

## Des barèmes génériques pour évaluer les impacts de la recherche sur la société : l'exemple des impacts politiques

Laurence Colinet<sup>1</sup>, Ariane Gaunand<sup>2,3,\*</sup>, Pierre-Benoit Joly<sup>3</sup> et Mireille Matt<sup>4,5</sup>

<sup>1</sup> INRA Collège de Direction, 75007 Paris, France

<sup>2</sup> INRA, Délégation à l'Évaluation, 75007 Paris, France

<sup>3</sup> Université Paris-Est, LISIS, CNRS, ESIEE Paris, INRA, UPEM, 77454 Marne-La-Vallée, France

<sup>4</sup> INRA, GAEL, 38000 Grenoble, France

<sup>5</sup> Univ.Grenoble Alpes, GAEL, 38000 Grenoble, France

**Résumé** – La recherche agricole doit contribuer à la résolution des multiples défis auxquels est confrontée l'agriculture du <sup>xxi</sup> siècle. Il est donc nécessaire d'évaluer son impact sociétal à cette aune. L'approche ASIRPA (Analyse socio-économique des impacts de la recherche publique agronomique) a été développée à l'Institut national de la recherche agronomique (Inra) dans cet objectif, et permet, par la réalisation d'études de cas standardisées, de mieux comprendre les mécanismes générateurs d'impact, et de rendre compte des impacts sociétaux de la recherche. Cet article présente la méthode développée par ASIRPA pour construire des barèmes de notation d'impact. Ces barèmes sont un des outils qui composent l'approche globale d'évaluation développée dans ASIRPA. Appliqués à une collection croissante d'études de cas, ces barèmes permettent de produire une image robuste de l'impact d'un organisme de recherche. La conception de ces barèmes s'appuie sur la construction d'un cadre analytique, sur son application à des études de cas pilotes, et sur une discussion de la notation de ces cas avec un panel d'experts. Le barème consiste en une échelle de notation de 1 à 5 de l'intensité de l'impact, accompagnée de critères génériques pour chacun des niveaux. Il permet une auto-notation des cas. La démarche est illustrée par le travail effectué pour l'établissement du barème de notation de l'impact politique. L'article conclut en évoquant quelques utilisations possibles de tels barèmes. Il évoque également certaines limites de l'approche et de futures pistes de développement.

**Mots clés** : évaluation *ex post* / impact sociétal / métrique / panel d'experts / ASIRPA

**Abstract** – **Grading scales to assess the impacts of research on society: the example of political impacts.** Agricultural research must contribute to addressing the diversity of challenges faced by agriculture today and in the future, which includes assessing its societal impact. The ASIRPA approach (Socio-economic analysis of public agricultural research impacts) was developed at the French National Institute for Agricultural Research (INRA) to this aim. Based on a series of standardized case studies, the method offers an understanding of impact-generating mechanisms and reports on the societal impacts of research. This paper presents the method developed by ASIRPA to build rating scales. These rating scales are one of the tools developed in the whole ASIRPA approach. They are applied to a growing number of cases to build a robust picture of the impact of a research organization. The design of these scales relied on i) building an analytical framework, ii) applying the framework to pilot cases, and iii) further discussing it with a panel of experts. The resulting metric is a 1 to 5 grading scale measuring the intensity of impact, with generic criteria associated to each level of the scale. We found that the metric could be used for the self-assessment of case studies. We illustrate the approach with the work carried out to develop the metric for political impact. We conclude by highlighting possible uses of the metric we developed and drawing attention to some limitations of this approach as well as to areas for future research.

**Keywords**: *ex post* assessment / societal impact / metrics / expert panel / ASIRPA

\*Auteur de correspondance : [ariane.gaunand@inra.fr](mailto:ariane.gaunand@inra.fr)

## 1 Introduction : les enjeux de la mesure des impacts

L'évaluation des impacts socio-économiques est à l'agenda des grandes institutions de recherche nationales et internationales. L'intérêt pour l'évaluation d'impact est actuellement renouvelé compte tenu d'un double mouvement :

- la pression croissante sur les budgets publics qui conduit les organismes publics à renforcer les arguments démontrant leur utilité sociale ;
- l'orientation d'une part substantielle de l'effort de recherche vers la réponse à de grands défis sociétaux, qui sont complexes et systémiques.

Cette orientation, si elle ne remet pas en cause l'utilité des connaissances, conduit néanmoins à ouvrir la boîte noire de l'impact pour en comprendre les mécanismes et les améliorer.

Cet article se fonde sur une recherche conduite au sein de l'Institut national de la recherche agronomique (Inra) et dont l'objectif est de concevoir une approche d'évaluation *ex post* rendant compte de la diversité des impacts sociétaux de la recherche, au-delà de la simple analyse économique qui constitue l'essentiel des évaluations (Bozeman et Melkers, 1993 ; Donovan et Hanney, 2011 ; Salter et Martin, 2001). L'approche est basée sur une méthode d'études de cas standardisées combinant analyse qualitative et quantitative : la méthode ASIRPA (Analyse socio-économique des impacts de la recherche publique agronomique), développée dans un autre article (Joly *et al.*, 2015a, b). Standardiser le format des cas permet à la fois une description narrative fine du processus d'innovation, et des enseignements génériques dès lors que le nombre d'études de cas est suffisant. Ce dernier objectif a été atteint à l'Inra par la mise en œuvre en routine de la méthodologie ASIRPA par les départements de recherche lors de leur évaluation.

Un premier enjeu d'ASIRPA est de mieux comprendre les mécanismes clés qui transforment les résultats de recherche en impacts, afin d'améliorer ces mécanismes. La méthodologie se donne également pour objectif une description fine et une mesure robuste des impacts – notamment dans leurs dimensions non économiques. Cinq dimensions d'impact, classiquement étudiées dans la littérature (voir Joly *et al.*, 2015a, b pour une revue de littérature complète) et significatives au regard des missions de l'organisme de recherche, ont été retenues et sont explorées systématiquement pour chaque étude de cas : économique, environnemental, territorial et social, politique, et sanitaire. La description qualitative de chacune de ces dimensions, et la mise en place d'une notation permet de comparer les intensités d'impacts entre les cas et ainsi d'enrichir l'analyse transversale ou typologique des chemins d'impacts des cas pour en tirer des enseignements à l'échelle de l'organisme. Adapté du CGIAR (Groupe consultatif pour la recherche agricole internationale), et également utilisé par le CIRAD (Barret *et al.*, 2017), le chemin d'impacts développé par ASIRPA est une représentation graphique des étapes de génération des impacts qui décrit le travail de recherche, le cheminement de la connaissance hors de la sphère académique, sa transformation et son utilisation par les acteurs socio-économiques.

Cet article présente le développement de l'outil ASIRPA de caractérisation des impacts des cas : les barèmes de notation des différentes dimensions d'impacts. Ces barèmes ont été produits avec le concours de panels d'experts. Les principaux résultats sont illustrés ici par des exemples issus du travail effectué sur les impacts politiques. L'article conclut en évoquant quelques utilisations du barème ainsi que certaines limites de l'approche et des pistes futures de développement.

## 2 L'approche ASIRPA et la problématique de la mesure de l'impact non économique

ASIRPA est une approche globale de l'évaluation d'un organisme de recherche par études de cas, qui conduit à collecter de façon systématique une série de données concernant le processus de génération d'impacts, et mobilise trois outils d'analyse complémentaires :

- une chronologie qui identifie les principaux événements du cas ;
- un chemin d'impacts qui représente graphiquement les étapes et acteurs de la génération des impacts à partir de connaissances scientifiques produites. Le chemin d'impact distingue et décrit les caractéristiques de la recherche (situation productive), ses produits primaires (outputs), les impacts de premier niveau, la généralisation de ces impacts au-delà du premier cercle de bénéficiaires (impacts 2), ainsi que les intermédiaires impliqués à chaque étape du chemin d'impact. Le chemin d'impact est essentiel pour déterminer la contribution spécifique de l'Inra au sein des réseaux d'acteurs du chemin, pour analyser le rôle des facteurs contextuels et pour identifier les mécanismes critiques sous-jacents à la génération de l'impact ;
- un vecteur d'impacts, synthétisé sous la forme d'un tableau et illustré par un radar, qui décrit la nature et l'intensité des impacts générés sur les cinq dimensions retenues.

Ces trois outils, conjugués à une narration standardisée par un plan détaillé de rapport, permettent de décrire et d'analyser les mécanismes qui génèrent les impacts, c'est-à-dire les chaînes d'opérations de « traduction » qui mettent en relation les mondes de la recherche et les mondes des utilisateurs (Joly *et al.*, 2015a, b).

ASIRPA a fait le choix d'une investigation systématique des cinq dimensions d'impacts pour chaque cas évalué, quels que soient les objectifs initiaux des recherches. Ainsi, des recherches visant à promouvoir le développement économique sont également observées sous le prisme de leur possible impact social, environnemental, sanitaire ou politique. En termes de collecte de données d'impacts, ASIRPA a également fait le choix de ne pas définir *a priori* une série d'indicateurs pour chacune de ces cinq dimensions, afin de capitaliser sur les données d'impacts effectivement disponibles et pertinentes pour les bénéficiaires eux-mêmes (Joly *et al.*, 2015a, b). L'analyse fine du chemin d'impact permet d'identifier les parties prenantes et leurs rôles dans l'innovation. Ces acteurs sont interrogés (5 à 12 entretiens semi-directifs par étude de cas), et fournissent les données brutes ou descripteurs locaux concernant les impacts de la recherche les plus significatifs.

Pour analyser l'impact à l'échelle d'un organisme (Matt *et al.*, 2017), sur la base d'un portefeuille de cas et des descripteurs locaux propres à chacun, il est nécessaire de développer un outil d'évaluation étalon, un système visant à donner une valeur à ces impacts.

Si des méthodologies existent pour mesurer et donner une valeur à un impact économique, les dimensions non économiques du vecteur d'impacts nécessitent le développement d'une méthodologie particulière (Bormann, 2013, confirmé par Joly *et al.*, 2015a, b). La littérature confirme que l'évaluation des impacts politiques des recherches est particulièrement mal équipée, les méthodes quantitatives n'étant applicables qu'aux politiques locales (Renkow et Byerlee, 2014) ou qu'à certaines dimensions d'impact politiques (Hazell et Slade, 2014).

En effet pour les dimensions non économiques, convertir des descripteurs des cas en des notations comparables de l'intensité d'impact est particulièrement ardu. La notation directe par les bénéficiaires du cas qui les concerne ou la notation à dire d'experts d'un portefeuille de cas ne permettent pas de satisfaire notre objectif d'une analyse à l'échelle de l'institut d'un portefeuille croissant d'études d'impacts. La notation directe par les bénéficiaires reflète leur propre système de valeur et n'est d'évidence pas comparable entre les cas ; la notation à dire d'experts du portefeuille de cas permet d'avancer sur la comparabilité de la notation, mais ne garantit pas une notation comparable lors de l'addition de nouveaux cas.

Notre méthodologie consiste à recourir à des panels d'experts, afin de construire des barèmes permettant de noter de façon homogène et transparente l'impact concernant les dimensions non économiques de notre radar d'impact. Le travail achevé sur la dimension de l'impact politique illustrera la démarche que nous reproduisons sur les autres dimensions (environnementale, sanitaire, territoriale, sociale).

### 3 Les enseignements de la littérature concernant l'utilisation des panels d'experts et l'évaluation des impacts socio-économiques

#### 3.1 Les enseignements de la littérature concernant l'utilisation des panels d'experts

Les panels d'experts sont fréquemment utilisés pour l'évaluation des projets et programmes de recherche, car ils sont gages de crédibilité (Arnold *et al.*, 2005 ; Boaz *et al.*, 2009). Les panels de pairs sont, par exemple, les modalités dominantes d'évaluation de la qualité académique de propositions de recherche. Les panels d'experts peuvent également être utilisés pour porter un jugement sur l'impact sociétal des recherches en l'absence de données facilement accessibles (Ruegg et Feller, 2003). Le dispositif d'évaluation européen fait ainsi appel à des panels mixtes, composés de chercheurs, gestionnaires de programmes, parties-prenantes (Georghiou et Roesner, 2000 ; Molas-Gallart et Davies, 2006) et utilisateurs finaux qui se livrent à ce que Ruegg et Feller (2003) appellent une « revue de la pertinence sociétale » des

projets de recherche européens. Ces panels mixtes fonctionnent de manière identique à celles des panels de pairs (Bozeman et Youtie, 2015).

Riche de ces expériences, la littérature fournit des recommandations concernant le fonctionnement des panels d'experts, pour s'assurer de l'objectivité et de la diversité des opinions exprimées durant la procédure d'évaluation. La littérature discute particulièrement trois points : la préparation du travail du panel, la sélection des experts, et l'explicitation du jugement rendu par le panel.

La préparation par un secrétariat d'une synthèse des informations disponibles (Ruegg et Feller, 2003) est reconnue par de nombreux auteurs comme un dispositif central à la qualité, car elle facilite l'interaction entre les membres du panel (Arnold *et al.*, 2005 ; Boaz *et al.*, 2009).

La sélection des membres du panel est une autre disposition critique, car la composition et la taille du groupe doivent être pensées pour s'assurer d'une discussion constructive entre les membres. Arnold *et al.* (2005), tout comme Langfeldt (2004) ont ainsi observé qu'un certain recouvrement des compétences était nécessaire entre les membres du panel. Dans le cas contraire, le groupe suivait l'avis des quelques personnes spécialistes de la question. Il en résultait une grande division des tâches au sein du panel et une faible interaction entre les experts qui nuisait à la qualité de l'évaluation. Boaz *et al.* (2009) en concluent qu'il ne faut pas tenter d'aborder un large champ thématique, car assurer un recouvrement de compétences impliquerait de construire un panel d'une taille excessive, qui rendrait les échanges plus difficiles. Pour faciliter le dialogue entre les experts, Bormann (2013) recommande d'associer aux panels des parties-prenantes ayant une expérience dans l'utilisation des résultats de la recherche. Il est également important que les experts puissent se prévaloir d'une crédibilité et d'une séniorité équivalente dans la thématique évaluée (Arnold *et al.*, 2005).

Le dernier enjeu est celui de la transparence concernant l'élaboration du jugement rendu par les experts. Arnold *et al.* (2005) remarque que les critères selon lesquels les experts rendent leur jugement ne sont souvent pas explicites et que cela nuit à la transparence et à la robustesse des évaluations (Langfeldt, 2004 ; Samuel et Derrick, 2015). Il conviendrait d'encourager les experts à quitter leur posture « d'autorité » pour expliciter et discuter leurs critères d'évaluation (Arnold *et al.*, 2005).

#### 3.2 La construction du cadre d'analyse des impacts fondé sur une revue de littérature

Notre cadre d'analyse conceptuel s'appuie principalement sur la littérature relative à l'évaluation des impacts sur la dimension considérée (ex. : politique, environnemental...). Des catégories et des sous-catégories d'analyse sont ainsi identifiées. Le cadre conceptuel est ensuite proposé à un panel d'experts, afin qu'ils le modifient, l'enrichissent à partir de leurs expériences et des observations issues d'une collection d'études de cas mise à leur disposition. L'encadré 1 détaille la construction du cadre conceptuel du barème d'impact politique.

### **Encadré 1. Cadre conceptuel de l'impact politique.**

Trois catégories d'impact politique ont été distinguées d'après Almeida et Báscolo (2006) : l'utilisation des connaissances dans le débat public et la négociation des politiques, l'utilisation dans le cycle politique, et les effets sur la percolation des idées à long terme.

Les travaux de Kingdon (1984) nous ont conduit à considérer l'importance sociétale du domaine politique comme une quatrième dimension de l'impact politique, arguant qu'une forte contribution à une politique d'ordre secondaire devait être relativisée par rapport à une forte contribution à une politique de haute importance sociétale.

Des sous-catégories ont été distinguées suite à une revue de littérature sur les processus d'élaboration des politiques (voir Gaunand *et al.*, 2017).

### **Encadré 2. Applications des recommandations théoriques à la consultation du panel d'experts politiques.**

*Composition du panel d'experts pour la construction de la métrique concernant l'impact politique*

Nous avons sélectionné un petit panel de cinq experts qui disposaient d'une expérience dans l'évaluation de l'utilité de la recherche pour les politiques publiques. Tous les experts avaient eu une expérience dans la recherche. Trois d'entre eux venaient de directions d'étude de ministères potentiellement utilisateurs des données de la recherche, deux experts étaient régulièrement impliqués dans l'évaluation de l'impact sociétal des recherches. À l'exception d'un seul, les experts avaient une large vision des politiques environnementales et agricoles conduites dans la dernière décennie, le dernier expert étant un chercheur en évaluation. Tous les experts avaient la même chance de participer aux débats dans la mesure où ils étaient tous familiers des processus d'élaboration des politiques publiques, avaient tous la même séniorité, et n'étaient pas liés par des relations hiérarchiques ou une dépendance vis-à-vis de l'Inra.

*Fournir aux membres du panel des informations complètes et facilement exploitables*

L'équipe d'ASIRPA a choisi cinq études présentant une large diversité d'impacts politiques, tant dans leur nature que dans leur intensité, selon les appréciations des bénéficiaires interviewés. Les descripteurs d'impact politique ont été extraits des études de cas et mis en regard des sous-catégories analytiques du cadre conceptuel construit à partir de la littérature. Les experts ont reçu, pour chacun des cas, outre le tableau du cadre conceptuel complété pour chaque cas avec ses descripteurs, un résumé d'une dizaine de pages présentant le chemin d'impact.

## **4 La méthode de construction du barème**

### **4.1 Composition du panel**

Concernant la sélection des membres des panels de chaque dimension d'impact, nous avons suivi les recommandations de la littérature, en prêtant attention à la taille du panel pour faciliter les échanges, à l'égale séniorité et crédibilité des membres, et au recouvrement de leurs compétences.

### **4.2 Matériel synthétique préparatoire**

L'équipe ASIRPA propose aux experts un cadre conceptuel d'analyse issu de l'analyse de la littérature, ainsi qu'une collection d'études de cas, reflétant une diversité de situations concernant la dimension d'impact pour laquelle la métrique doit être construite. Pour chacune des études de cas, les experts disposent d'un résumé long, décrivant le chemin d'impact et les effets constatés par les bénéficiaires, ainsi que d'une extraction des observations en rapport avec les dimensions du cadre conceptuel d'analyse de l'impact.

Ces deux premières étapes sont détaillées dans l'encadré 2 sur l'exemple du panel d'experts jugeant les impacts politiques.

### **4.3 Travail préparatoire des experts : notation des cas**

Avant la réunion, il est demandé à chaque expert de noter chaque cas sur chaque catégorie du cadre conceptuel sur une échelle de 1 à 5, les uns relativement aux autres. Il leur est également demandé de donner les arguments qui justifient selon eux chacune de leur note (cette note ne sera que provisoire). Nous avons souhaité privilégier les échanges autour de la notation de cas réels plutôt que des échanges autour de critères d'évaluation isolés.

### **4.4 Réunion des experts et critères d'évaluation**

Les experts sont invités à une réunion d'une journée. L'objectif de la réunion n'est pas d'obtenir des experts un consensus sur une note unique pour chaque cas, mais de les amener à expliciter leur logique d'évaluation et de notation en vue de construire un barème. Dans un premier temps, les

experts sont invités à discuter la pertinence des catégories et sous-catégories proposées dans le cadre d'analyse qui leur a été soumis. Dans un second temps, le travail procède par une revue collective des cas. Chaque expert prend connaissance, pour chacun des cas, de la note accordée et des arguments avancés par les autres experts. Ils sont invités à comparer les mérites de chaque cas et à commenter les notes qu'ils ont octroyées individuellement, sur la base des informations et observations disponibles concernant l'impact. La confrontation des arguments, les explications et les reformulations qui en résultent permettent le développement d'arguments plus divers et plus robustes pour fonder l'évaluation.

### **4.5 Analyse des données et construction de l'échelle de notation**

La réunion débouche sur la stabilisation des sous-catégories d'impacts à considérer. Concernant les impacts politiques, certaines des catégories et sous-catégories proposées dans le cadre d'analyse ont été subdivisées ou reformulées par le panel, puis collectivement validées (voir note sous Tab. 1). La dispersion des notes et les arguments avancés par les experts sont analysés pour chaque sous-catégorie.

**Tableau 1.** Le barème d'évaluation de l'impact politique.**Table 1.** *The grading scale for political impact.*

Catégorie	Sous-catégorie	Critères
Utilisation dans le débat public et la formulation de politiques	Qualité et puissance des messages relayés*	Originalité, excellence et nouveauté Traçabilité Robustesse Adaptation aux audiences visées Crédibilité et légitimité
	Concomitance du débat et des agendas politiques	Modification des agendas politiques Production ou médiatisation des connaissances dans une fenêtre d'opportunité politique Portée de la fenêtre de politique créée ou exploitée
	Intensité et qualité de la médiatisation	Volume, cible, diversité, pertinence de la couverture médiatique Intégrité et sélectivité des messages relayés
	Intensité et qualité du débat	Volume du débat Diversité et pertinence des parties du débat
Utilisation dans le cycle politique	Étapes du cycle politique affectées : mise sur agenda, formulation, mise en œuvre, évaluation	Nombre d'étapes du cycle impactées Ampleur des impacts à chaque étape
	Échelles territoriales des impacts sur les différentes étapes du cycle Pertinence et nouveauté de la solution politique conçue*	Exhaustivité et homogénéité des échelles pertinentes affectées Intensité des impacts sur les différentes échelles Nouveauté et originalité de la solution politique Rapidité et niveau d'adoption Niveau d'incorporation des connaissances scientifiques dans la solution
Percolation des idées sur le long terme	Importance de la connaissance dans le débat	Nature et pluralité des effets de la connaissance : évolution des termes, des positions, des tensions Nature du débat (public/interne, ponctuel/croissant)
	Percolation des idées dans des débats et études ultérieures et arènes distantes Pertinence et intégrité des idées à long terme	Vitesse et durabilité de la percolation Importance et exhaustivité des arènes influencées Résilience des idées sur le long-terme Facteurs protégeant l'intégrité des idées : renommée des chercheurs/institut, continuité de l'implication, traçabilité des idées
Importance sociétale du domaine politique considéré	Gravité potentielle et caractère systémique des enjeux	Multiplicité et imbrication des enjeux Gravité
	Ampleur de la population et politique affectée Émotion publique	Taille et nature de la population affectée Échelle du domaine politique Ampleur de l'émotion publique et politique Occurrence de crises Existence d'un consensus social sur le problème et/ou la solution Régularité et ampleur de la médiatisation

\* Ces deux sous-catégories ont été ajoutées par rapport au cadre d'analyse initialement proposé aux experts.

Sur la base de la revue collective et de la confrontation par les experts des arguments fondant leur jugement, nous dérivons des critères d'évaluation génériques pour chaque sous-catégorie (voir [Tab. 1](#) pour un exemple sur les critères d'évaluation de l'impact politique). Les experts justifient généralement leur notation par plusieurs arguments combinés.

Le barème de notation final est ce [tableau 1](#), auquel est appliquée une échelle de Likert, à partir des critères détaillés établis par les experts pour chaque sous-catégorie d'impact (exemple : originalité, excellence et nouveauté → note de 1 à 5 ; Traçabilité → note de 1 à 5...).

L'application du barème d'impact politique est illustrée sur le cas de l'alerte sur le bisphénol A, dans l'[encadré 3](#).

**Encadré 3. Application du barème d'impact politique sur un cas ASIRPA : « l'alerte sur le bisphénol A ».**

Les recherches de l'Inra ont montré que le bisphénol A, perturbateur endocrinien reconnu, utilisé dans la fabrication de plastiques et résines, peut s'extraire de ces matériaux et traverser les barrières intestinales, cutanées et hépatiques. Cette alerte sur le bisphénol A a eu des répercussions politiques importantes. Les impacts politiques sur la catégorie « Utilisation dans le débat public et la formulation de politiques » ont été notés 4 sur 5. Les arguments jugés à l'aune des critères du barème sont les suivants :

Qualité et puissance des messages relayés	<p>Les résultats de recherche mettent à l'agenda politique de nouvelles questions : alerte concernant le passage du BPA par la barrière cutanée.</p> <p>Les résultats de recherche de l'Inra sur les barrières cutanées, hépatiques et digestives sont explicitement mentionnés dans les médias et débats.</p> <p>La faiblesse de certains messages scientifiques affecte leur crédibilité : l'ANSES juge nécessaire la réplication des travaux. Mais les résultats servent de caution scientifique: leur statut scientifique est utilisé par les députés pour interroger les méthodes d'évaluation des agences.</p> <p>L'INSERM est le principal référent d'expertise pour le ministère de la santé, ce qui affecte la légitimité de l'Inra.</p>
Concomitance du débat et des agendas politiques	<p>Les résultats de recherche sont délivrés et médiatisés dans une fenêtre d'opportunité politique nationale et européenne : les résultats de l'Inra sont publiés en 2010, alors que la première évaluation des risques par l'EFSA a été réalisée en 2006 et qu'une série de saisines de l'ANSES est en cours (2009–2010).</p> <p>Ces résultats contribuent au déclenchement d'une commande d'une expertise collective commandée à l'INSERM sur la Reproduction et l'Environnement (en 2010).</p>
Intensité et qualité de la médiatisation	<p>Forte couverture médiatique (150 articles dans l'année de publication des résultats) à destination d'un public très large, professionnel, grand public et politique.</p> <p>Faible distorsion des messages, malgré une confusion entre les notions de danger et de risque.</p>
Intensité et qualité du débat	<p>Forte contribution des résultats de recherche aux débats au-delà de la sphère sectorielle, au niveau national et européen.</p> <p>Prise de positions des parties prenantes principales : françaises (parlement, ministères, ANSES, ONG – ex. : Réseau Environnement santé, secteur privé – ex. : PlasticEurope) ou européennes. Renforcement des positions des acteurs souhaitant une révision des modalités d'évaluation des perturbateurs endocriniens.</p>

**Tableau 2.** Notes (/5) obtenues pour l'impact politique sur l'ensemble des 23 cas ASIRPA notés en 2016.

**Table 2.** Marks of political impact assigned to the 23 ASIRPA cases scored in 2016.

	Moy	Min	Max	Nb Cas > 1/5
Impact politique global	3,0	1,5	4,8	23
Utilisation dans le débat public	2,6	1	5	16
Mise en œuvre de la politique publique	3,6	2	5	23
Percolation à long terme des idées	2,4	1	4	17
Importance sociétale du domaine considéré	3,1	1	5	21

Le barème de notation est ensuite inclus dans le guide méthodologique ASIRPA pour l'étude d'un cas et mis en œuvre en routine au sein de l'Inra pour la conduite des études de cas d'impact, au même titre que les autres barèmes établis autour des cinq dimensions d'impact (économique, politique, environnemental, sanitaire, sociétal et territorial). La notation n'est pas réalisée par le chercheur directement concerné par les recherches décrites dans le cas, mais par un chercheur tiers sur la base des informations fournies. Nous vérifions que le barème de notation offre une discrimination de notation selon l'intensité de

l'impact et qu'il est assez robuste pour s'appliquer à la diversité des situations produites par les études de cas.

Par exemple, sur les 23 études de cas présentant un impact politique et notées en 2016, le barème d'impact politique s'est révélé opérationnel et assez générique pour couvrir la diversité des situations rencontrées (depuis l'approbation des produits pour le biocontrôle, jusqu'à la conception de nouveaux emballages alimentaires non toxiques). Il a permis de discriminer les 23 cas selon les quatre dimensions de l'impact politique (voir [Tab. 2](#)).

## 5 Enseignements de l'utilisation des barèmes

### 5.1 L'utilisation pour l'auto-évaluation

Utiliser un barème commun pour auto-évaluer l'impact sociétal des recherches, à partir de descripteurs fournis par les utilisateurs et bénéficiaires finaux, fournit un cadre indépendant de la variabilité individuelle d'un comité d'experts *ad hoc*. Certes, si les mêmes cas pilotes avaient été fournis à un panel composé d'experts différents, ce panel aurait produit un barème différent. Néanmoins, le cadre d'analyse conceptuel issu d'une analyse de la littérature et fournit aux experts en amont constitue un ancrage qui limite la subjectivité des analyses conduites. Par ailleurs, utiliser un barème invariable permet de s'affranchir de la consultation systématique de panels à chaque nouveau cas, ce qui renforce la robustesse de l'évaluation entre les cas puisque le biais de jugement lié à la composition de panels différents est évité.

Le barème, co-construit avec les experts, améliore la transparence des critères fondant l'évaluation, qui deviennent alors visibles donc discutables. Ces barèmes ne réduisent pas l'impact à une note issue d'une boîte noire mais permettent au contraire d'enrichir l'appréhension des dimensions d'impacts non économiques. Cette quantification, au même titre que celle de l'impact économique couramment pratiquée, ne peut effectivement pas rendre compte de la complexité des processus de production et circulation des connaissances. Ceci-dit, standardiser la quantification grâce aux barèmes augmente la lisibilité des dimensions d'impacts non économiques.

Ces barèmes d'évaluation ont d'autres avantages : ils limitent les coûts des évaluations puisqu'ils évitent la consultation systématique d'un panel d'évaluateurs à chaque nouveau cas étudié ; il permet une évaluation plus rapide puisque les cas supplémentaires peuvent être évalués sans attendre qu'un nombre minimal d'études soit disponible pour convoquer un panel d'experts.

Toutefois, ces barèmes ne doivent pas être considérés comme définitifs : l'évaluation régulière de nouveaux cas peut fournir des situations imprévues et donc des opportunités d'amélioration. Un deuxième panel sera consulté lorsque 50 cas auront été évalués pour réviser le barème établi, avec le double objectif de vérifier la qualité de l'application des critères et la robustesse des évaluations conduites d'une part, et d'améliorer le barème en cas de nécessité d'autre part.

### 5.2 L'intérêt d'une explicitation des valeurs sous-tendant l'évaluation

Les experts, comme n'importe quelle partie prenante, tendent à traduire dans l'évaluation de l'impact sociétal leurs points de vue personnels concernant les rôles de la recherche et les valeurs de ses résultats pour la société. Il est donc virtuellement impossible pour des parties prenantes différentes d'avoir exactement le même classement de la valeur de différents impacts sociétaux (Spaapen et Van Drooge, 2011). Ce raisonnement milite en faveur de notre objectif visant à la construction d'un barème d'évaluation des impacts sociétaux qui soit indépendant du biais introduit par les intérêts de panels

d'experts différents. Dans la construction du barème, l'étape de discussion parmi les membres du panel permet de réduire les incohérences dans l'évaluation due à une mauvaise compréhension des faits, tout en conservant la divergence due à l'expression des valeurs personnelles des experts (Wooding *et al.*, 2014).

La méthode de consultation que nous avons développée conduit les experts à révéler le système de valeurs complexe enchâssé dans le processus d'évaluation, et assis sur une combinaison de critères d'évaluation interdépendants (voir l'encadré 4 pour les valeurs révélées par le panel d'experts sur les impacts politiques). Ces valeurs ont permis la définition des critères de notation. Ainsi, les critères d'évaluation retenus dans ces barèmes sont transparents, et donc rendus discutables par l'ensemble de la société, contrairement à des jugements portés à dire d'experts au moyen d'indicateurs quantitatifs par exemple. L'utilisation de ces barèmes ne s'oppose pas à la mise en œuvre de méthodes participatives d'évaluation, mais permet éventuellement de les équiper (Chevalier et Buckles, 2013). En effet, conjuguée à la plus grande lisibilité des dimensions d'impacts non économiques conférée par un système de notation comparable sur l'ensemble des dimensions d'impacts, cette transparence de jugement favorise le développement d'un débat démocratique autour des rôles et des impacts de la recherche publique. Le dialogue sociétal ouvert peut aborder le jugement porté sur les impacts d'un cas donné, les parties prenantes interrogées, les descripteurs collectés ; le débat peut également porter sur les critères du barème utilisé pour juger une dimension d'impact donné sur l'ensemble des cas.

### 5.3 L'utilisation pour l'apprentissage

L'instruction des cas ASIRPA associe en pratique des équipes de chercheurs, d'ingénieurs et de techniciens des unités de recherche directement concernées par l'étude de cas. Cette équipe est responsable de la reconstitution historique du chemin d'impact et de la collecte des données d'impact auprès des bénéficiaires finaux extérieurs à l'institut.

Pour chaque étude de cas, les impacts sont explorés dans les cinq dimensions retenues par ASIRPA, à l'aide des barèmes associés. Ces barèmes contribuent à approfondir les périmètres de définition de chacune des dimensions d'impact socio-économique. Ainsi, l'application systématique du barème politique aux recherches dont l'impact politique n'était pas la finalité première, permet de rendre compte de retombées inattendues de la recherche. La méthodologie d'auto-évaluation des impacts selon un barème fourni par l'équipe donne des résultats intéressants pour ce qui est de la compréhension par les chercheurs de la nature et de la diversité des impacts (voir l'encadré 5 pour une illustration sur l'utilisation du barème d'impacts politiques). Les chercheurs sont par ailleurs encouragés à enregistrer l'impact tel que le décrivent les personnes affectées, et à ne pas se limiter aux descripteurs actuels, afin de fournir les éléments pour un enrichissement ultérieur du barème.

La série d'études de cas conduite à ce jour semble toutefois suggérer quelques difficultés persistantes dans l'utilisation de ces barèmes par les chercheurs. Comme le montre le tableau 2 sur les impacts politiques, la catégorie « utilisation dans le cycle politique » a été plus systématiquement explorée, et a

**Encadré 4. Les valeurs concernant le rôle de la recherche attachées dans le barème de mesure de l'impact politique.**

Les experts privilégient dans l'échelle de notation les valeurs suivantes concernant la contribution de la recherche :

- l'affiliation visible des résultats de recherche à l'organisme ;
- l'intégrité des messages de la recherche (exactitude, exhaustivité, actualité) au fil de leur diffusion dans le débat et la mise en œuvre des politiques publiques, et le maintien au fil de temps de cette intégrité. Affiliation et intégrité sont facilitées par le maintien sur un temps assez long de l'implication des mêmes acteurs de recherche, qui protègent l'intégrité et la crédibilité du message scientifique d'une possible distorsion tactique par des parties prenantes. La révision fréquente des objectifs de recherche permet également de prendre en considération les évolutions du contexte et ainsi de maintenir la pertinence des résultats scientifiques au fil du temps. De même, le soutien continu apporté par l'institut aux chercheurs impliqués dans la médiation scientifique et la valorisation des résultats, sont ainsi reconnus;
- l'implication de la recherche sur l'ensemble du cycle de la politique publique. La recherche sur le long terme est susceptible d'apporter un éclairage aux différentes étapes, depuis la mise sur l'agenda de nouveaux sujets, jusqu'à l'évaluation des effets de la politique, en incluant également la formulation de politiques, la prise de décision et la mise en œuvre. Les interventions politiques «isolées» sont moins valorisées par les experts.

La pertinence des résultats de recherche par rapport au niveau de décision de la politique. Ceci inclut la position de l'organisme de recherche et sa crédibilité dans le domaine de politique considéré, la pertinence des actions de diffusion des connaissances : pertinence temporelle vis-à-vis de l'agenda politique, spatiale et hiérarchique vis-à-vis de l'échelle de décision. La contribution de la recherche à l'impact est plus forte si la connaissance scientifique est diffusée sous une forme actionnable au moment où les décisions sont prises, ce qu'[Almeida et Báscolo \(2006, p. 9\)](#) décrivent comme des points d'entrée de la recherche dans le cycle politique. Un facteur essentiel est la crédibilité et la force des messages de la recherche, et la diversité des parties-prenantes auprès desquelles la connaissance est diffusée (grand public, responsables et artisans des politiques sectorielles, décideurs publics et élus).

reçu en moyenne une note plus élevée que les autres catégories. La catégorie relative à la percolation à long terme des idées a en particulier reçu des notes plus faibles, ce qui peut être dû au moindre intérêt des chercheurs dans l'exploration d'un impact plus distant, pour lequel la contribution de la recherche est plus diluée et leur légitimité à revendiquer un rôle plus faible, mais

**Encadré 5. Élargir la vision de la contribution de la recherche à l'impact politique.**

Pour la majorité des chercheurs, la définition de l'impact politique de la recherche se résumait à une utilisation directe des résultats par les fonctionnaires en charge de la politique sectorielle en vue de la mise en œuvre des politiques. L'application systématique du barème d'impacts politiques construit aux 41 cas ASIRPA disponibles, a conduit les chercheurs à considérer les quatre catégories analytiques décrites dans le barème ([Tab. 1](#)). Les chercheurs ont ainsi réalisé que leurs recherches pouvaient avoir des effets auxquels ils ne s'attendaient pas, en particulier pour les catégories d'impact qui ne s'inscrivaient pas dans leurs objectifs initiaux de recherche. Dans certaines situations, ils ont pu mettre en lumière des impacts sur une plus grande diversité d'utilisateurs finaux.

également aux difficultés relatives à la collecte des informations pour analyser cette catégorie. De même, les données relatives à l'utilisation dans le débat public, et en particulier la médiatisation, peuvent être coûteuses à rassembler lorsque aucune revue de presse n'a été effectuée.

#### 5.4 Un retour sur les enjeux liés à l'établissement des barèmes

L'exploration systématique pour chaque étude de cas de l'ensemble des dimensions d'impact les plus significatives pour la mission de l'organisme de recherche, la description qualitative pour chacune de ces dimensions de la nature des impacts, et la mise en place d'une notation permettant de comparer les intensités d'impacts entre les cas a pour principal objectif d'enrichir l'analyse des types de chemins d'impacts représentatifs de l'organisme. Une telle analyse concernant l'Inra a été récemment conduite ([Matt \*et al.\*, 2017](#)).

Les barèmes que nous avons développés sont génériques et portables ; Ils peuvent être appliqués pour l'évaluation des impacts sociétaux d'un organisme de recherche dans la mesure où ses missions sont relatives aux cinq dimensions que nous avons retenues (économique, politique, sanitaire, environnemental, territorial-social). Néanmoins, la mise en œuvre d'ASIRPA à l'Inra démontre qu'une certaine pratique est nécessaire pour appliquer les barèmes aux études de cas : des « référents impacts » ont été formés à la méthodologie ASIRPA et accompagnent les chercheurs en particulier dans la collecte des descripteurs d'impacts auprès de bénéficiaires extérieurs, et l'application des barèmes.

Il doit être souligné que les barèmes que nous avons construits n'ont pas vocation à éclairer l'allocation des fonds à la recherche dans la mesure où le voyage entre l'initiation de la recherche et sa traduction en un impact socio-économique prend plusieurs décennies (19 ans en moyenne sur le portefeuille des 41 cas analysés à l'Inra). De plus, l'analyse des chemins d'impacts met en lumière l'omniprésence des ruptures de trajectoire dans ces voyages d'innovation,



souvent liées à des éléments facilitateurs ou inhibiteurs apparaissant dans le contexte ou dans l'agencement du réseau d'acteurs, qui rendent tout à fait inopérants l'utilisation de la notation d'impact *ex post* à des fins d'allocation de ressources *ex ante*.

## 6 Conclusions

Nous avons développé une méthodologie qui permet de construire un barème pour l'évaluation de l'impact sociétal des recherches, à partir d'une collection d'études de cas, de données concernant l'impact fournies par les utilisateurs finaux, et d'une revue de la littérature. Un barème ainsi construit permet de comparer les études de cas autour de critères similaires concernant les impacts, d'enrichir ainsi l'analyse des chemins d'impact, et d'en tirer des enseignements à l'échelle d'un organisme de recherche. La procédure pour la construction du barème suit les étapes suivantes :

- construction d'un cadre conceptuel à partir d'une revue de la littérature et mise en forme des données extraites d'une collection d'études de cas pilotes selon ce cadre conceptuel ;
- composition et consultation d'un panel d'experts pour raffiner le cadre conceptuel d'abord, exprimer lors de l'examen des études de cas pilotes les éléments qui sous-tendent leur jugement ensuite, et exprimer des notations accompagnées de critères de jugement explicites ;
- traitement des données issues de cette consultation pour la production d'un barème de notation ;
- validation par le panel du barème et test du caractère opérationnel sur une collection élargie de cas.

Dans l'état actuel de nos recherches les barèmes ainsi construits se sont révélés opérationnels en routine et assez génériques pour une auto-évaluation de la diversité des situations rencontrées. Ils devront toutefois être régulièrement réexaminés par des panels d'experts externes, pour vérifier la qualité des évaluations conduites, et faire évoluer les barèmes si des situations non prévues sont rencontrées ultérieurement.

## Références

- Almeida C, Báscolo E. 2006. Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cadernos de Saúde Pública* 22: S7–S19. DOI: [10.1590/S0102-311X2006001300002](https://doi.org/10.1590/S0102-311X2006001300002).
- Arnold E, Clark J, Muscio A. 2005. What the evaluation record tells us about European Union Framework Programme performance. *Science and Public Policy* 32: 385–397. DOI: [10.3152/147154305781779335](https://doi.org/10.3152/147154305781779335).
- Barret D, Blundo-Canto G, Dabat MH, Devaux-Spatarakis A, Faure G, Hainzelin E, *et al.* 2017: IMPRESS (IMPact des REchercheS au Sud). Guide méthodologique sur l'évaluation des impacts de la recherche agronomique dans les pays du sud. CIRAD, Montpellier. Disponible sur <http://agritrop.cirad.fr/586223>.
- Boaz A, Fitzpatrick S, Shaw B. 2009. Assessing the impact of research on policy: a literature review. *Science and Public Policy* 36: 255–270. DOI: [10.3152/030234209X436545](https://doi.org/10.3152/030234209X436545).
- Bornmann L. 2013. What is societal impact of research and how can it be assessed? a literature survey. *Journal of the American Society for Information Science and Technology* 64: 217–233. DOI: [10.1002/asi.22803](https://doi.org/10.1002/asi.22803).
- Bozeman B, Melkers J. 1993. Evaluating R&D impacts: methods and practice. New-York: Springer Science + Business Media, 303 p.
- Bozeman B, Youtie J. 2015. Socio-economic impacts and public value of government-funded research: lessons from four US National Science Foundation initiatives. Presented at the ImpAR Conference. INRA : Paris, France.
- Chevalier JM, Buckles D. 2013. Participatory action research: theory and methods for engaged inquiry. New-York: Routledge, 474 p.
- Donovan C, Hanney S. 2011. The “Payback Framework” explained. *Research Evaluation* 20: 181–183. DOI: [10.3152/095820211X131118583635756](https://doi.org/10.3152/095820211X131118583635756).
- Gaunand A, Colinet L, Joly PB, Matt M. 2017. Counting what really counts? Assessing the political impact of science. *Journal of Technology Transfer*. DOI: [10.1007/s10961-017-9605-9](https://doi.org/10.1007/s10961-017-9605-9).
- Georghiou L, Roesner D. 2000. Evaluating technology programs: tools and methods. *Research Policy* 29: 657–678. Available from [https://doi.org/10.1016/S0048-7333\(99\)00094-3](https://doi.org/10.1016/S0048-7333(99)00094-3).
- Hazell P, Slade R. 2014. Policy Research: the search for impact. In: *Workshop on best practice methods for assessing the impact of policy-oriented research: summary and recommendations for the CGIAR, Independent Impact Assessment Report*. Washington D.C.
- Joly PB, Colinet L, Gaunand A, Lemarié S, Laredo P, Matt M. 2015a. Évaluer l'impact sociétal de la recherche pour apprendre à le gérer : l'approche ASIRPA et l'exemple de la recherche agronomique. *Gérer et Comprendre – Les Annales des Mines. L'épreuve des faits* 122: 31–42.
- Joly PB, Gaunand A, Colinet L, Larédo P, Lemarié S, Matt M. 2015b. ASIRPA: a comprehensive theory-based approach to assessing the societal impacts of a research organization. *Research Evaluation* 24: 1–14. DOI: [10.1093/reseval/rvv015](https://doi.org/10.1093/reseval/rvv015).
- Kingdon J. 1984. Bridging research and policy: agendas, alternatives, and public policies. New-York: Harper Collins. ed. New-York: Longman.
- Langfeldt L. 2004. Expert panels evaluating research: decision-making and sources of bias. *Research Evaluation* 13: 51–62. DOI: [10.3152/147154404781776536](https://doi.org/10.3152/147154404781776536).
- Matt M, Gaunand A, Joly PB, Colinet L. 2017. Opening the black box of impact–Ideal-type impact pathways in a public agricultural research organization. *Research Policy* 46: 207–218. DOI: [10.1016/j.respol.2016.09.016](https://doi.org/10.1016/j.respol.2016.09.016).
- Molas-Gallart J, Davies A. 2006. Toward theory-led evaluation: the experience of European Science, Technology, and Innovation policies. *American Journal of Evaluation* 27: 64–82. DOI: [10.1177/1098214005281701](https://doi.org/10.1177/1098214005281701).
- Renkow M, Byerlee D. 2014. Assessing the impact of policy-oriented research: a stocktaking. In: *Workshop on best practice methods for assessing the impact of policy-oriented research: summary and recommendations for the CGIAR, Independent Impact Assessment Report*. Washington D.C.
- Ruegg R, Feller I. 2003. A Toolkit for Evaluating Public R&D Investment: models, methods, and findings from ATP's first decade. (Grant/Contract Report). Gaithersburg: National Institute of Standards and Technology.
- Salter AJ, Martin BR. 2001. The economic benefits of publicly funded basic research: a critical review. *Research Policy* 30: 509–532. DOI: [10.1016/S0048-7333\(00\)00091-3](https://doi.org/10.1016/S0048-7333(00)00091-3).
- Samuel GN, Derrick GE. 2015. Societal impact evaluation: exploring evaluator perceptions of the characterization of impact under the

- REF2014. *Research Evaluation* 24(3): 229–241. DOI: [10.1093/reseval/rvv007](https://doi.org/10.1093/reseval/rvv007).
- Spaapen JM, Van Drooge L. 2011. Introducing “productive interactions” in social assessment. *Research Evaluation* 20(3): 211–218. DOI: [10.3152/095820211X12941371876742](https://doi.org/10.3152/095820211X12941371876742).
- Wooding S, Hanney SR, Pollitt A, Grant J, Buxton MJ, on behalf of the Project Retrosight Team. 2014. Understanding factors associated with the translation of cardiovascular research: a multinational case study approach. *Implementation Science* 9: 47. DOI: [10.1186/1748-5908-9-47](https://doi.org/10.1186/1748-5908-9-47).

**Citation de l'article** : Colinet L, Gaunand A, Joly P-B, Matt M. 2017. Des barèmes génériques pour évaluer les impacts de la recherche sur la société : l'exemple des impacts politiques. *Cah. Agric.* 26: 65006.