B and Matthews B (2003) An Exploratory Study of Process Use: Findings and Implications for Future Research. *American Journal of Evaluation*, 24(4), 423–442.
- Rickles D, Hawe P and Shiell A (2007) A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health*, 61(11), 933–937.
- Rodriguez-Campos L (2011) Stakeholder involvement in evaluation: three decades of the American journal of evaluation. *Journal of MultiDisciplinary Evaluation*, 8(17), 57–79.

- Rog DJ (2012) When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation*, 2012(135), 25–40.
- Rog DJ, Fitzpatrick JL and Conner RF (2012) *Context: A Framework for Its Influence on Evaluation Practice*, *New Directions for Evaluation*, Number 135. John Wiley & Sons.
- Rogers PJ (2008) Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29–48.
- Rolland C and Pierru F (2013) Les Agences Régionales de Santé deux ans après : une autonomie de façade. *Santé Publique*, Vol. 25(4), 411–419.
- Ross S, Lavis J, Rodriguez C, et al. (2003) Partnership experiences: involving decision-makers in the research process. *Journal of Health Services Research & Policy*, 8(suppl 2), 26–34.
- Rossi PH, Lipsey MW and Freeman HE (2003) *Evaluation: A Systematic Approach*. Édition : 7th Revised edition. Thousand Oaks, CA: SAGE Publications Inc.
- Roué-LeGall A, Lucotte M, Carreau J, et al. (2005) Development of an Ecosystem Sensitivity Model Regarding Mercury Levels in Fish Using a Preference Modeling Methodology: Application to the Canadian Boreal System. *Environmental Science & Technology*, 39(24), 9412–9423.
- Rychetnik L, Frommer M, Hawe P, et al. (2002) Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*, 56(2), 119–127.
- Sabatier Paul (1988) An advocacy coalition framework of policy change and the role of policy-oriented learning therein - Springer. Available from: <http://link.springer.com/article/10.1007/BF00136406#page-1> (accessed 9 April 2014).
- Sanderson I (2000) Evaluation in Complex Policy Systems. *Evaluation*, 6(4), 433–454.
- Saunders M (2012) The use and usability of evaluation outputs: A social practice approach. *Evaluation*, 18(4), 421–436.
- Scemama O and Rumeau-Pichon C (2009) L'évaluation en santé publique à la Haute autorité de santé, *Actualité et dossier en santé publique*, 69, 35-36.
- Scriven M (1980) *The logic of evaluation*. Edgepress Inverness, CA.

- Scriven M (1986) New frontiers of evaluation. *Evaluation Practice*, 7(1), 7–44.
- Sheldon TA (2005) Making evidence synthesis more useful for management and policy-making. *Journal of health services research & policy*, 10 Suppl 1, 1–5.
- Shiell A, Hawe P and Gold L (2008) Complex interventions or complex systems? Implications for health economic evaluation. *BMJ*, 336(7656), 1281–1283.
- Shulha LM and Cousins JB (1997) Evaluation Use: Theory, Research, and Practice Since 1986. *American Journal of Evaluation*, 18(3), 195–208.
- Sicotte C, Champagne F, Contandriopoulos AP, et al. (1998) A Conceptual Framework for the Analysis of Health Care Organizations' Performance. *Health Services Management Research*, 11(1), 24–41.
- Skolits GJ, Morrow JA and Burr EM (2009) Reconceptualizing Evaluator Roles. *American Journal of Evaluation*, 30(3), 275–295.
- Société française de l'évaluation (2006) Charte de l'évaluation des politiques publiques. Available at: http://www.sfe-asso.fr/sfe-evaluation.php?menu_id=191.
- Société française de l'évaluation (2007) Voisinage : de la confusion à la complémentarité. Available at: <http://www.sfe-asso.fr/index.php>.
- Stame N (2004) Theory-Based Evaluation and Types of Complexity. *Evaluation*, 10(1), 58–76.
- Stame N (2010) What Doesn't Work? Three Failures, Many Answers. *Evaluation*, 16(4), 371–387.
- Stevahn L (2005) Establishing Essential Competencies for Program Evaluators. *American Journal of Evaluation*, 26(1), 43–59.
- Stuffelbeam DL (1973) An introduction to the PDK book: Educational evaluation and decision-making. In: Worthen BR and Sanders JR (eds.) *Educational evaluation: Theory and practice*. Belmont, CA: Wadsworth, pp.128-142.
- Tabuteau D (2010) L'expert et la décision en santé publique. *Les Tribunes de la santé*, 27(2), 33–48.
- Tabuteau D (2013) 1983-2013 : les évolutions de la politique de santé. *Journal de gestion et d'économie médicales*, Vol. 31(1), 53–67.

- Touati N (2011) L'évaluation des interventions complexes: quelle peut être la contribution des approches configurationnelles? *Canadian Journal of Program Evaluation*, 26(3), 17-35.
- Tourmen C (2009) Evaluators' Decision Making The Relationship Between Theory, Practice, and Experience. *American Journal of Evaluation*, 30(1), 7-30.
- Trosa S (2010) Comment profiter des expériences étrangères ? L'évaluation des démarches au niveau international. In: Mouterde F and Trosa S (eds.) *Les nouvelles frontières de l'évaluation, 1989-2009, vingt ans d'évaluation des politiques publiques en France et demain ?* Paris: L'Harmattan, pp.124- 135.
- Trosa S and Perret B (2005) Vers une nouvelle gouvernance publique ? La nouvelle loi budgétaire, la culture administrative et les pratiques décisionnelles. *Esprit*, (312), 65-85.
- Valovirta V (2002) Evaluation Utilization as Argumentation. *Evaluation*, 8(1), 60-80.
- Vedung E (2010) Four Waves of Evaluation Diffusion. *Evaluation*, 16(3), 263-277.
- Viveret P (1989) *L'évaluation des politiques et des actions publiques*. Rapport au Premier Ministre. Paris: La Documentation française.
- Walton M (2014) Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning*, 45, 119-126.
- Ward VL, House AO and Hamer S (2009) Knowledge brokering: Exploring the process of transferring knowledge into action. *BMC Health Services Research*, 9(1), 12.
- Weiss CH (1980) Knowledge Creep and Decision Accretion. *Science Communication*, 1(3), 381-404.
- Weiss CH (1997) *Evaluation: Methods for Studying Programs and Policies, 2nd Edition*. 2nd edition. Upper Saddle River, N.J: Prentice Hall.
- Weiss CH (1998) Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, 19(1), 21-33.
- Weiss CH (2005) An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. *American Journal of Evaluation*, 26(1), 12-30.
- Weiss CH and Bucuvalas MJ (1980) Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research. *American Sociological Review*, 45(2), 302.

- Whitehead M, Petticrew M, Graham H, et al. (2004) Evidence for public health policy on inequalities: 2: Assembling the evidence jigsaw. *Journal of Epidemiology and Community Health*, 58(10), 817–821.
- Wholey (2004) Using Evaluation to improve Performance and Support Policy Decision Making. In: Alkin MC (ed.) *Evaluation Roots*, Sage Publications, pp. 267–275.
- World Health Organization (2004) *World report on knowledge for better health: strengthening health systems*. Geneva: World Health Organization.
- Yin RK (2009) *Case Study Research: Design and Methods*. SAGE.
- Zimmerman BJ, Dubois N et al. (2011) En quoi la complexité a-t-elle des répercussions sur l'évaluation ?, *Revue canadienne d'évaluation de programme*, 26, 3, xiii-xx.

RESUME

Contexte : Malgré une volonté affichée traduite dans des textes et procédures, l'évaluation en France n'est que peu intégrée dans le processus de décision. Pourtant, dans le secteur de la santé confronté à de multiples défis, l'évaluation pourrait être une approche utile à la détermination des choix. L'utilisation de l'évaluation dépend d'une pluralité de facteurs, dont les connaissances produites et leur mode de production, les caractéristiques des décideurs et autres utilisateurs et le contexte sociopolitique et institutionnel de l'évaluation.

Objectifs : Les objectifs de ce travail sont d'apprécier l'utilisation de l'évaluation dans les politiques de santé, d'identifier les leviers associés à cette utilisation et d'apprécier la capacité de l'évaluation à satisfaire les attentes et favoriser l'utilisation de ses productions.

Méthode : Une revue de littérature a permis d'enrichir la compréhension du concept d'utilisation, d'identifier les facteurs influents et de faire émerger les problématiques associées. Considérant l'évaluation comme un système d'action complexe, un modèle basé sur une approche systémique a été construit et mis à l'épreuve à travers quatre étapes d'analyse des évaluations des plans régionaux de santé publique (PRSP) : (1) analyse globale de 16 évaluations ; (2) lien processus/utilisation dans une région ; (3) résultats à court/moyen terme dans 9 régions ; (4) études de cas approfondies et analyse multicritères dans 5 régions.

Résultats : Une première analyse a éclairé les enjeux du contexte et la contribution des PRSP à la cohérence des politiques régionales de santé. L'examen du processus d'évaluation dans une région a montré la relation entre finalités, démarche et utilisations de l'évaluation. L'observation des suites de l'évaluation dans neuf régions a identifié différentes formes d'utilisation et les principaux facteurs intervenant. La dynamique de changement a été appréhendée dans la globalité des interactions entre le contexte, les acteurs et l'évaluation dans cinq régions.

Discussion : Les retombées de l'évaluation sont plus à chercher du côté des savoirs accumulés et des évolutions des pratiques que dans des décisions formelles transformant de façon radicale les politiques. Le contexte, la crédibilité de l'évaluation, l'engagement et la motivation des acteurs sont des facteurs déterminants. Des pistes de recherche sont envisagées en vue d'approfondir les conditions du développement des capacités et de la culture d'évaluation, elles-mêmes indispensables à une meilleure exploitation de l'évaluation.

Mots clés : complexité, décision, diffusion des connaissances, évaluation, politiques de santé, utilisation.

ABSTRACT

Title: Evaluation of public policies: a conceptual framework and a study of its use by the decision-makers of the health regional institutions in France

Context: Despite the willingness showed in texts and procedures, evaluation in France is little integrated in the decision making process. However, in the health sector which faces multiple challenges, evaluation should be a useful approach to select choices. The use of evaluation depends on multiple factors such as, knowledge and its production process, characteristics of decision makers and others users, and the political and institutional context of the evaluation

Objective: The objectives were: to assess the use of evaluation on health policies; to identify levers associated with use; to assess the capacity of evaluation to fit with decision-makers needs and to enhance usability.

Method: Literature review allowed to enrich the understanding of the concept of use and to identify the main influent factors as well as the related issues. Regarding evaluation as a complex system, a model based on a systemic approach was built and tested in evaluations of regional public health plans (PRSP) in 4 steps: (1) global analysis of 16 evaluations; (2) relation process/use in on region; (3) use at short/medium term in 9 regions; (4) case studies and multicriteria analysis in 5 regions.

Results: A first analysis enlightened the context and the contribution of the PRSP to the coherence of regional policies. A deeper process analysis carried out in one region pointed out the relation between final aims, management and evaluation use. The examination of evaluation consequences conducted in nine regions identified different forms of use and the major factors associated with them. The dynamic of change has been apprehended as a whole through the interactions between context, users and evaluation in five regions.

Discussion: The effects of evaluation are more obvious in terms of knowledge building and evolution of practice than in radical change of policies. Context, evaluation credibility, actors' commitment and motivation are key factors. Future research should help to better understand how to foster the culture and the capacities of evaluation. These are important prerequisites to a wider use of evaluation.

Key words: complexity, decision-making, evaluation use, health policy, knowledge dissemination.