

# Comparaison internationale des pratiques d'évaluation *ex post*

## *Résultats préliminaires*

France Stratégie / DITP / Sciences Po  
« Printemps de l'évaluation »  
Assemblée nationale, le 28 juin 2018

# Contexte

- L'évaluation, une des missions de France Stratégie et de ses partenaires
- Quatre évaluations de politiques publiques *ex post*, confiées à des commissions au sein de France Stratégie :
  - Le comité de suivi et d'évaluation du CICE
  - Le comité de suivi des aides publiques aux entreprises (COSAPE)
  - La commission nationale d'évaluation des politiques d'innovation (CNEPI)
  - Le comité d'évaluation des ordonnances relatives au dialogue social et aux relations de travail
- Réflexions menées par France Stratégie avec le SGPI sur les méthodes d'évaluation socio-économique des projets d'investissements publics (guides)
- Organisation de conférences sur les méthodes d'évaluation et leurs enjeux : par exemple, le cycle de six séminaires sur l'évaluation d'impact des politiques publiques (en partenariat avec l'IPP, J-PAL, le LIEPP)
- 18 % des moyens de France Stratégie dévolus à sa mission d'évaluation (en pilotage et en crédits d'études)



# L'étude en cours

- Une comparaison des pratiques d'évaluation des politiques publiques dans sept pays de l'OCDE
  - Évaluations *ex post* des politiques publiques à visée scientifique
    - Intérêt : mesure de l'effet propre – causal – de la politique publique
    - Méthodes : expérimentales et quasi-expérimentales
    - Limites : délai de réalisation, disponibilité et accès aux données, précautions dans l'interprétation des résultats
  - Deux axes de questionnement :
    - Quantitatif : combien d'évaluations de ce type produites par pays ? (axe I)
    - Qualitatif : quelle est leur place dans la décision ? (axe II)
- **Comparer les pays et leurs pratiques et situer la France afin d'en tirer des enseignements en termes de bonnes pratiques (recommandations)**

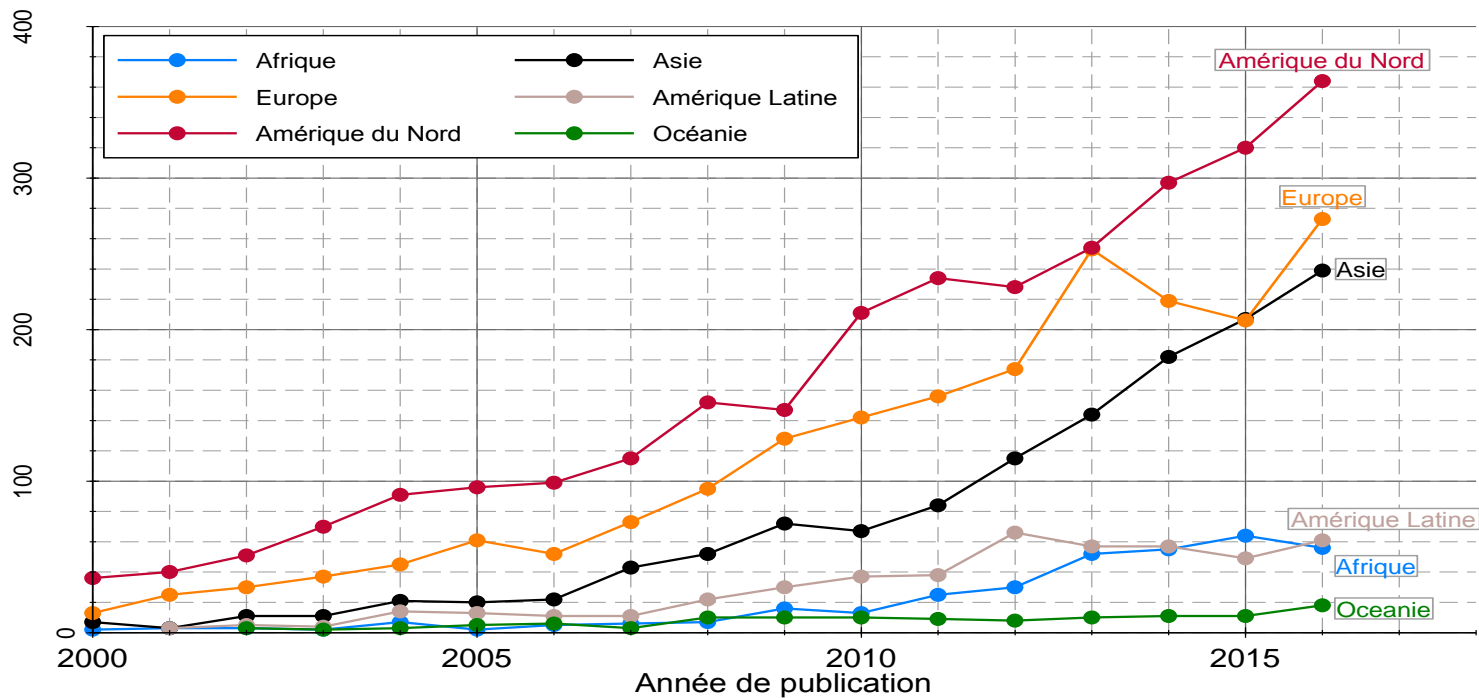


# I. Quantifier le volume d'évaluations

- Pays : Allemagne, Canada, États-Unis, France, Pays-Bas, Royaume-Uni et Suède
  - À partir de la base [Econlit](#) : articles d'économie publiés dans des revues à comité de lecture (majorité en anglais)
  - Période d'analyse : de la date d'entrée des revues dans la base jusqu'en 2016
  - Évaluations portant sur un pays donné (quels que soient les auteurs) et couvrant tous les secteurs (éducation, emploi, santé, logement, etc.)
  - Recherche par méthode (mots-clés), dans le titre et dans le résumé
- [8 669 articles](#) identifiés au total dans la base



# Nombre d'articles par région et par année

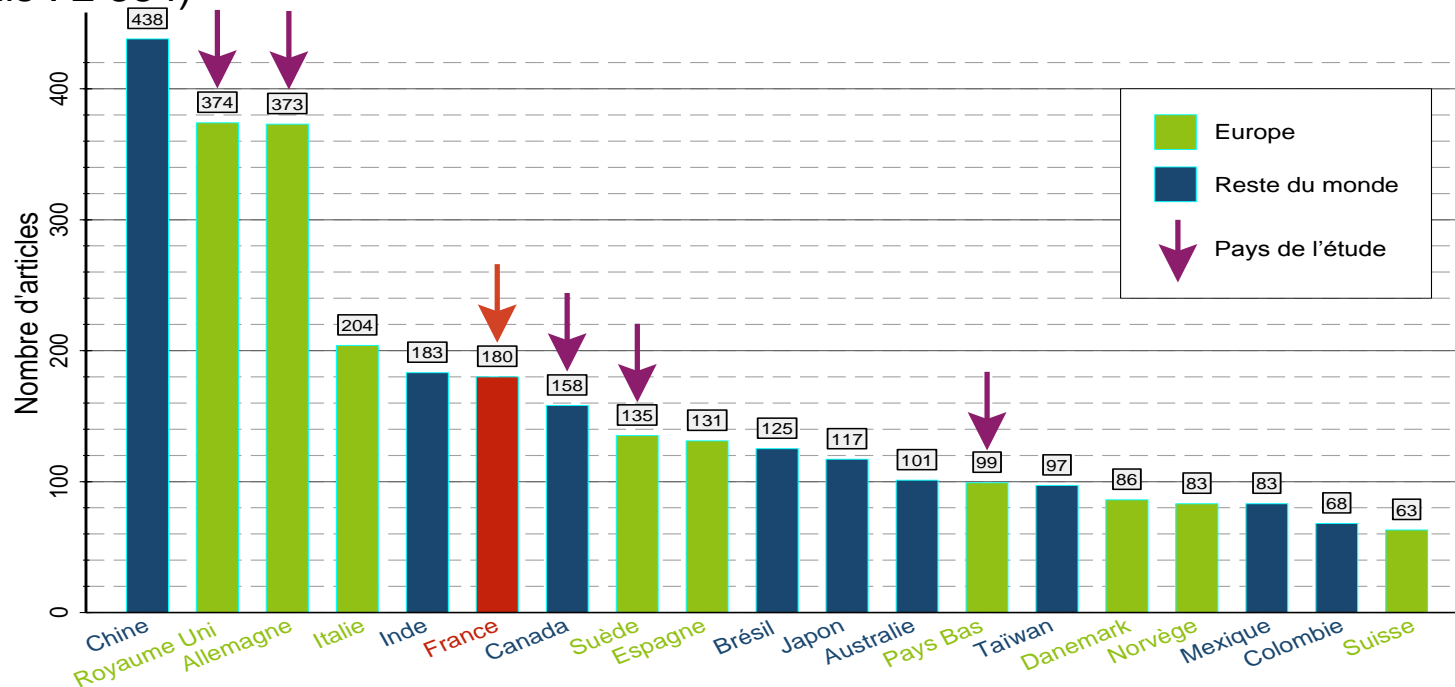


Source : base Econlit.



# Nombre d'articles par pays sur l'ensemble de la période (hors États-Unis) : le G20

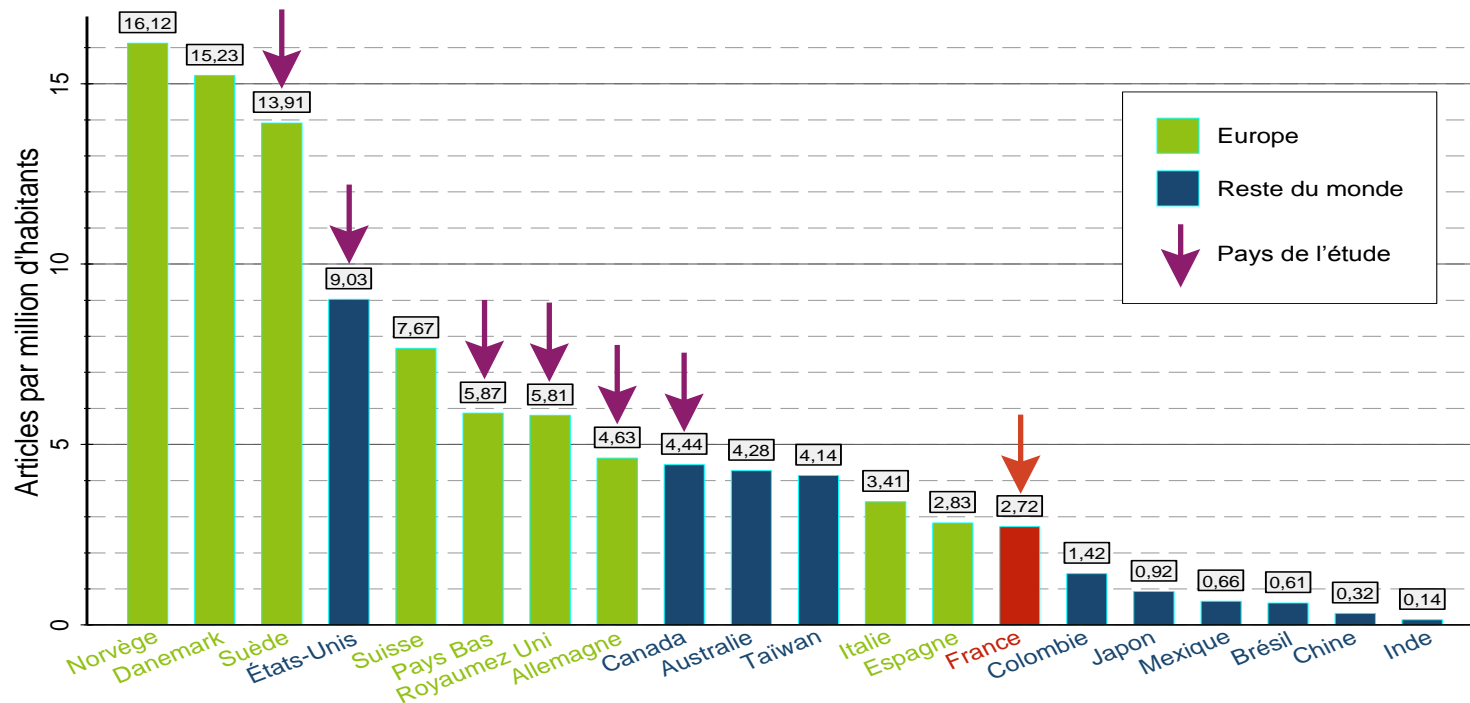
(États-Unis : 2 884)



Source : base Econlit.



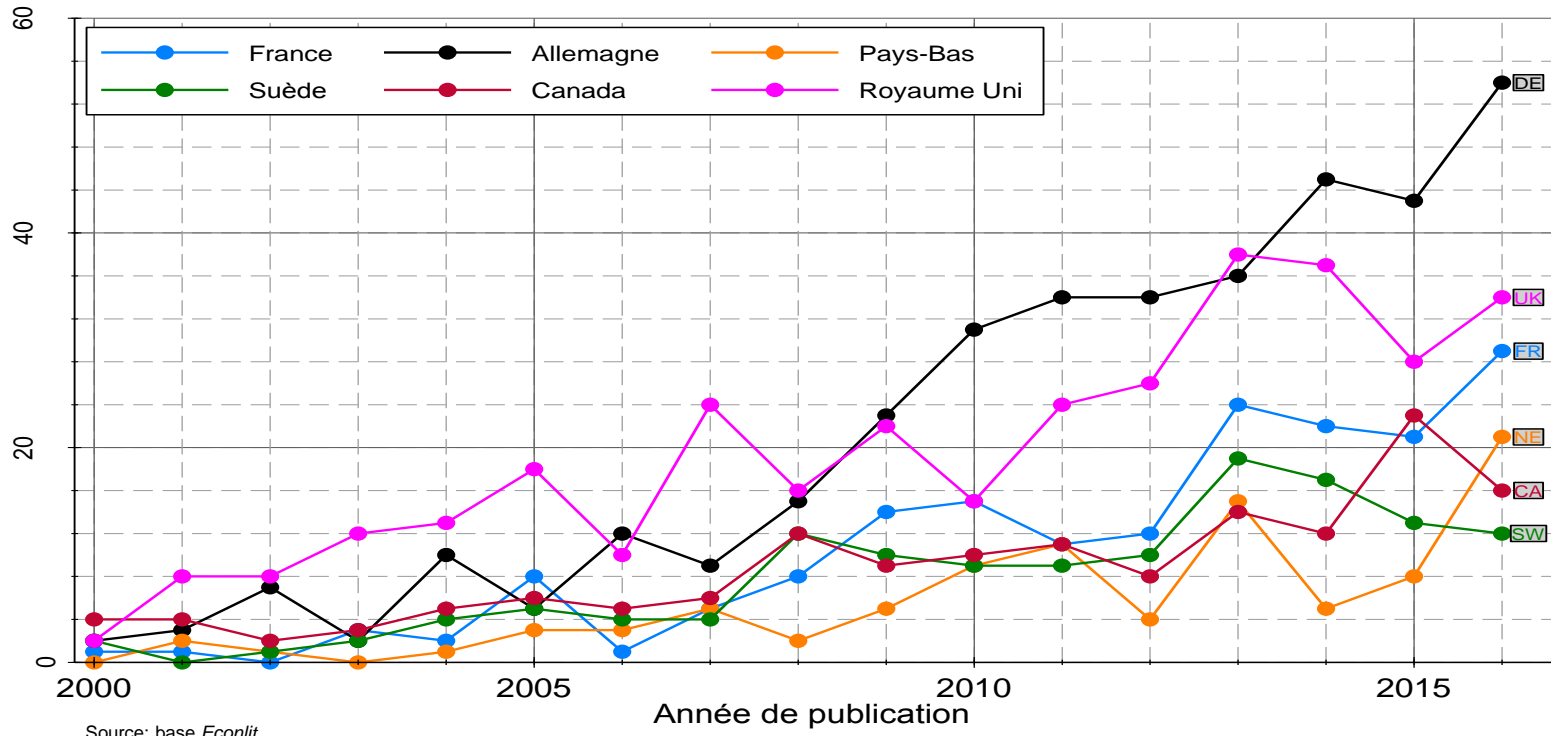
# Nombre d'articles par pays sur l'ensemble de la période (par tête)



Source : base Econlit.

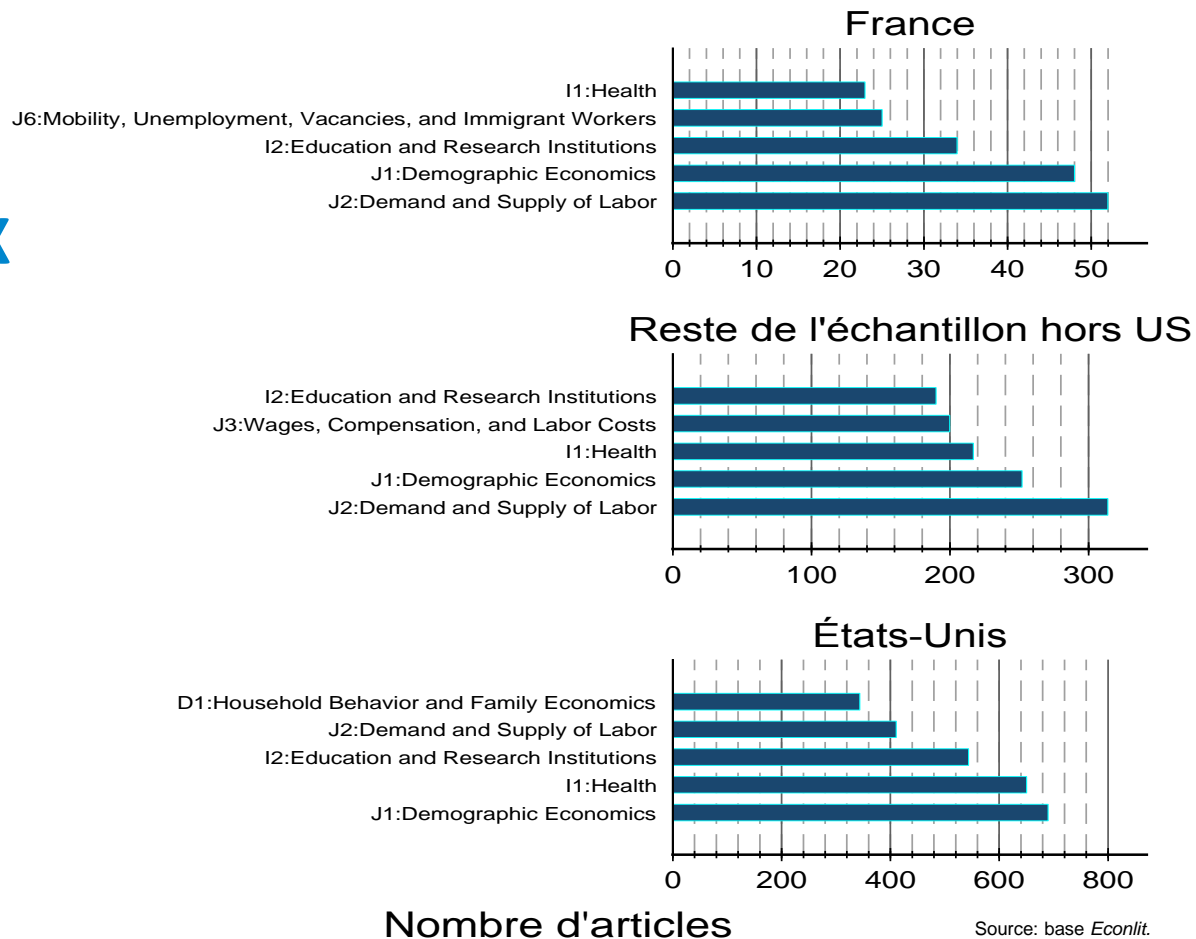


# Nombre d'articles par pays et par année (hors États-Unis)





# Cinq principaux domaines (codes JEL) France vs. autres pays de l'étude



# Conclusions provisoires de l'axe I

- Principaux résultats :
  - La France a démarré avec retard, en raison possiblement de difficultés dans l'accès aux données
  - Elle comble son retard en fin de période
  - Elle se distingue des États-Unis par une concentration des évaluations sur les questions d'emploi, mais pas des autres pays de l'échantillon
- Ce qu'il reste à faire :
  - Tests de robustesse sur la France
  - Travail d'identification des principales institutions productrices d'évaluations



## II. Place des évaluations dans la décision

- Travail qui repose sur une analyse documentaire et des entretiens avec des acteurs variés
- Travail moins avancé que l'axe I
- Bonnes pratiques identifiées selon trois critères :
  - Critère 1 : degré d'ouverture des milieux décisionnaires aux chercheurs (procédures RH, partenariats, etc.)
  - Critère 2 : qualité des dispositifs formels favorisant la conduite et l'influence des évaluations (instances, programmation, dispositions juridiques, financement)
  - Critère 3 : diffusion et influence des évaluations



# Bonnes pratiques : quelques exemples (axe II)

<b>Critères/Pays</b>	<b>États-Unis</b>	<b>Royaume-Uni</b>
<b>Critère 1</b>	<ul style="list-style-type: none"><li>• Intergovernment Personnel Act Mobility Program (IPA)</li><li>• Social and Behavioral Insights Team (SBST)</li></ul>	<ul style="list-style-type: none"><li>• Policy Profession</li><li>• Behavioural Insights Team (BIT)</li></ul>
<b>Critère 2</b>	<ul style="list-style-type: none"><li>• Chief Evaluation Officer (DOL)</li><li>• Sommes réservées pour l'évaluation (« set-asides »)</li></ul>	<ul style="list-style-type: none"><li>• Rôle du Trésor britannique (guides)</li></ul>
<b>Critère 3</b>	<ul style="list-style-type: none"><li>• Clearinghouses</li><li>• Tiered-evidence grants</li></ul>	<ul style="list-style-type: none"><li>• What Works Network</li></ul>



# Quelques pistes de réflexion pour la France

- Critère 1 (ouverture aux chercheurs)
  - Renforcer la coordination des acteurs de l'évaluation, et les partenariats entre chercheurs et décideurs publics
- Critère 2 (dispositifs formels)
  - Prévoir dans la loi instaurant un dispositif majeur le principe de son évaluation et réserver une partie de ses crédits à son financement
- Critère 3 (diffusion et influence)
  - Créer un répertoire des évaluations scientifiques *ex post* existantes pour la France, et rendre les résultats plus accessibles aux décideurs



# Annexe : mots-clés

<b>Expérience aléatoire</b>	<b>Expérience naturelle</b>	<b>Expérience quasi-aléatoire</b>
randomized trial clustered randomized trial randomized evaluation randomized experiment randomized field experiment randomized controlled trial RCT	field experiment natural experiment regional experiment quasi-randomly assigned	quasi-experiment quasi-experimental study quasi-experimental method
<b>Régression discontinue</b>	<b>Méthode d'appariement</b>	<b>Variables instrumentales</b>
regression discontinuity regression discontinuity approach discontinuity analysis	matched comparisons matched-comparison design matching propensity score	instrumental variable
<b>Double différence</b>	<b>Régression</b>	<b>Autres</b>
difference-in-differences double difference triple difference	regression approach	causal impact causal effect treatment effect counterfactual



Retrouvez l'ensemble des travaux de France Stratégie sur  
l'évaluation des politiques publiques sur :

[www.strategie.gouv.fr/thematiques/evaluation-de-politiques-publiques](http://www.strategie.gouv.fr/thematiques/evaluation-de-politiques-publiques)