

ANALYSE ET ÉVALUATION DES POLITIQUES PUBLIQUES

PSL Week Mineure Action Publique

Novembre 2024

OBJECTIFS DU COURS

- 1- Comprendre le rôle des sciences économiques pour l'analyse de l'action publique
- 2- Pouvoir lire et (globalement) comprendre une analyse de politiques publiques
- 3- Pouvoir (commencer à) identifier les forces et faiblesses d'une analyse de PP

1 COMPRENDRE LE RÔLE DES SCIENCES ÉCONOMIQUES POUR L'ANALYSE DE L'ACTION PUBLIQUE

PLAN DE LA PARTIE 1

1 Les sciences économiques, c'est quoi ?

2 Le cadre néoclassique et l'absence de déficiences de marchés

3 Justifier l'intervention publique

QU'EST-CE QUE L'ÉCONOMIE

Sciences économiques: Analyse de l'allocation optimale des ressources

- Comment les ressources (comme le travail, ou les matières premières) sont produites, échangées et réparties
- Expliquer et améliorer la gestion des ressources pour répondre aux besoins humains

Les sciences économiques se divisent en deux grandes branches :

- **Microéconomie** : Étudie les comportements individuels (ménages, entreprises) et les mécanismes de marché.
- **Macroéconomie** : Analyse les phénomènes économiques globaux (croissance, chômage, inflation).

QU'EST-CE QUE L'ÉCONOMIE

L'économie fait appel à des outils:

1/ Théoriques

Soit pour rationaliser/poser un cadre pour comprendre les comportements, les équilibres

Soit pour proposer des hypothèses à tester

2/Empiriques: avec de l'utilisation de données

Utilisation de méthodes statistiques et de l'économétrie

QU'EST-CE QUE L'ÉCONOMIE

L'économie a deux approches :

1/ Une approche dite « positive » (= descriptive) → observations, corrélations, causalité...

2/ Une approche dite « normative » → Suite à l'observation, la description et l'identification de causalité ou à l'aide de modèles théoriques : comment maximiser l'utilité, atteindre un objectif, guider les agents vers un comportement souhaitable...

STATISTIQUES ET ÉCONOMÉTRIE

L'économétrie est la rencontre de la statistique et de l'économie.

L'économétrie est l'application de la statistique à la science économique.

L'économétrie cherche essentiellement à:

- Estimer des paramètres
- Identifier des relations causales
- Inférence : connaître un paramètre de la population globale à partir de l'étude d'un échantillon

Différent de la « statistique descriptive » qui a aussi un intérêt pour la compréhension des politiques publiques

LE CADRE NEOCLASSIQUE

Economie concurrentielle hypothétique

Efficacité du marché dans cette économie : pas d'intervention nécessaire (sauf redistribution)

Pour analyser les interventions, partir des déviations par rapport aux hypothèses de concurrence

PREMIER THÉORÈME DU BIEN-ÊTRE

Si:

- 1) Ménages et entreprises prennent les prix comme donnés
- 2) Tous les marchés existent
- 3) L'information est parfaite
- 4) La mobilité est parfaite (les coûts de transaction sont nuls)

alors l'équilibre concurrentiel, s'il existe, est Pareto-optimal

Equilibre: chaque agent (ménage, entreprise, travailleurs...) est à son optimum (= maximise sa fonction objectif : utilité, profit, arbitrage travail/loisir...) étant donné le système de prix

Pareto-optimal: on ne peut pas améliorer la situation d'un agent sans détériorer celle d'un autre

L'équilibre concurrentiel "Walrasien" est la formalisation mathématique de la "main invisible" d'Adam Smith
On peut y voir les fondements mathématiques du libéralisme
Influence très forte sur la doctrine libérale au sens large

Autres hypothèses:

- Chaque agent est "atomique": suffisamment petit pour ne pas avoir d'influence à lui seul sur les prix (\neq monopole ou milliardaire)
- Les agents sont "rationnels"
- Pas d'interdépendance entre les agents

DEUXIÈME THÉORÈME DU BIEN-ÊTRE

Si:

- 1) Ménages et entreprises prennent les prix comme donnés
- 2) Tous les marchés existent
- 3) L'information est parfaite
- 4) La mobilité est parfaite
- 5) L'économie est convexe
- 6) On peut réaliser des transferts forfaitaires

Alors n'importe quelle allocation Pareto-optimale peut être obtenue comme un équilibre concurrentiel

=> Une intervention publique ne permettrait pas de faire mieux, et n'importe quelle répartition peut être obtenue avec un changement de dotations initiales

TRANSFERTS FORFAITAIRES : UNE IDÉE ABSTRAITE

Doivent être indépendants des comportements d'optimisation

Par exemple, ne peuvent pas être déterminés par les revenus observés

Dépend de « paramètres » théoriques (e.g. potentiel productif inné) inobservés

Transferts non-forfaitaires = distorsions potentielles

Le second théorème évite le problème d'arbitrage équité-efficacité. S'il ne s'applique pas, on retrouve le trade off entre distribution et Pareto efficacité.

QUAND EST-CE QUE LE CADRE NÉOCLASSIQUE NE S'APPLIQUE PAS?

1. Existence des marchés
2. Information imparfaite
3. Externalités
4. Biens publics
5. Concurrence imparfaite
6. Rationalité limitée

EXISTENCE DES MARCHÉS

Pour discuter les imperfections des marchés, il faut encore que les marchés existent

Suppose un cadre institutionnel permettant de faire respecter les contrats, par exemple droits de propriété

Peut aussi résulter d'un équilibre social sans intervention directe de l'Etat (menaces, réputation)

INFORMATION IMPARFAITE

Première dimension: on observe imparfaitement la qualité des biens ou les « types » des agents
= Sélection adverse

- Exemple classique : Le marché des voitures d'occasion (George Akerlof, *The Market for Lemons* (1970))
 - Les vendeurs connaissent la véritable qualité des voitures (bonnes voitures ou "lemons", c'est-à-dire des voitures défectueuses).
 - Les acheteurs ne peuvent pas observer directement cette qualité et doivent s'appuyer sur le prix pour évaluer la voiture.
 - Les acheteurs, craignant de payer trop cher pour une voiture défectueuse, sont prêts à payer un prix moyen.
 - Les vendeurs de bonnes voitures, estimant que le prix moyen est trop bas pour leur produit, se retirent du marché.
 - Le marché se retrouve dominé par des voitures de mauvaise qualité.
- Risque de réduction, voire d'effondrement du marché
- Secteurs les plus concernés: Assurance, banque
- Solutions: Certification qualité, normes de sécurité, assurances obligatoires, etc.

INFORMATION IMPARFAITE

Deuxième dimension: on observe imparfaitement les actions des agents
= aléa moral

Il se produit après la réalisation d'un contrat ou d'une transaction

- Risque de réduction, voire d'effondrement du marché
- Exemples classiques: assurance (prise de risque des assurés), banque (e.g. « too big to fail »), marché du travail (effort au travail), etc.
- Solutions: transparence, contrôle, régulation (e.g. bancaire), contrats incitatifs (franchise + bonus/malus pour les assurances automobiles)

EXTERNALITÉS

Les actions d'un agent modifient le bien-être d'un autre agent, sans donner lieu à une compensation

Exemples ?

Conduit les agents à produire ou consommer des quantités de biens qui sont leur optimum privé étant donnés les prix, mais ne sont pas l'optimum social (car les prix de marché ne reflètent pas le coût ou le bénéfice social)

THÉORÈME DE COASE

Idée : dire qu'il y a une externalité revient à dire qu'il manque des marchés : on pourrait imaginer un marché où j'achète mon droit à polluer les autres

Problème : il faut des marchés parfaits et des coûts de transaction nuls

Il n'y a pas de marché décentralisé pour définir le contrat : valoriser l'externalité, définir le contrat, distribuer les compensations

De telles négociations seraient compliquées, notamment si beaucoup d'agents (+ free riding : si plusieurs victimes, chacune compte sur les autres pour payer)

Interventions de politiques publiques:

- Réglementation des quantités
- Taxes/subventions

BIENS PUBLICS

Biens publics purs:

- **Non rival:** la consommation par un individu n'empêche pas la consommation par un autre
- **Non-exclusif:** quand le bien est produit, on ne peut pas empêcher une personne de consommer ce bien

Exemples ?

BIENS PUBLICS: INEFFICACITÉS

Non-rival

On ne produit pas assez de bien public, parce que le marché ne tient pas compte de l'externalité (ma consommation profite aussi aux autres)

Non-exclusif

Crée un problème de free-riding : les entreprises privées peuvent avoir du mal à financer la production de leurs biens

BIENS PUBLICS: INTERVENTIONS

Produire le bien public a des niveaux socialement optimaux

Taxer pour financer la production du bien public (le subventionner)

Mais difficile de connaître la valeur que les consommateurs donnent au bien public (leur consentement à payer)

CONCURRENCE IMPARFAITE

Monopole permet d'imposer un prix supérieur à l'optimum concurrentiel

- Monopole naturel : coûts fixes importants, coûts unitaires décroissants, une seule firme domine
 - Production publique
 - Production privée réglementée
 - Production réglementée de la part qui génère le coût fixe (infrastructure), concurrence pour le reste
- Monopoles générés par entente ou abus de position dominante (barrière à l'entrée)
 - Droit de la concurrence, autorités de contrôle

RATIONALITÉ LIMITÉE

Dans quelle mesure peut-on considérer qu'un agent économique prend des décisions contraires à son intérêt?

Révolution plus récente (mais initiée par Kahneman et Tversky) : revenir sur le modèle de rationalité des agents (behavioral economics)

- Information : décisions rationnelles supposent une très grande quantité d'information, coûteuse à réunir
- Biais cognitifs : la complexité des décisions (comparer tous les choix possibles) conduit à prendre des raccourcis
- Incohérence temporelle : n'agit pas demain comme il le prévoit aujourd'hui (exemple de l'addiction, taxer les cigarettes, etc.)

CONCLUSION DE CETTE PREMIÈRE PARTIE TRÈS THÉORIQUE

- Des marchés imparfaits...
- Qui justifient l'action publique et le rôle de l'Etat d'un point de vue théorique (stabilité des marchés, contrats, déficience des marchés, redistribution)
- Mais aussi très éloigné des considérations pratiques sur les politiques publiques
 - Secteur public très important :
 - dépense publique = 58,2% du PIB en 2022 (en incluant les prestations sociales)
 - Secteur public = 21% de l'emploi total (5,7 millions de salariés)
 - Des pans entiers de l'économie sont publics : éducation, police, etc.
 - Sans évaluation, pas de façon probante de savoir si la politique a atteint ses objectifs, si elle est efficace, etc.

OBJECTIFS DU RESTE DU COURS

1 Principe:

- Il faut évaluer les politiques publiques !

2 concepts:

- Causalité vs corrélation
- La notion du contrefactuel

6 méthodes:

- RCTs
- Différence en différence (DiD)
- Régression de discontinuités (RDD)
- Méthodes d'appariement / contrôles synthétiques
- Régressions linéaires et analyse descriptive
- Micro-simulation

3 distinctions à retenir et comprendre:

- Données administratives et données d'enquête
- Variables observables et non observables
- Analyse d'impact versus analyse descriptive

PLAN

1. Cadre général pour l'évaluation des politiques publiques
2. Evaluations d'impact
3. Méthodes

Analyse et évaluation des politiques publiques : cadre général

POLITIQUE PUBLIQUE: DÉFINITION

- Définition d'une politique publique (PP) :
« Une politique publique est une intervention de l'État ou d'une autre autorité publique (banque centrale, collectivité territoriale, ...) afin d'avoir un effet sur la société ou un territoire » (Source : Wikipédia)
- Une politique publique peut être :
 - Au niveau européen, national, des collectivités territoriales
 - Temporaire ou permanente
 - Ciblée ou non ciblée

A QUOI SERT L'ÉVALUATION ? POURQUOI ÉVALUER ?

A quoi sert l'évaluation ?

- Accompagner la transformation de l'intervention publique
- Rendre les actions des pouvoirs publics plus efficaces (lien entre objectifs et résultats)
- Rationaliser les dépenses publiques
- Responsabiliser les décideurs publics
- Favoriser la participation citoyenne au contrôle et à la fabrique des politiques publiques

DÉFINITIONS DE L'EPP

- L'évaluation d'une politique publique (EPP) consiste à émettre un jugement de valeur sur une intervention publique
- Définitions de l'EPP tournent souvent autour du triptyque « objectifs - moyens – résultats ». Plusieurs définitions de l'EPP, **plus ou moins restrictives** :
 - Décret n° 90-82 du 22 janvier 1990 relatif à l'évaluation des politiques publiques : l'EPP « 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 »
 - Décret du 18 novembre 1998 : « L'évaluation d'une politique publique consiste à comparer ses résultats aux moyens qu'elle met en œuvre, qu'ils soient juridiques, administratifs ou financiers, et aux objectifs initialement fixés. Elle se distingue du contrôle et du travail d'inspection en ce qu'elle doit aboutir à un jugement partagé sur l'efficacité de cette politique et non à la simple vérification du respect de normes administratives ou techniques »
 - Distinction entre la notion d'EPP d'une part, et les notions de contrôle et d'inspection d'autre part. L'EPP **n'est pas** le contrôle et le travail d'inspection

DÉFINITIONS DE L'EPP

Définition plus restrictive donnée par France stratégie :

« toute publication visant à éclairer un champ de politique publique ou la conduite d'une intervention publique et s'appuyant sur une expertise reconnue et sur des données ayant valeurs de preuves (statistiques, témoignages, etc.) »

LES CRITÈRES D'ÉVALUATION DANS L'EPP

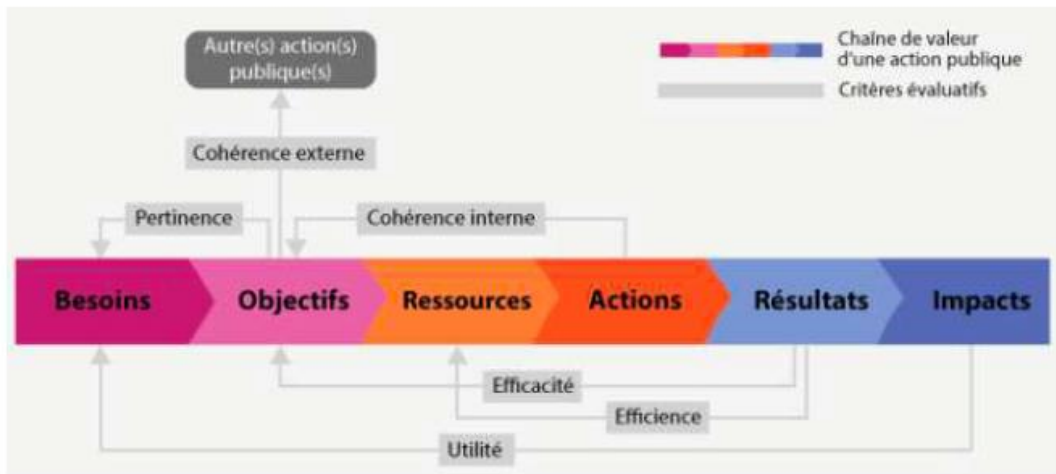
Principaux critères : la pertinence, la cohérence, l'efficacité et l'efficience.

L'évaluation d'une PP permet d'apprécier si cette PP est :

- **pertinente** : on évalue l'adéquation, l'adaptation, entre les objectifs explicites de la PP d'une part, et la réalité sociale (les besoins ou les problèmes que la PP est supposée résoudre) d'autre part
- **cohérente** dans sa conception et dans son déploiement
 - La cohérence interne a trait à l'adéquation entre les objectifs assignés à une politique et les moyens qui lui sont alloués
 - La cohérence externe désigne l'adéquation entre l'action évaluée et d'autres politiques
- **efficace** : on mesure les résultats obtenus au regard des objectifs et on met en évidence les rapports de causalité entre les objectifs, implicites et explicites, et les effets réels des actions publiques
 - Se pose la question des effets pervers et des externalités de la politique
- **efficience** : on s'intéresse au rapport entre les moyens mis en œuvre (les coûts) et les résultats de l'action publique

Critères : schéma récapitulatif

Schéma 1
Les critères d'évaluation d'une politique publique



Source : SG-MAP (2015)

Source : France stratégie : https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/guide_methodologique_20160906web.pdf

Note : Le SGMAP est le Secrétariat général pour la modernisation de l'action publique.

UN PEU D'HISTOIRE : DANS LE MONDE

Source sur l'histoire de l'EPP : Vie publique (2022).

L'EPP apparaît vers le début du XXe siècle aux **États-Unis**

- A partir de 1970, l'acteur central de l'EPP est le *General Accounting Office*
- 1960-1980 : premier âge d'or de l'évaluation, porté par la méfiance à l'égard de toute action étatique

Après la Seconde guerre mondiale, l'évaluation se répand aux pays anglo-saxons et à l'Europe du Nord

Cela explique l'avance des É-U et de ces pays dans l'EPP aujourd'hui

UN PEU D'HISTOIRE : EN FRANCE

En France, **le développement de l'EPP est tardif**

Il faut attendre les années 1970 et 1980 pour que soit mise en place la « rationalisation des choix budgétaires » (RCB) qui permet et préconise l'évaluation principalement ex-ante

- L'EPP devient un outil de modernisation de l'Etat

Dans les années 1980, l'évaluation est surtout confiée au Commissariat général du Plan

Fin des années 1980 : impulsion donnée par le gouvernement de M. Rocard

- Sont créées plusieurs institutions au sein des ministères en charge de faire de l'évaluation

Mais finalement, un faible nombre d'évaluations ont lieu et elles débouchent rarement sur des décisions importantes. Elles sont davantage pratiquées par les régions, soutenues dans cette démarche par la Commission Européenne

HISTOIRE

Depuis la révision constitutionnelle de 2008 : il est inscrit dans l'article 24 de la Constitution que l'EPP est une mission du Parlement (Assemblée nationale et Sénat) (comme le vote de la loi et le contrôle du gouvernement)

Décembre 2012 : le gouvernement lance une démarche d'évaluation ex-post de l'ensemble des politiques publiques, dans le cadre de la modernisation de l'action publique

- Déclaration de J-M Ayrault (2012) : « toutes les politiques publiques, sur l'ensemble du quinquennat, feront l'objet d'une évaluation »
- La tendance à la maîtrise des dépenses + parfois une certaine défiance à l'égard des institutions politiques → Renforcent la légitimité de l'EPP

Décret du 22 avril 2013 : création du Commissariat général à la stratégie et à la prospective (qui est connu sous le nom de France stratégie). L'EPP est au cœur de ses missions

ACTEURS DE L'ÉVALUATION EN FRANCE

Acteurs de l'évaluation en France :

- **Parlement** (article 24 Constitution) (comité d'évaluation et de contrôle des politiques publiques (CEC))
- **Cour des comptes** (article 47-2 Constitution)
- **France stratégie**
- Autres acteurs publics: Conseil économique, social et environnemental, Conseil d'Etat, Conseil d'Analyse Economique ...
- **Les grands corps** (IGF, IGAS, etc.)
- Institut national de la statistique et des études économiques (Insee) et les services statistiques ministériels (SSM)
- Autorités publiques ou administratives indépendantes:
 - Haute autorité de santé (HAS): a été créé par une loi de 2014 ; l'HAS a plusieurs missions, dont une mission d'évaluation des produits, actes , prestations et technologies de santé
 - Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES): a été créé par une loi de 2013 ; est chargé de l'évaluation de l'enseignement supérieur et de la recherche publique
- Universités et laboratoires spécifiques (IPP, LIEPP, TEPP, CREST, INSA ...)
- Chercheurs (économistes, chercheurs en sciences sociales, etc.)

ACTEURS DE L'ÉVALUATION EN FRANCE

Les institutions publiques qui réalisent des évaluations ont souvent recours à des « **chargés d'évaluation** » (comme des économistes, des chercheurs, des consultants, etc.) qui réalisent l'évaluation avec les fonctionnaires ou dans le cadre d' « Appels à projet de recherche »

Spécificité française : les EPP sont souvent réalisées par **administrations**

- alors que dans d'autres pays, la place des universités est plus grande

TECHNIQUES UTILISÉES LORS D'UNE EPP

Une EPP s'appuie sur

- Une expertise
- Des données quantitatives (statistiques) et qualitatives (témoignages, etc.)
 - Complémentarité entre l'approche quantitative et l'approche qualitative
 - NB : Dans ce cours, je vous présente les techniques ayant recours à des données quanti

Les évaluations peuvent être présentées dans des

- Articles académiques publiés et documents de travail
- Articles de presse
- Rapports d'information parlementaires
- Enquêtes réalisées par des instituts de sondage
- Etc.

PRINCIPES POUR UNE BONNE ÉVALUATION

- **Expertise**
 - Accès aux données
 - Protocole d'évaluation en amont
 - Compétence et réputations reconnus des évaluateurs
- **Indépendance**
 - Publication des résultats
 - Approche non partisane (respecter les choix démocratiques)
- **Communication**
 - Publication au grand public
 - Participer au débat public

EVALUATIONS D'IMPACT

SOURCES SUR L'ÉVALUATION D'IMPACT

- Givord P (2014). Méthodes économétriques pour l'évaluation de politiques publiques. Économie & prévision, 204-205, 1-28.
 - Lien : <https://www.cairn.info/revue-economie-et-prevision-2014-1-page-1.htm?ref=doi>
 - Voir en particulier l'introduction et la partie intitulée « La formalisation du problème d'évaluation : Le cadre de Rubin »
- Behaghel L (2006) Lire l'économétrie. Collection Repères. La Découverte.
 - Voir en particulier le tableau IV, page 56, intitulé « Typologie des principales stratégies empiriques »

ÉVALUATION D'IMPACT

- Les évaluations d'impact se distinguent des évaluations générales qui répondent à plusieurs types de questions (évaluation de la pertinence, de la cohérence, de l'utilité, etc.)
- Les évaluations d'impact sont un type particulier d'évaluations qui porte sur la **causalité**. La dimension **causale** est primordiale. Contrairement aux évaluations générales, les évaluations d'impact sont structurées autour d'un type particulier de question : **quel est l'impact d'un programme sur une variable de résultat ?**
- **Définition :** Variable de résultat = variable d'étude = *outcome* = le phénomène économique que l'on cherche à analyser
 - Exemples : les résultats scolaires (la performance scolaire) des élèves ; le fait de trouver un emploi ou pas, pour des chômeurs
- **Vocabulaire :** on utilise les mots suivants comme des synonymes
 - Programme, traitement, politique
 - Effet causal, impact
 - Variable de résultat, variable d'étude, *outcome*

QUESTIONS QUE L'ON SE POSE QUAND ON RÉALISE UNE ÉVALUATION D'IMPACT

(1) La PP a-t-elle un effet causal (sur une ou des variable(s) de résultat) ? Oui / non

(1-bis) Question liée : quel est la taille de l'effet ?

- Les évaluations d'impact visent **généralement** à estimer l'impact **moyen** du programme (sur une ou des variable(s) de résultat) : quel est l'impact moyen de la PP ?
- Mais on peut enrichir l'analyse en regardant l'ensemble de la distribution conditionnelle de la variable d'intérêt et non seulement la moyenne de la variable
- On s'intéresse à la taille de l'impact pour différents **quantiles** de la variable de résultat (*quantile treatment effect*)
 - On se penche donc sur une forme d'hétérogénéité
- Intuition : une PP peut avoir un effet **moyen** limité, mais permettre d'augmenter suffisamment **les niveaux les plus faibles** de la variable d'intérêt, pour que l'implémentation de la PP soit souhaitable
 - Exemple : Il peut ainsi être souhaitable de mettre en œuvre une PP éducative qui permet de réduire la proportion d'élèves **en grande difficulté** même si cette PP n'a qu'un effet négligeable sur le niveau moyen de l'ensemble des élèves

QUANTILES

Les quantiles les plus couramment utilisés en économétrie sont la **médiane**, le premier et le dernier **quartiles** et le premier et le dernier **déciles**

Définition du premier décile (de revenu, par exemple) : le premier décile désigne *stricto sensu* le seuil de revenus en-dessous duquel 10 % exactement de la population se situe

- Le « premier décile » de revenu est donc un niveau de **revenu**
- Erreur : souvent, on parle du « premier décile » pour désigner les 10 % de la population les moins riches. Cette désignation est incorrecte. Cette erreur est fréquente dans les articles de presse

Définition du premier quartile d'une variable de résultat : seuil de la variable de résultat tel que $\frac{1}{4}$ de la population soit en-dessous du seuil

Définition de la médiane : seuil de la variable de résultat tel que la moitié de la population soit en-dessous de ce seuil (et l'autre moitié de la population soit au-dessus)

Exemple : calculer l'effet de la PP pour le 1er quartile de la variable d'intérêt (i.e. notes à l'école)

- Exemple : l'analyse va nous indiquer l'effet de la PP pour le 1^{er} quartile de la distribution des notes, le 2^{ème}, le 3^{ème} et le 4^{ème}
- On verra si l'effet de la PP varie entre ces quartiles

QUESTIONS QUE L'ON SE POSE... (SUITE)

(2) La PP a-t-elle un effet hétérogène dans la population ?

- Ici on s'intéresse à de l'hétérogénéité selon d'autres critères que le niveau de la variable de résultat)
- Exemple : on teste si l'effet de la PP (sur une variable de résultat) est différent...
 - Entre femmes et hommes
 - Selon l'âge (entre groupes d'âge)
 - Selon le niveau d'éducation
 - Selon le niveau de revenu

EXEMPLES D'ÉVALUATION D'IMPACT

Développement (pays à faible revenu)

- Cadre : débat, ou questionnaire, sur l'efficacité de l'aide publique au développement
- Quel est l'impact des politiques de « développement » et en particulier des politiques de lutte contre la pauvreté ?

Développement, santé

- Un programme d'accès à l'eau potable (dans un pays à bas revenu) permet-il d'améliorer l'état de santé des bénéficiaires ?
- Les campagnes de lutte contre le paludisme (distribution de moustiquaires, démoustication, etc.) améliore-t-elles l'état de santé des enfants ?
 - ... et les résultats scolaires et le niveau de revenu à l'âge adulte ?

EXEMPLES D'ÉVALUATION D'IMPACT

Éducation

- Est-ce qu'une politique de diminution de la taille des classes engendre une augmentation du niveau scolaire ?
- Quel est l'effet de la suppression de la carte scolaire ?

Économie urbaine

- Les systèmes de transports publics (de type métro) améliorent-ils l'attractivité des villes ?

EXEMPLES D'ÉVALUATION D'IMPACT

Travail, emploi

- Quel est l'impact d'un programme de suivi des chômeurs ?
- Quel est l'impact d'un programme de formation des chômeurs ? Est-ce qu'il facilite le retour à l'emploi des chômeurs ?
- Une politique d'aide aux retours à l'emploi a-t-elle effectivement entraîné des sorties du chômage plus fréquentes ?
- Un programme de formation des jeunes (peu qualifiés) permet-il de développer l'entrepreneuriat et d'augmenter les revenus ?

ÉVALUATION DES PP → QUAND? 2 (3) MOMENTS:

1- Ex-ante (avant la mise en place de la PP)

Analyser les impacts potentiels de la mesure.

Plusieurs outils:

- Etude de la littérature académique (politiques publiques similaires ailleurs dans le temps ou dans l'espace)
- Etude théorique → se basant sur un ou plusieurs modèles théoriques
- (Micro)Simulations → utilisation d'outils de simulation qui utilisent de vraies données

2- Ex-post (après la mise en place de la PP)

Analyser les impacts réels après son entrée en vigueur

Outils utilisés:

- Econométrie sur données
- Parfois éclairage « structurel » qui allie modèle théorique et données

3- In-itinere (sur la longueur)

- Parfois une politique a des effets de court, moyen et long terme. C'est pour cela que certaines évaluations commencent vite (parfois même ex-ante) et durent longtemps
- Parfois une politique est déployée sur une certaine partie du territoire ou sur certains bénéficiaires et où la montée en gamme de la politique peut-être conditionné au succès sur l'échantillon test.

ÉVALUATION EX ANTE

Évaluation **ex ante** : le but ici est d'**anticiper** les résultats

Comment sont faites les évaluations ex ante ?

Étude de la littérature académique

- PP ayant eu lieu dans d'autres pays, si possible assez « comparables » et récemment

Traditionnellement, on utilisait une approche **structurelle** : on explicite un modèle **théorique**, qui spécifie « les règles de comportement individuel, et, si possible, les interactions entre ces comportements » (source : Blanchet et al., 2016)

- Pour calibrer le modèle, on a besoin de connaître la valeur de certains paramètres. On va alors utiliser des estimations empiriques de ces paramètres
- Le modèle va permettre de prévoir le résultat d'une politique
- Approche structurelle est utile quand on veut évaluer ex ante une politique totalement inédite, qui n'a jamais été appliquée dans aucun pays, pour laquelle on ne possède pas de référence préexistante
- Force de l'approche structurelle : elle prend en compte les effets de bouclage ; elle donne aux décideurs des vues globales de l'ensemble des effets des PP
- Faiblesse de l'approche structurelle : il arrive que les modèles de comportements soient critiquables
- Micro-simulation sans réponses comportementales est utile aussi!

ÉVALUATION EX ANTE

Aujourd'hui, les modèles structurels n'ont pas disparu, mais les évaluations ex ante se font **souvent grâce à de la simulation**

- Définition actuelle de l'évaluation ex ante : évaluation de l'impact d'une politique qui n'est pas encore en œuvre, au moyen le plus souvent de simulations
- La simulation utilise de vraies données et de vrais barèmes et elle se base sur des hypothèses plausibles

ÉVALUATION IN ITINERE

Évaluation **in itinere** : quand l'évaluation est concomitante ou parallèle à l'action

But de ce type d'évaluations : apporter un diagnostic d'étape, faire du *monitoring*

ÉVALUATION EX POST

Évaluation d'impact **ex post** = évaluation après la mise en œuvre de la PP

- Plus ou moins longtemps après (ex : 1 an, 5 ans, 10-20 ans)

Pour évaluer, on utilise des données (microéconomiques) quantitatives collectées après la mise en œuvre de la mesure

Ce type d'évaluation se fait à l'aide d'outils économétriques

Exemples d'évaluations ex post

- L'approche expérimentale
- Les méthodes d'appariement (*matching*)
- Les méthodes pseudo-expérimentales (différence-de-différences, variables instrumentales, RDD)

ÉVALUATION D'IMPACT : DEUX ERREURS COURANTES

Dans une évaluation d'impact, on cherche à mesurer l'effet causal d'un programme

- On s'intéresse à la relation de cause à effet

2 erreurs sont souvent commises

Une première erreur = un « **piège** » de l'évaluation = erreur qui consiste à évaluer l'impact en comparant la variable de résultat des bénéficiaires du programme et la variable de résultat des non-bénéficiaires

- Erreur qui consiste à comparer les résultats scolaires des élèves qui ont bénéficié d'un programme et les résultats scolaires des élèves qui n'en ont pas bénéficié
- Erreur qui consiste à comparer le niveau de retour à l'emploi des chômeurs qui ont bénéficié d'un traitement et le niveau de retour à l'emploi des chômeurs qui n'en ont pas bénéficié

Cette comparaison ne permet pas de mesurer l'effet causal, en général

LE PIÈGE DE L'EFFET DE SÉLECTION

Pourquoi ? A cause d'un possible **effet de sélection** (*selection effect*)

- Cette expression appartient au jargon statistique

Qu'est-ce qu'un effet de sélection ? Il est dû à de la sélection ou à de l'auto-sélection

- **Sélection** : Les programmes **ciblent** souvent certaines populations : en général, les bénéficiaires d'un programme sont les individus pour lesquels le programme est particulièrement utile
 - Les bénéficiaires peuvent avoir de profils socio-économiques particuliers (indépendamment du programme évalué). Les programmes **ciblent** souvent certaines populations selon un ou plusieurs critères bien définis (pauvreté, niveau initial, etc.)
 - Exemple : une politique de réduction de la taille des classes est mise en œuvre dans des classes où les élèves sont vraiment en difficulté
 - **Auto-sélection** : Les gens choisissent ou non de participer au programme. Ceux qui choisissent de participer ont peut-être plus d'information ou sont peut-être plus motivés que les non-bénéficiaires
- Des différences subtiles entre bénéficiaires et non-bénéficiaires. Ces différences entre les caractéristiques des personnes peuvent être **observables** ou **inobservables** dans les données (par exemple, la motivation ou l'accès à l'information sont généralement inobservables)

LE PIÈGE DE L'EFFET DE SÉLECTION

Un des pièges de l'évaluation d'impact : trouver une corrélation (significative) entre la variable de programme et la variable de résultat, et en conclure que le programme a un effet causal sur la variable de résultat

Le problème est que la corrélation peut être (complètement ou en partie) due à un effet de sélection

Faire cette erreur revient à attribuer au programme ce qui relève des spécificités des bénéficiaires, c'est-à-dire ce qui relève d'un effet de sélection

La présence d'effets de sélection rend l'évaluation difficile...

Une bonne évaluation nécessite de comprendre les mécanismes de sélection (s'il y en a) pour pouvoir **neutraliser les effets de sélection**

VUE PLUS GLOBALE DU « PIÈGE » : LA CONFUSION ENTRE CORRÉLATION ET CAUSALITÉ

Dans une évaluation d'impact, on cherche à mesurer **l'effet causal** d'un programme

Nous venons de présenter un « piège » de l'évaluation d'impact dû au biais de sélection

Ce « piège » s'inscrit dans un « piège » plus général : la confusion entre corrélation et causalité. Corrélacion ne veut pas dire causalité !

EFFETS DE SÉLECTION

Le problème de la confusion possible entre effet causal du programme et effet de sélection se pose dans d'autres contextes que celui de l'évaluation d'impact d'une politique publique

Exemple : supposons que l'on veuille étudier les effets de pairs dans le milieu éducatif

- Définition des effets de pairs = un effet causal = idée que les résultats scolaires des élèves sont influencés par les résultats scolaires de leurs pairs
 - Les pairs auraient un effet causal sur les individus
- Mais si on observe une corrélation entre les résultats scolaires des élèves et les résultats scolaires de leurs pairs, il faut bien avoir en tête que cette corrélation peut-être (complètement ou en partie) due à d'autres phénomènes qu'un effet causal
- Cette corrélation peut être due à un effet de sélection
 - Exemple : il y aura un effet de sélection si les individus sont sélectionnés pour entrer dans les établissements scolaires en fonction de leurs résultats scolaires antérieurs
 - Cet effet de sélection conduit à une concentration des élèves ayant le même niveau dans certains établissements. La sélection génère une corrélation entre les résultats des élèves et les résultats de leurs pairs

ÉVALUATION D'IMPACT : DEUX ERREURS COURANTES (SUITE)

Une seconde erreur : Erreur consistant à comparer la situation des bénéficiaires « avant / après » la PP : erreur qui consiste à comparer la variable de résultat après le programme avec la variable de résultat avant le programme

Ce n'est pas une bonne idée, parce que de nombreux facteurs peuvent expliquer des variations (de la variable de résultat) à travers le temps, indépendamment du programme

- Exemple : la conjoncture économique peut s'améliorer ou se dégrader

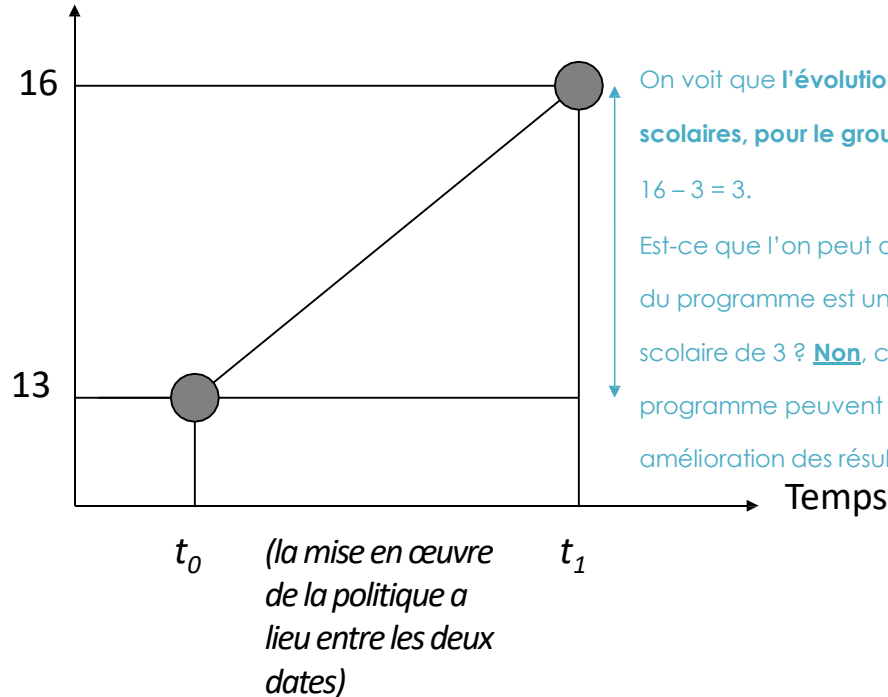
EXEMPLE

Par exemple, dans le domaine de l'éducation, la mise en œuvre d'un certain programme entraîne-t-elle une amélioration des résultats scolaires ?

La figure représente l'évolution des résultats scolaires **pour le groupe de bénéficiaires** (les traités)

On observe seulement ce qui est représenté par des boules.

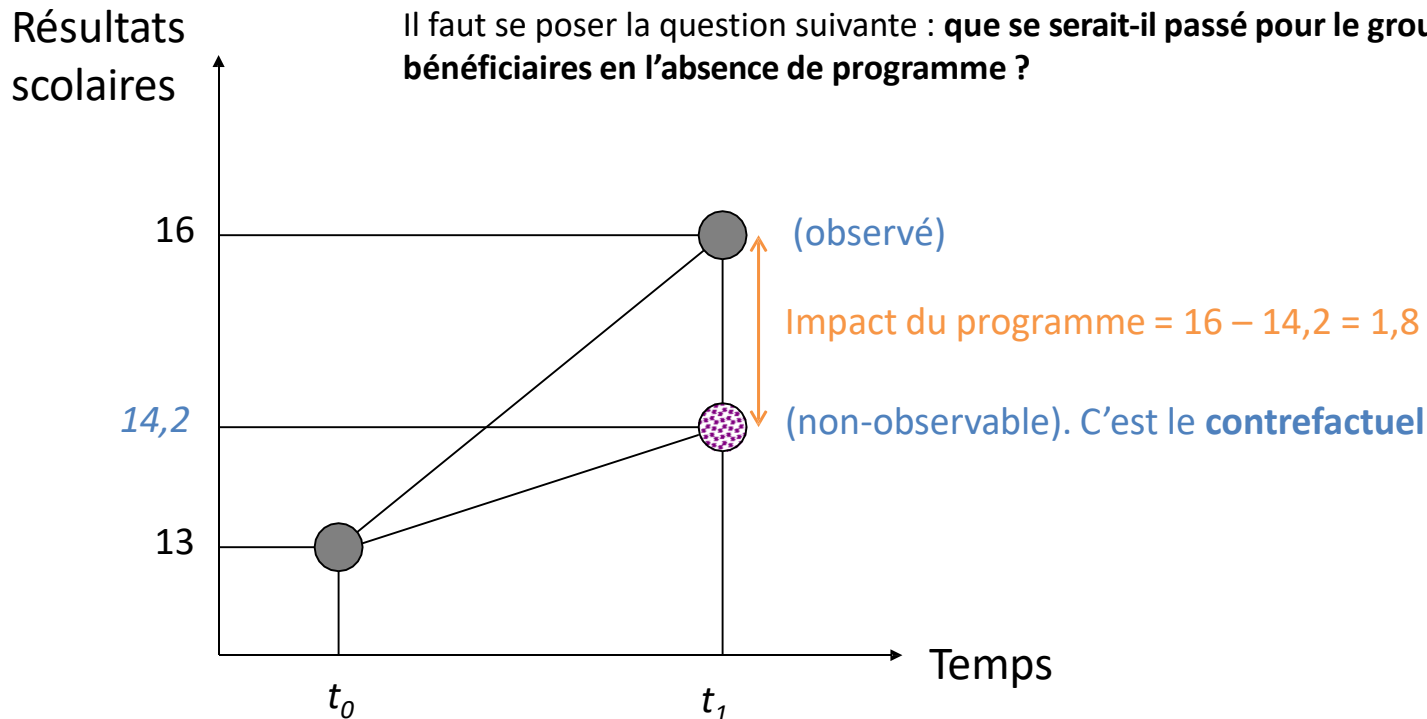
Résultat scolaire



INTRODUCTION À LA NOTION DE CONTREFACTUEL

Que faut-il faire alors pour évaluer l'impact d'un programme ?

Il faut se poser la question suivante : **que se serait-il passé pour le groupe de bénéficiaires en l'absence de programme ?**



CADRE D'ANALYSE DE RUBIN : LE CONTREFACTUEL

L'évaluation d'impact vise ainsi à déterminer quels changements peuvent être attribués exclusivement au programme

On se place dans le cadre d'analyse dit « de **Rubin** »

- qui systématise le recours à un **contrefactuel hypothétique** pour comparer la situation des bénéficiaires de la mesure que l'on souhaite évaluer avec celle qui aurait prévalu en son absence

L'effet causal est estimé en comparant

- (a) la situation observée des bénéficiaires de la PP (après la mise en œuvre de la PP)
- avec
- (b) la situation qu'auraient connus ces bénéficiaires si la PP n'avait pas existé
 - c-à-d si la politique n'avait pas été mise en œuvre

(b) est appelé la situation contrefactuelle ou le contrefactuel. C'est une situation hypothétique = la situation des bénéficiaires en l'absence de politique

LE CONTREFACTUEL ET LE GROUPE DE CONTRÔLE

La situation contrefactuelle = les résultats qu'auraient connus les bénéficiaires d'une PP si ladite PP n'avait pas existé

Il est fondamental d'avoir un bon contrefactuel

Le contrefactuel est une situation qui n'est **pas observable**, donc il faut le construire ou l'estimer

Pour cela, on va chercher (ou construire) un groupe d'**individus qui n'ont pas bénéficié de la PP mais qui sont les plus comparables possibles aux bénéficiaires de la PP**

Ces individus sont appelés **le groupe de contrôle** (ou le groupe témoin)

La situation contrefactuelle est estimée en utilisant la situation de ce groupe de contrôle

ILLUSTRATION

On considère un programme qui consiste à proposer des stages de formation à des chômeurs

Est-ce que le programme augmente les chances de retour à l'emploi ?

- On veut savoir si le programme a un effet sur la variable de résultat, qui désigne ici le fait d'avoir un emploi ou pas (après la mise en œuvre du programme)

Prenons un bénéficiaire :

- Y_i^1 représente la variable de résultat du bénéficiaire. On observe cette variable (c-à-d que l'on sait si l'individu a retrouvé un emploi ou pas, après le programme)
- Y_i^0 représente la variable de résultat du bénéficiaire s'il n'avait pas bénéficié du programme. C'est le contrefactuel. Cette situation est inobservable par nature

Pour les bénéficiaires, on observe Y_i^1 mais on n'observe pas Y_i^0

Pour évaluer l'impact du programme, on a besoin d'estimer la situation contrefactuelle

AVEC DES MATHÉMATIQUES

On s'intéresse à l'évaluation d'une PP / d'un programme / d'un traitement T / d'une mesure

On distingue

- les bénéficiaires du programme ($T = 1$)
- des personnes qui n'en bénéficient pas ($T = 0$)

On s'intéresse à l'effet de ce programme sur une variable de résultat

On considère que chacun a « virtuellement » deux résultats potentiels, selon qu'il bénéficie ou non du programme

- On note Y^0 la variable de résultat associée à $T = 0$ (pas de traitement)
- On note Y^1 la variable de résultat associée à $T = 1$ (traitement)

Pour **un** individu i , l'effet causal du programme est la différence entre les deux résultats potentiels :

$$\text{Effet causal} = Y_i^1 - Y_i^0$$

MATHÉMATIQUES (SUITE)

Point important :

Pour un individu bénéficiaire du programme, on va observer Y_i^1 mais on ne va pas observer Y_i^0

NB : Pour un individu non bénéficiaire du programme, on pourrait observer Y_i^0 mais on ne pourrait pas observer Y_i^1

L'EFFET MOYEN DU TRAITEMENT SUR LES TRAITÉS

Auparavant, on a écrit l'effet du traitement pour **UN** individu i . Cet effet était $Y_i^1 - Y_i^0$
Pour évaluer l'impact de la mesure, on veut évaluer l'effet **moyen** de la mesure sur **les** individus qui en ont bénéficié

- C'est l'**effet moyen du traitement sur les traités** (*Average Treatment effect on the Treated, ATT*), qui s'écrit

$$\begin{aligned} ATT &= E(Y_i^1 - Y_i^0 \mid T_i = 1) \\ ATT &= E(Y_i^1 \mid T_i = 1) - E(Y_i^0 \mid T_i = 1) \end{aligned}$$

E est la notation de l'espérance, qui est l'autre nom donné à la moyenne, en mathématiques
L'ATT est la différence entre l'*outcome* moyen avec traitement et l'*outcome* moyen que les traités auraient obtenu s'ils n'avaient pas été traités

L'ATT est la variation (dans l'*outcome*) due au traitement, pour les personnes traitées

Le premier terme ($E(Y_i^1 \mid T_i = 1)$) est observé

Le second terme ($E(Y_i^0 \mid T_i = 1)$) n'est pas observé. Donc il faut trouver un **estimateur** de ce second terme

Tout l'enjeu des méthodes d'évaluation est de proposer une estimation crédible de cette situation virtuelle

ATT ET EFFETS DE SÉLECTION

On revient à la première erreur dont on a parlé auparavant : il s'agit de la mauvaise idée consistant à évaluer l'effet causal en calculant la différence entre l'*outcome* des bénéficiaires $E(Y_i^1 \mid T_i = 1)$ et l'*outcome* des non-bénéficiaires

Cette erreur consiste à utiliser l'*outcome* des non-bénéficiaires comme estimateur du contrefactuel $E(Y_i^0 \mid T_i = 1)$

C'est souvent un mauvais contrefactuel, en général, parce qu'il y a des effets de sélection

Cet estimateur « naïf » va donc conduire à attribuer au programme ce qui relève des spécificités des bénéficiaires, c'est-à-dire ce qui relève d'un **effet de sélection**

→ En général, cet estimateur « naïf » n'est donc pas convaincant

Une bonne évaluation nécessite de comprendre les mécanismes de sélection pour pouvoir **neutraliser les effets de sélection**

HYPOTHÈSE : L'ABSENCE D'EXTERNALITÉS

Dans les évaluations d'impact, on fait généralement l'hypothèse **d'absence d'externalités**, c'est-à-dire l'hypothèse que le programme a un effet seulement sur les bénéficiaires et n'a pas d'effet sur d'autres personnes

Une externalité peut être positive (bonne) ou négative (nuisible)

Attention : parfois l'hypothèse d'absence d'externalités n'est pas valable. Exemple d'une politique qui a une externalité positive :

- Campagnes de distribution de moustiquaires pour lutter contre le paludisme. Les moustiquaires sont imprégnées d'insecticide qui tue les moustiques. Il y a diverses façons de faire ces distributions. Exemple : porte-à-porte ; distribution aux femmes lorsqu'elles viennent au centre de santé (pour un suivi de grossesse ou une vaccination d'enfant, par exemple) ; etc.
- Les distributions peuvent potentiellement protéger les **voisins** des ménages qui ont reçu des moustiquaires, même s'ils n'en ont pas reçues eux-mêmes

Si on sait qu'une politique génère des externalités, alors on doit essayer de les prendre en compte dans l'évaluation (par exemple, en raisonnant à un niveau plus agrégé dans une expérimentation)

- NB : dans un modèle structurel, on peut tenir compte des externalités

COMMENT TROUVER OU CONSTRUIRE UN « BON » CONTREFACTUEL ?

Supposons que l'on veuille évaluer l'impact du programme. Pour trouver ou construire le contrefactuel, plusieurs méthodes possibles
Ces méthodes neutralisent plus ou moins bien les biais de sélection (dus aux caractéristiques observables et inobservables)

CONTREFACTUELS ET MÉTHODES : EXPÉRIENCES CONTRÔLÉES

* **Méthode expérimentale.** La méthode de l'**expérimentation contrôlée** :

Vocabulaire : *randomised controlled trials (RCT)*

L'approche de l'expérimentation contrôlée est issue du champ médical. Elle a ensuite été utilisée dans le champ social

Le principe, dans le cas le plus simple :

On **tire au hasard** les personnes qui vont bénéficier du programme et les personnes qui ne vont pas en bénéficier

Le groupe de traitement (*treatment group*) = ceux qui bénéficient du programme

Le groupe de contrôle (*control group*) = ceux qui ne bénéficient pas du programme

L'appartenance aux groupes résulte d'un **tirage au sort**, d'un processus aléatoire, du hasard

Golden standard de l'évaluation

EXPÉRIENCES CONTRÔLÉES

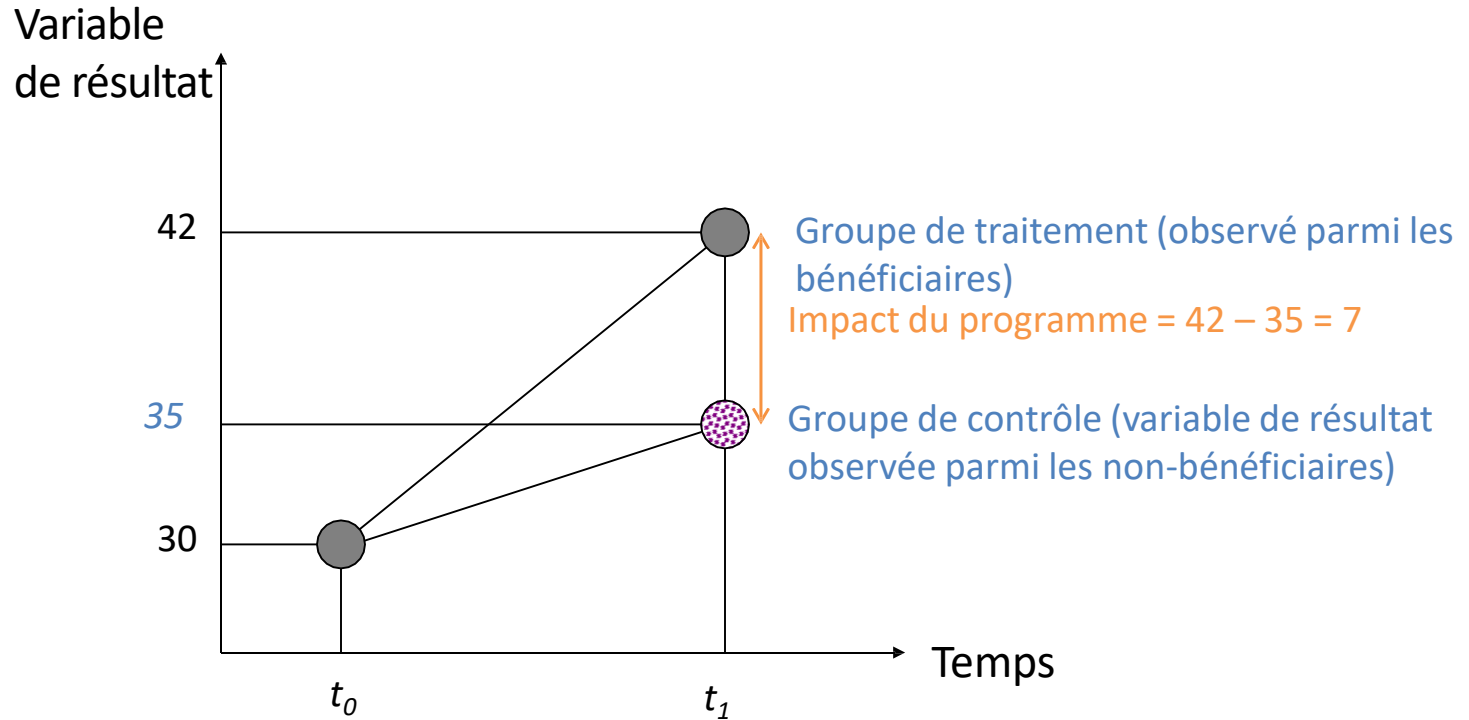
Le tirage au sort garantit que les 2 groupes, s'ils sont assez grands, sont comparables

On peut le vérifier si on a des données pour cela

Puis on met en œuvre le programme

Le contrefactuel correspond à la situation du groupe de contrôle

On mesure l'effet causal en comparant (après le programme) la situation du groupe de traitement et celle du groupe de contrôle



CONTREFACTUELS ET MÉTHODES : EXPÉRIENCES CONTRÔLÉES

On a parlé d'une expérience avec 1 groupe de contrôle et 1 groupe de traitement. Mais une expérience peut être plus complexe

- Par exemple, il peut y avoir 1 groupe de contrôle et 2 groupes de traitement (correspondant à 2 traitements différents). Comme cela fait 3 groupes, on parlerait en anglais de « three-arm trial » (expérience aléatoire à trois bras).

Exemples de RCT

- Les travaux d'Esther Duflo
- Le réseau de chercheurs J-PAL (*Abdul Latif Jameel Poverty Action Lab*)
 - Lien vers J-PAL Europe : <https://www.povertyactionlab.org/fr/europe>

Un des intérêts des approches expérimentales : expérimenter les effets des politiques sur de petits échantillons d'individus aussi soigneusement contrôlés que possible

- avant d'envisager la mise en œuvre de la politique en vraie grandeur (si l'efficacité du programme est démontrée)

Une expérimentation nécessite une intervention très en amont du lancement du programme à proprement parler

CONTREFACTUELS ET MÉTHODES : EXPÉRIENCES CONTRÔLÉES

NB : il existe aussi des expériences menées en laboratoire

- Exemple : il y a un laboratoire à la Maison des sciences économiques (MSE), Paris 1-Sorbonne, Boulevard de l'hôpital, 75013 Paris

Aujourd'hui, ce que l'on appelle « économie expérimentale » comprend

- les expériences en laboratoire
- les expériences randomisées (RCTs)
- les expériences naturelles

Souvent, les statisticiens évaluent des politiques ex post (après leur mise en œuvre). Dans ce cas, l'expérimentation n'est plus possible et il faut se tourner d'autres méthodes

CONTREFACTUELS ET MÉTHODES : MÉTHODES QUASI-EXPÉRIMENTALES

Ces méthodes présentées dans la suite du cours (**e.g. différence-de-différences, variables instrumentales et régressions sur discontinuité**) sont des méthodes **quasi-expérimentales** sur données observées

- On parle aussi d'« expériences naturelles » ou de « quasi-expériences » car ces méthodes tentent de reproduire le cadre des expériences contrôlées
- Tout le travail est de reconstruire un « bon » groupe témoin pour mesurer le contrefactuel en l'absence de tirage au sort
- Estimation de l'impact causal du traitement ou de la politique repose sur une **hypothèse d'identification** : le groupe traité ou impacté par la réforme aurait été similaire (ou évolué de la même manière) que le groupe témoin en l'absence du traitement ou de la réforme

CONTREFACTUELS ET MÉTHODES : MÉTHODES QUASI-EXPÉRIMENTALES

Les méthodes que nous allons voir sont une boîte à outils mais la qualité de l'évaluation repose sur:

- **Le bon sens** : certaines idées, même avec la bonne méthode, ne sont pas convaincantes en pratique
- **Une bonne connaissance du contexte institutionnel et des données** : appliquer une méthode presse-bouton à une réforme qu'on ne comprend pas amène à de grosses erreurs (écueil répandu chez les économistes)
- **De l'imagination**: souvent évaluer semble impossible, mais avec de bonnes idées on peut s'en sortir
 - **Exemples**: effet de la taille des classes sur les résultats scolaires, effet d'avoir un enfant sur l'offre de travail des femmes

Une méthode statistique/économétrique permet la mise en œuvre d'une évaluation d'impact mais elle ne suffit pas à résoudre les problèmes d'identification d'effet causal

⇒ Très grande source de confusion, y compris chez des chercheurs chevronnés

Exemple: l'emploi du terme analyse « ceteris paribus » ou « toutes choses égales par ailleurs » pour une régression avec variables de contrôles

VALIDITÉ INTERNE ET EXTERNE

QUALITÉ D'UNE ÉVALUATION D'IMPACT : VALIDITÉ INTERNE ET EXTERNE

La qualité d'une évaluation d'impact dépend de sa validité interne et de sa validité externe

La **validité interne** : rigueur de la méthodologie pour capter l'effet causal

- Qualité des données : un échantillon suffisamment grand, un échantillon représentatif
- Qualité de la méthode d'estimation : est-ce qu'on capte vraiment un effet causal du dispositif, et non pas un effet causal mélangée à d'autres effets (comme un effet de sélection) ?
 - En particulier, l'approche doit neutraliser les effets de sélection, s'il y en a
- La validité interne d'études précédentes/anciennes est souvent questionnée

QUALITÉ D'UNE ÉVALUATION D'IMPACT : VALIDITÉ INTERNE ET EXTERNE

La **validité externe**

- Question de la généralisation des conclusions (dans une perspective d'élargissement de l'application de la PP), c-à-d de la possibilité d'extrapoler les résultats à d'autres contextes ou populations
 - Est-ce que la PP peut s'appliquer à des groupes et contextes autres que ceux qui ont été étudiés dans l'évaluation ?
- La question de la validité externe est très importante pour les **décideurs**
- Les résultats d'une évaluation d'impact sont particulièrement utiles lorsque les conclusions peuvent être appliquées à une population plus large

**BIAIS DE PUBLICATION (QUAND ON
CHERCHE, ON TROUVE...)**

LES PLANS DE PRÉANALYSE

Dans certains cas, principalement pour les expériences contrôlées, on peut rédiger un plan de préanalyse (*pre-analysis plan*)

- Qui décrit les étapes et le protocole de recherche
- Dans ce cas, on va spécifier :
 - Les variables utilisées
 - Les procédures pour vérifier et recoder les variables
 - Les modèles statistiques qu'on utilisera

Le plan de préanalyse peut être enregistré, c-à-d déclaré dans un registre public

Certaines revues (en éco, psycho, etc.) apprécient qu'il y ait un plan de préanalyse
Un plan de préanalyse n'est pas indispensable de nos jours

Critique du plan : il est difficile d'anticiper toutes les possibilités

Avantages du plan : avoir un plan de préanalyse permet de

- limiter les pressions ou incitations sur les chercheurs (exemple : pressions politiques pour trouver un certain résultat)
- réaliser des études plus fiables et transparentes

LE PHÉNOMÈNE DU *P-VALUE HACKING*

Pourquoi plus fiables et transparentes ?

Dans le monde scientifique/académique, il existe un **biais de publication** (*publication bias*)

- Les études scientifiques dont les résultats sont statistiquement non significatifs (càd celles qui ne peuvent conclure à l'existence d'effets) ne sont généralement pas aussi facilement acceptées pour publication dans les revues scientifiques (avec comité de lecture) que les études qui trouvent des résultats significatifs
 - Les études qui concluent à l'absence d'effets restent souvent « dans le tiroir » / ne sont pas publiées. Parfois, personne ne sait qu'elles existent...

Ce contexte de biais de publication crée, pour les chercheurs, une incitation insidieuse à un comportement inapproprié : adapter leurs méthodes ou « torturer » les données, afin d'obtenir des résultats statistiquement significatifs

- Ces manipulations visent à obtenir des « *p values* » significatives → Ce phénomène est appelé le *p-value hacking* ou *p-hacking*
- Synonymes : *data dredging*, *data fishing*, *data butchery*, etc.
- Cette pratique est très difficile à repérer

BIAIS DE PUBLICATION ET *P-HACKING*

C'est gênant pour guider les décisions publiques

- Ex : les publications scientifiques laissent penser qu'une politique a eu des effets positifs... alors qu'en réalité, en général, elle n'a pas eu d'effet, mais les travaux de recherche montrant cette absence d'effet ne sont pas publiés...
- → Il est recommandé de faire des revues de littérature systématiques : il faut lire les articles publiés dans les grandes revues, mais aussi les documents de travail (non publiés dans des revues)
- Il faut s'intéresser aux méta-analyses et revues de littérature existantes et voir si elles prennent cela en compte

Avoir un **plan de préanalyse** permet de limiter ce problème de *p-hacking* (on ne peut pas modifier la façon de préparer les données ou le modèle statistique) → Analyse plus fiable et transparente

P-VALUE: DÉFINITION

La **p-value** est un outil utilisé en statistiques pour répondre à la question: « Les résultats que j'observe sont-ils dus au hasard, ou y a-t-il une raison spécifique (une vraie différence, un effet, etc.) ? »

On fait une hypothèse (hypothèse nulle):

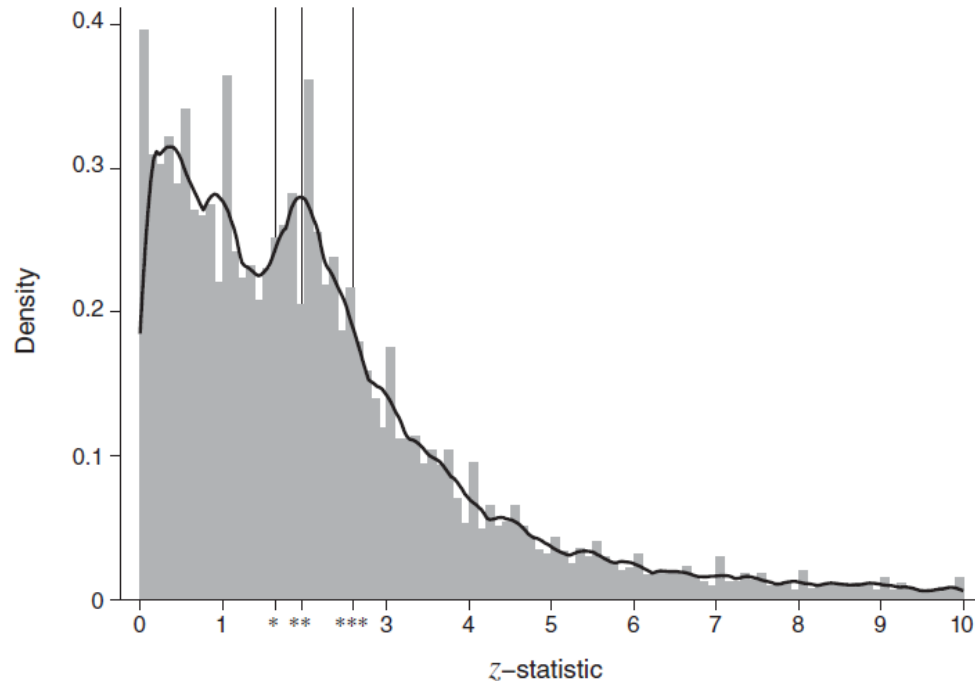
- Par exemple: la politique n'a pas d'effet = le paramètre d'intérêt que j'estime pour capturer l'effet de la politiques avec mes données vaut 0

La p-value donne la probabilité d'observer la valeur estimée du paramètre sous l'hypothèse nulle

- C'est aussi la probabilité de se tromper en rejetant l'hypothèse nulle (= en concluant que la politique a un effet)
- Convention: effet significatif si $p < 0,05$.
- Les seuils de 10% et 1% sont aussi utilisés dans les articles et indiqués avec des « * »

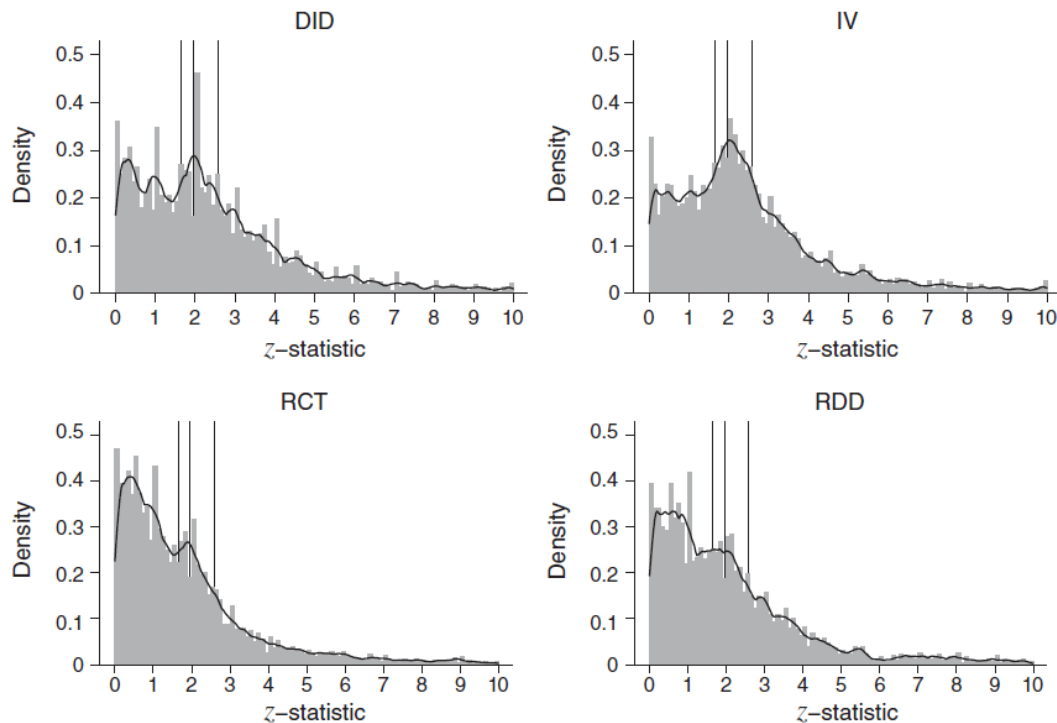
Capture l'incertitude liée au fait qu'on a seulement un nombre limité d'observations pour mesurer l'effet d'intérêt (une sous-population d'une population exhaustive ou infinie)
et/ou l'incertitude qu'on mesure un traitement (une réforme) dans un contexte spécifique non nécessairement reproductible

DISTRIBUTION DES STATISTIQUES DE TEST



Source: Brodeur, Abel, Nikolai Cook, and Anthony Heyes.
"Methods matter: P-hacking and publication bias in causal analysis in economics."
American Economic Review 110.11 (2020): 3634-60.

QU'EST CE QUE LE PUBLICATION BIAS



Source: Brodeur, Abel, Nikolai Cook, and Anthony Heyes. "Methods matter: P-hacking and publication bias in causal analysis in economics." *American Economic Review* 110.11 (2020): 3634-60.

FIGURE 2. Z-STATISTICS BY METHOD

EXEMPLE: SALAIRE MINIMUM ET EMPLOI (CARD ET KRUGER, 1995)

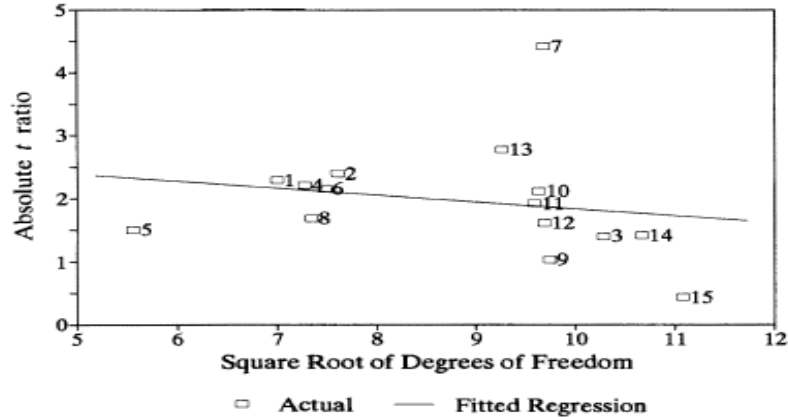


FIGURE 1. RELATION OF ESTIMATED t RATIOS TO SAMPLE SIZE

Note: The numbers in the figure refer to the following studies: 1) Hyman Kaitz, 1970; 2) Jacob Mincer, 1976; 3) Edward Gramlich, 1976; 4) Finis Welch, 1976; 5) James Ragan, 1977; 6) Michael Wachter and Choongsoo Kim, 1979; 7) George Iden, 1980; 8) Ragan, 1981; 9) John M. Abowd and Mark R. Killingsworth, 1981; 10) Charles Betsey and Bruce Dunson, 1981; 11) Brown et al., 1983; 12) Daniel Hamermesh, 1981; 13) Solon, 1985; 14) Wellington, 1991; 15) Klerman, 1992.

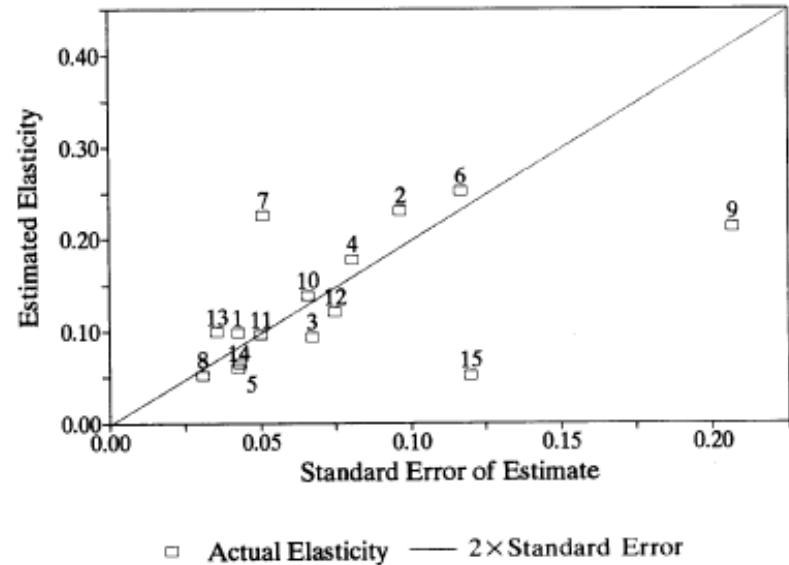


FIGURE 2. RELATION OF ESTIMATED EMPLOYMENT ELASTICITY TO STANDARD ERROR

EXEMPLE DU SALAIRE MINIMUM (ANDREWS AND KASY (2019))

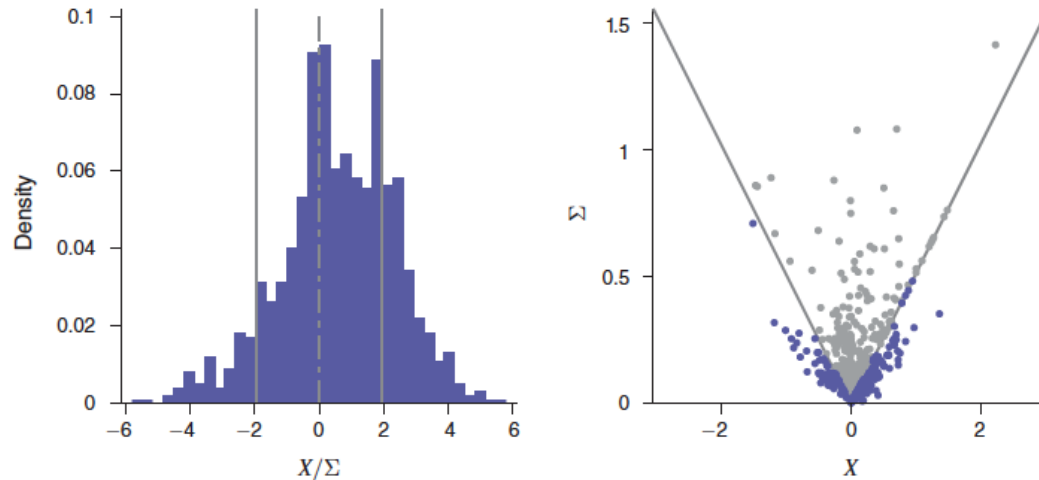


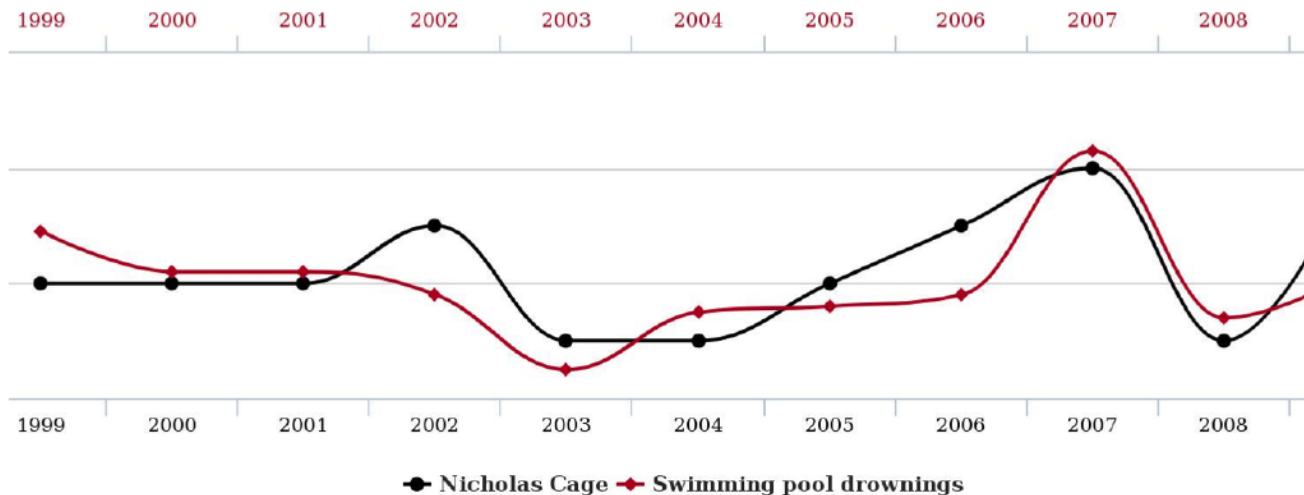
FIGURE 9. WOLFSON AND BELMAN (2015) DATA

Notes: The left panel shows a binned density plot for the z -statistics X/Σ in the Wolfson and Belman (2015) data. The solid gray lines mark $|X|/\Sigma = 1.96$, while the dash-dotted gray line marks $X/\Sigma = 0$. The right panel plots the estimate X against its standard error Σ . The gray lines mark $|X|/\Sigma = 1.96$.

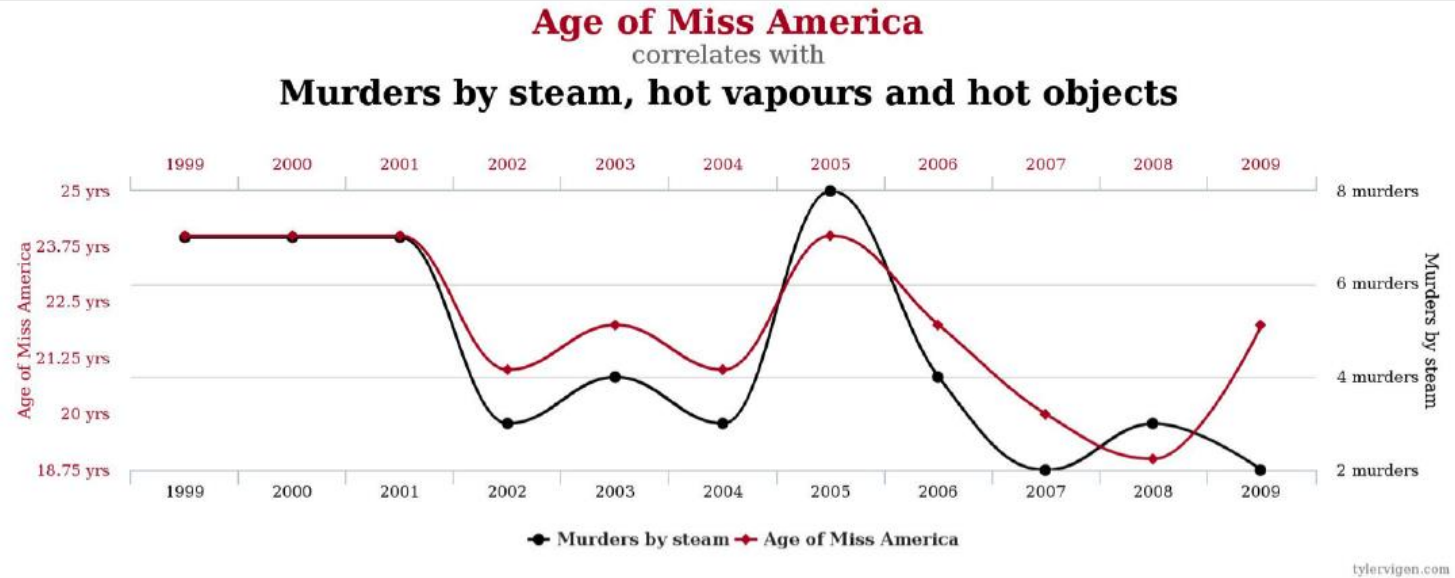
BIAIS DE SELECTION ET D'INTERPRETATION

CORRÉLATION ET CAUSALITÉ

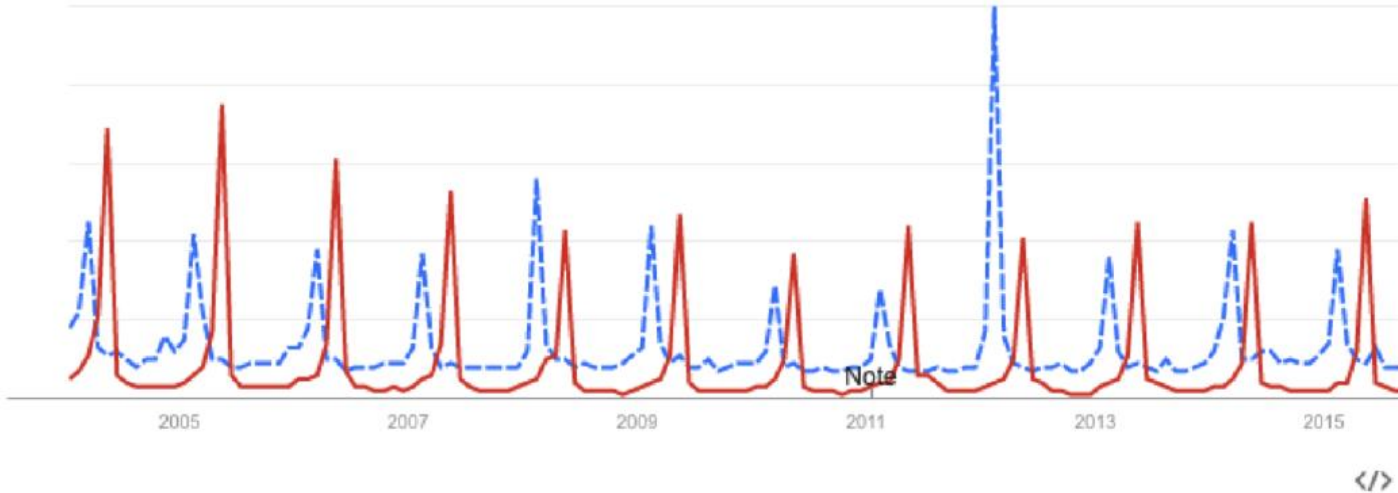
Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in



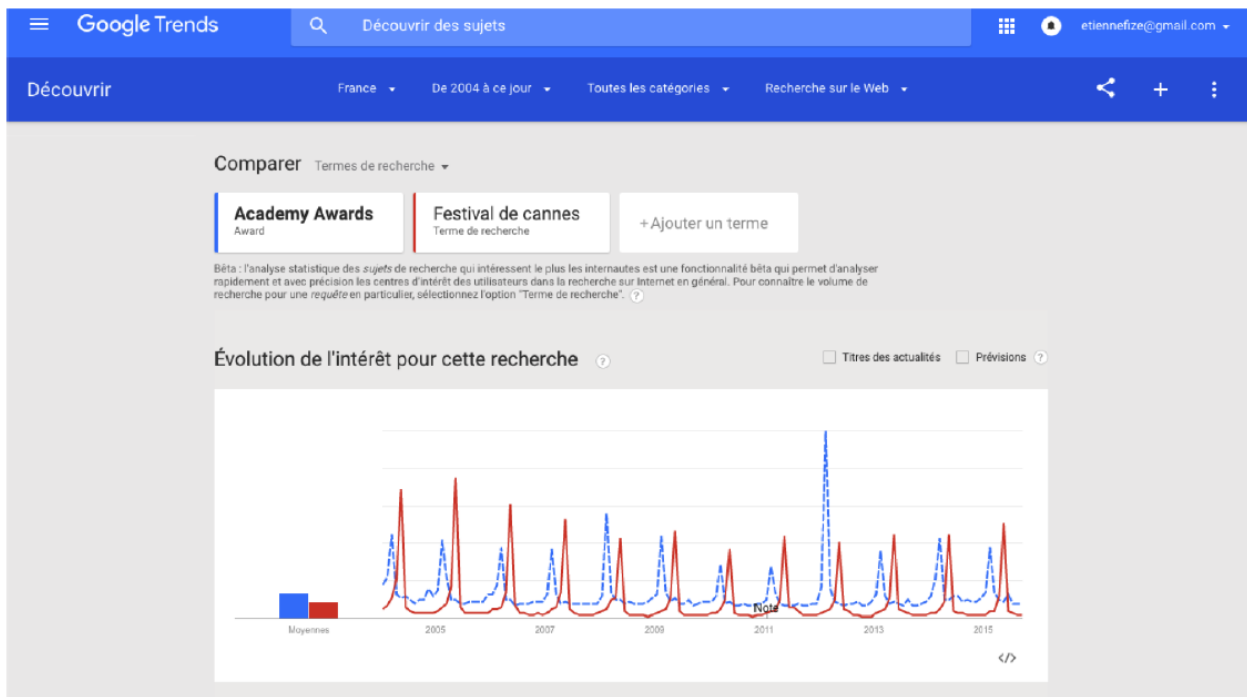
CORRÉLATION ET CAUSALITÉ



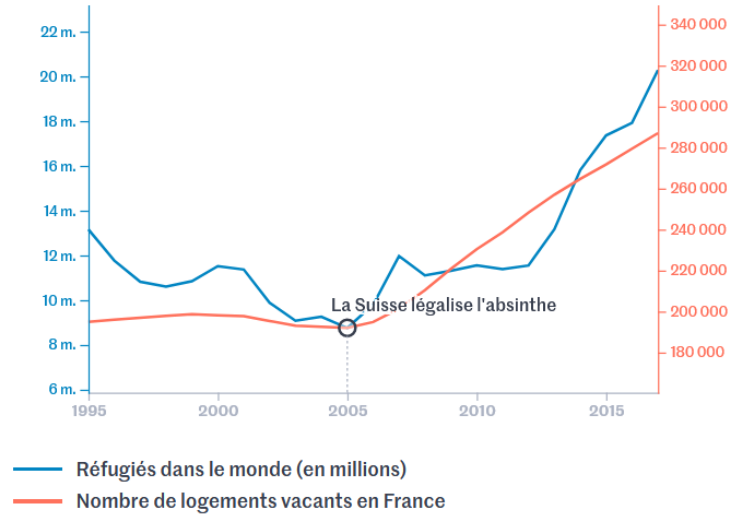
CORRÉLATION ET CAUSALITÉ



CORRÉLATION ET CAUSALITÉ



Cliquez sur le bouton pour générer une nouvelle comparaison au hasard : [ici](#), nous avons un coefficient de corrélation de 86,5% entre les deux courbes.



Une autre corrélation aléatoire

Les lycées privés en moyenne mieux classés que les lycées publics



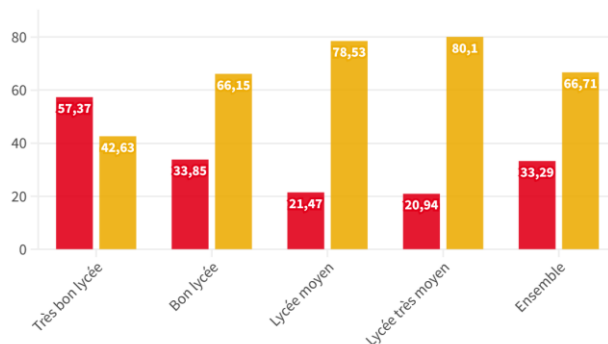
Cliquez à gauche pour comparer les établissements privés et publics selon les différents critères utilisés dans notre classement 2023 des lycées.

1 of 2

Les lycées privés en moyenne mieux classés que les lycées publics

Pourcentage d'établissements publics et privés en fonction de leur positionnement dans le classement 2023 des lycées de l'Etudiant.

Type d'établissement : ■ Privé sous contrat ■ Public



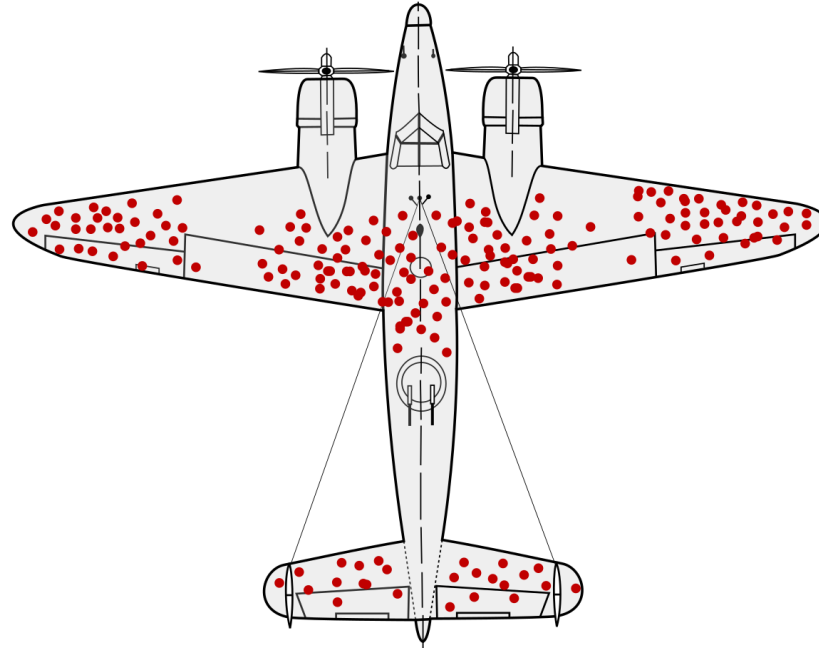
CORRÉLATION ET CAUSALITÉ

Corrélation= lien statistique entre deux variables

- Hasard et signification statistique.
- Variables omises.
- Causalité inverse.
- Multiples biais possibles ...
 - On parle d'endogénéité de la variable explicative, de « biais d'endogénéité »
 - On cherche des sources de variation « exogènes »

EXEMPLE N°1 ABRAHAM WALD LORS DE LA SECONDE GUERRE MONDIALE

Lorsque les avions revenaient de la bataille voici où ils étaient touchés:



EXEMPLE N°1

Que faire en conclusion:

- Renforcer les zones les plus touchées?
- Renforcer les autres zones?

Pourquoi?

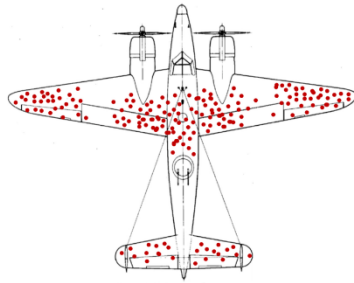
BIAIS DU SURVIVANT

Les avions qui rentraient ...
... étaient ceux qui avaient survécu

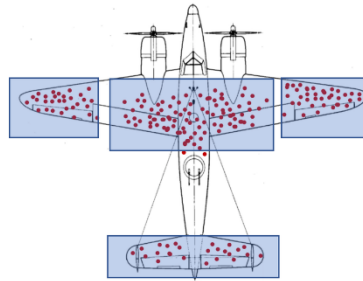
Et donc survécu malgré leurs impacts → ce qui peut signifier que ces zones n'étaient pas aussi vitales que les autres

Il faudrait à l'inverser repêcher les avions échoués ou détruit et voir là où **eux** avaient été touchés

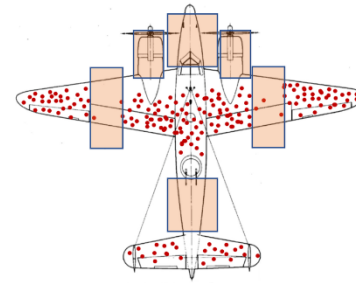
BIAIS DU SURVIVANT



Our data is only from returning flights. Here we have a visualization of the places that bullet holes were observed.



And initial guess at how to fix this might be to apply additional armor plating to the parts of the plane with the most holes...



.... However this is where planes that *returned* had bullet holes. The planes we want to protect are the ones that did *not* return, so we should place armor there.

BIAIS DU SURVIVANT

Autres applications:

- « C'était mieux avant » → on se souvient que des meilleures œuvres ou des outils qui marchent encore (voiture, vêtements...)
- « Les entreprises qui ont connu une crise s'en sortent mieux que les nouvelles entreprises » (sous entendu: donc une crise rend les entreprises plus résilients)
- Ce n'est pas parce qu'elles ont connues la crise qu'elles sont productives, mais l'inverse c'est parce qu'elles étaient productives qu'elles ont survécu

MARCHE AUSSI POUR AGE ET SOUTIEN À LA RÉFORME DES RETRAITES

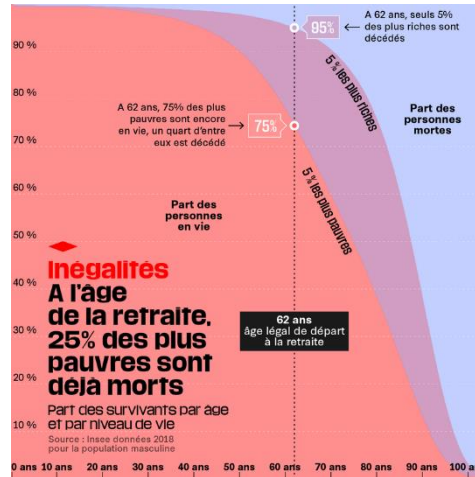
Détail des résultats

Par catégories sociodémographiques et professionnelles

Aujourd'hui, l'âge légal de départ à la retraite est de 62 ans. Selon vous, faudrait-il augmenter ou baisser l'âge légal de départ à la retraite ?

En %	Il faudrait l'augmenter	Il devrait rester à 62 ans	Il faudrait l'abaisser	Sans opinion
Ensemble	20	48	32	<1
SEXE DU REPONDANT				
Homme	24	46	30	
Femme	16	49	34	1
ÂGE DU REPONDANT				
18-24 ans	8	28	64	
25-34 ans	10	51	39	
35-49 ans	11	50	39	
50-64 ans	18	53	29	
65 ans et plus	43	47	9	1
PROFESSION DU REPONDANT				
Actifs	11	51	38	
Artisan, commerçant (*)	22	45	33	
CSP+	14	54	32	
Cadre, profession intellectuelle sup.	22	50	27	1
Profession intermédiaire	9	57	34	
CSP-	8	48	44	
Employé	8	47	45	
Ouvrier	6	51	42	1
Inactifs	31	44	24	1
Retraité	40	45	14	1
Autre Inactif	12	42	46	

MARCHE AUSSI POUR AGE ET SOUTIEN À LA RÉFORME DES RETRAITES

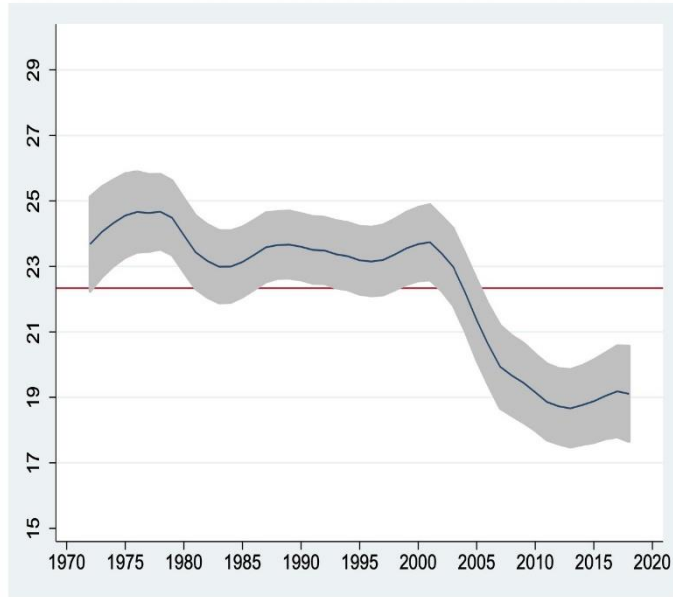


MARCHE AUSSI POUR AGE ET SOUTIEN À LA RÉFORME DES RETRAITES

Ici l'effet du soutien à la réforme des retraites selon l'âge est surement (au moins en partie) dû à un **effet de sélection** qui fait que les personnes les plus riches ont tendance à vivre plus longtemps, et donc la préférence des personnes âgées capture aussi la préférence des personnes riches.

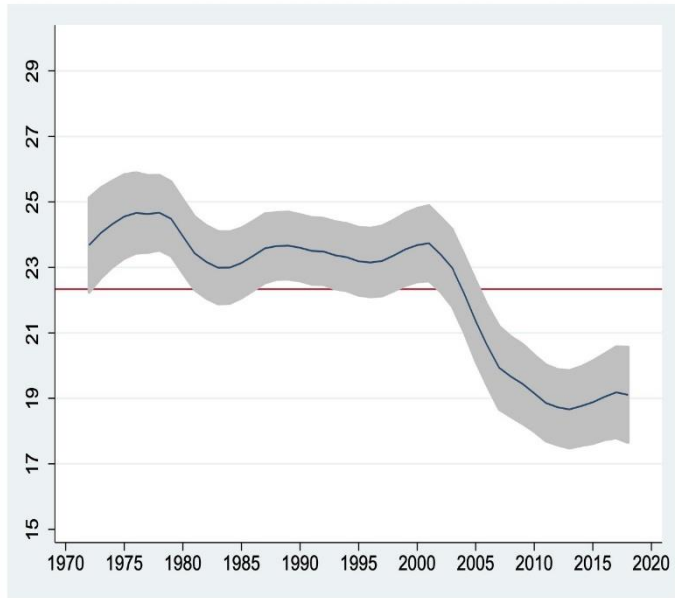
BIAIS DE COMPOSITION – PARADOXE DE SIMPSON

Figure 1. Happiness. 1972-2018



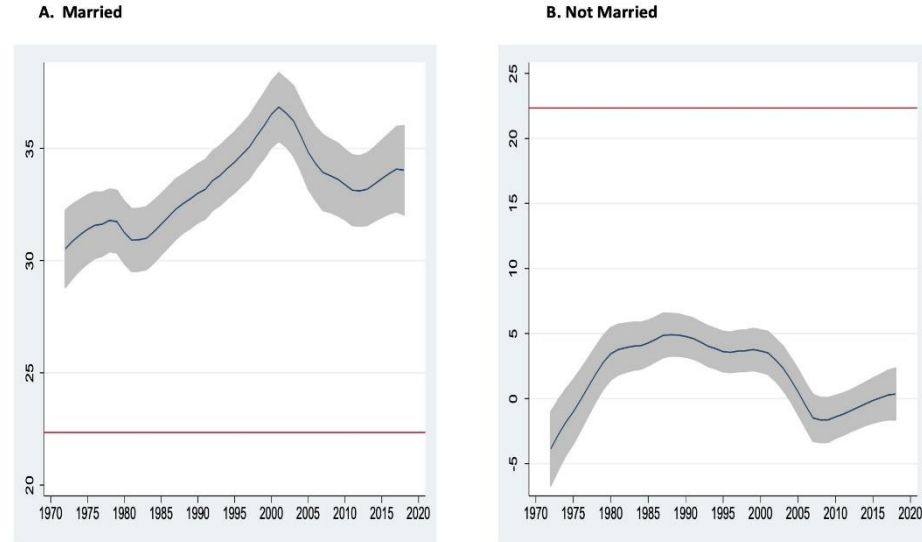
BIAIS DE COMPOSITION – PARADOXE DE SIMPSON

Figure 1. Happiness. 1972-2018



Mais →

Figure 3. Happiness by Marital Status. 1972-2018



BIAIS DE COMPOSITION – PARADOXE DE SIMPSON

Effet de composition:

- Si parmi les mariés et parmi les non-mariés : bonheur croissant
- Mais translation de beaucoup de mariés vers non-mariés (ou des nouveaux pas mariés)
- Donc bonheur moyen décroît, alors que par catégorie ça croît

EXEMPLE N°2: LE PARADOXE DE SIMPSON

Lorsque que le résultat sur tout l'échantillon est incohérent avec les résultats sur des parties de l'échantillon

Exemple:

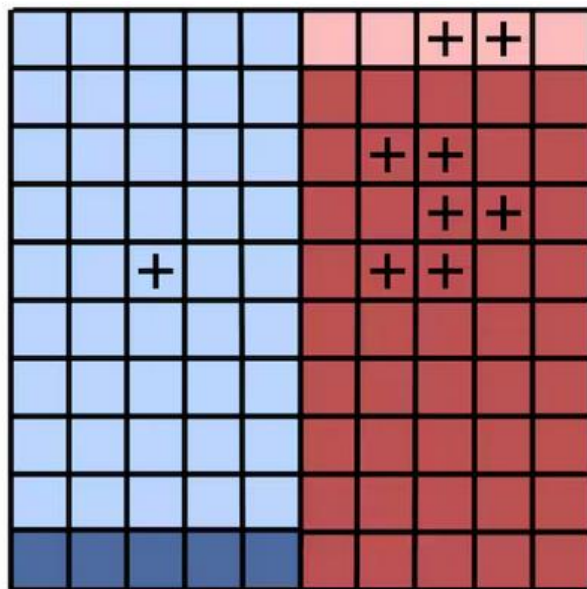
1. parmi les moins de 50 ans, le taux de mortalité est environ 1,8 fois *plus* élevé chez les non-vaccinés que chez les vaccinés (ou partiellement vaccinés) ;
2. parmi les plus de 50 ans, le taux de mortalité est environ 6,3 fois *plus* élevé chez les non-vaccinés que chez les vaccinés (ou partiellement vaccinés) ;
3. par contre, dans la population prise dans son ensemble, le taux de mortalité est environ 1,3 fois *moins* élevé chez les non-vaccinés que chez les vaccinés (ou partiellement vaccinés).

EXEMPLE N°2: LE PARADOXE DE SIMPSON

Explication de l'exemple:

- La partition n'est pas égale en termes de nombre d'individu
- Le taux de mortalité diffère largement entre les deux groupes
- Le taux de vaccination aussi

EXEMPLE N°2: LE PARADOXE DE SIMPSON








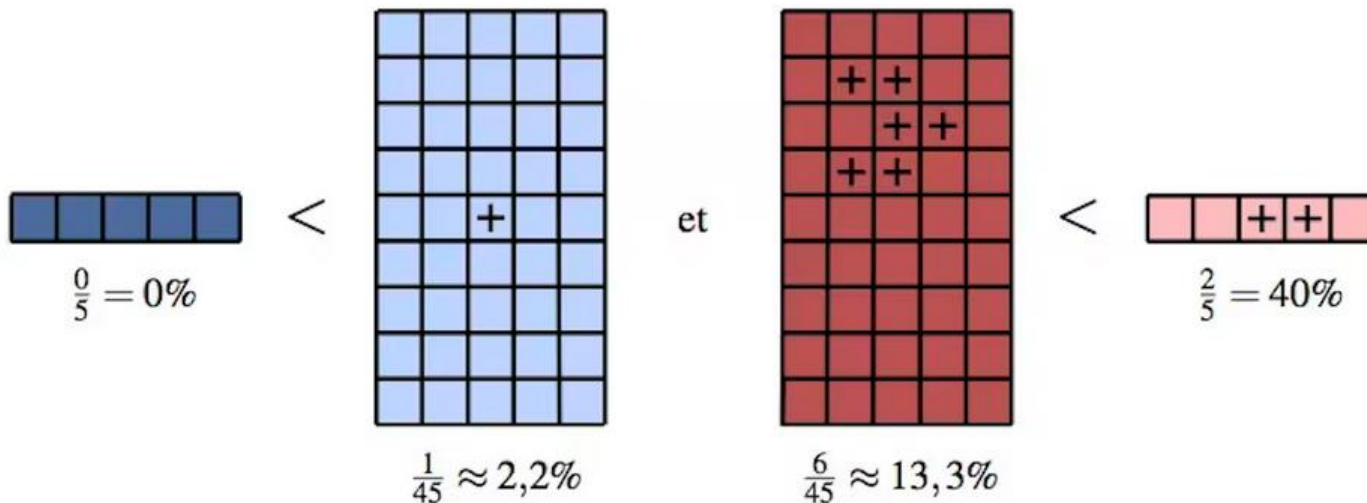
-  < 50 ans, non vaccinés
-  < 50 ans, vaccinés
-  ≥ 50 ans, non vaccinés
-  ≥ 50 ans, vaccinés
-  décès

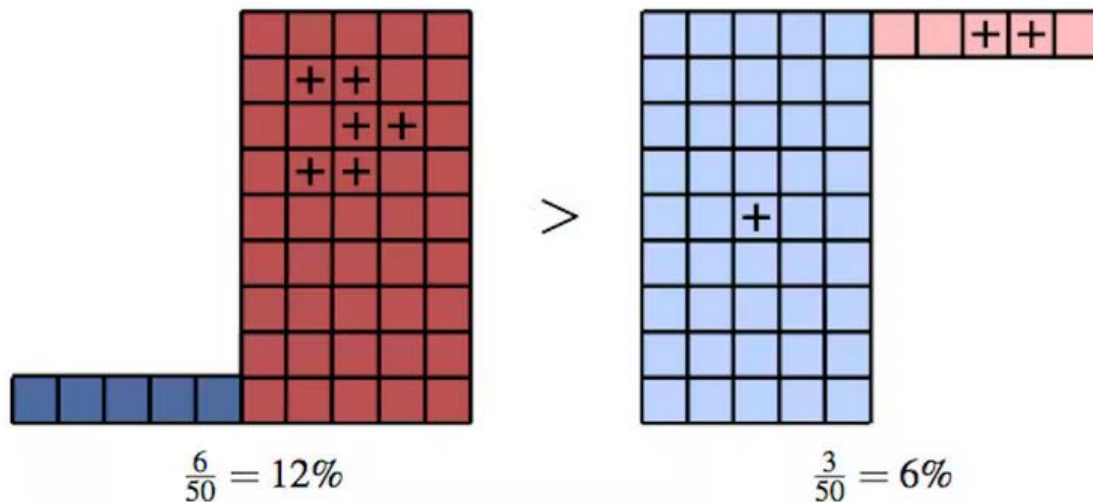
Illustration du paradoxe de Simpson avec des données caricaturales : chaque personne est représentée par un carré, la couleur du carré correspondant à une classe d'âge, la teinte foncée ou claire au statut de vaccination ; une petite croix représente un décès. On peut comparer facilement les taux de mortalité dans chaque classe d'âge et par statut de vaccination. Quentin Berger et Francesco Caravenna, Fourni par l'auteur

EXEMPLE N°2: LE PARADOXE DE SIMPSON



Pour le groupe des moins de 50 ans (bleu), le taux de mortalité est plus élevé chez les non-vaccinés (2,2 %) que chez les vaccinés (0 %). Pour le groupe des plus de 50 ans (rouge), le taux de mortalité est plus élevé chez les non-vaccinés (40 %) que chez les vaccinés (13,3 %). Quentin Berger et Francesco Caravenna, Fourni par l'auteur

EXEMPLE N°2: LE PARADOXE DE SIMPSON



Pour la population dans son ensemble, le taux de mortalité est plus élevé parmi les vaccinés (bleu foncé et rouge foncé, 12 %) que parmi les non-vaccinés (bleu clair et rouge clair, 6 %). Quentin Berger et Francesco Caravenna, Fourni par l'auteur

EXEMPLE N°2: LE PARADOXE DE SIMPSON

Conclusion à tirer:

Pas toujours facile de tirer des conclusions des données si on n'est pas attentif

Ici:

- Biais de composition:
 - Les plus vaccinés sont les plus vieux, pour qui le Covid est le plus léthal
 - Ainsi on capture l'effet d'être vieux (plus que l'effet d'être vacciné) lorsque l'on fait la comparaison

Solutions?

- Variables de contrôle
- Sélection d'un groupe de contrôle
- Analyser au sein du même groupe
- ...

LES ÉVALUATIONS D'IMPACT EN FRANCE ET DANS LE MONDE

CONTEXTE ET MÉTHODE

- Niveau de dépenses élevé ; défiance de certains citoyens par rapport aux institutions publiques → les évaluations d'impact a pour but d'administrer la preuve qu'une PP atteint son objectif
- Etude de France Stratégie sur l'évaluation d'impact dans le monde et dans 7 pays en particulier (dont la France)
 - Objectif : quelle est la production scientifique en termes d'évaluations d'impact de la France au regard des autres pays ?
 - Les évaluations d'impact considérées dans cette étude = expériences + études quasi-expérimentales
- Etude bibliométrique pour quantifier la production d'évaluations d'impact concernant la France et six autres pays
 - Recherche sur la base de données *Econlit*
 - Se focalise sur les travaux publiés comme articles de recherche ou documents de travail académiques (exclut les rapports d'évaluation internes ou non publiés comme documents académiques)

RÉSUMÉ

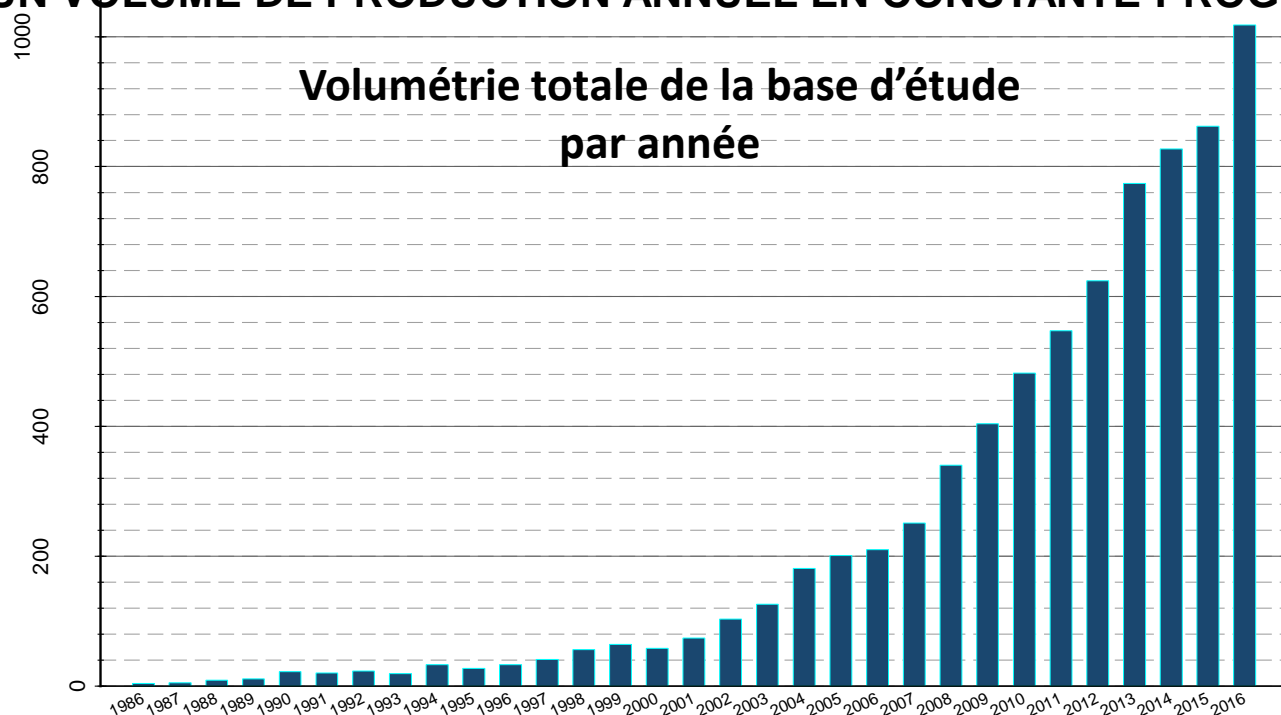
Résumé des résultats :

« [...] la France présente un certain retard dans la production d'évaluations d'impact, qu'elle comble cependant progressivement depuis peu ».

« Les questions d'emploi sont les plus traitées par les évaluations d'impact en France, comme dans les autres pays continentaux (Allemagne, Suède) ».

« Le pays se démarque cependant par le rôle important que tiennent des acteurs à mi-chemin entre le monde de la recherche et celui de l'administration dans la conduite des évaluations d'impact ».

UN VOLUME DE PRODUCTION ANNUEL EN CONSTANTE PROGRESSION

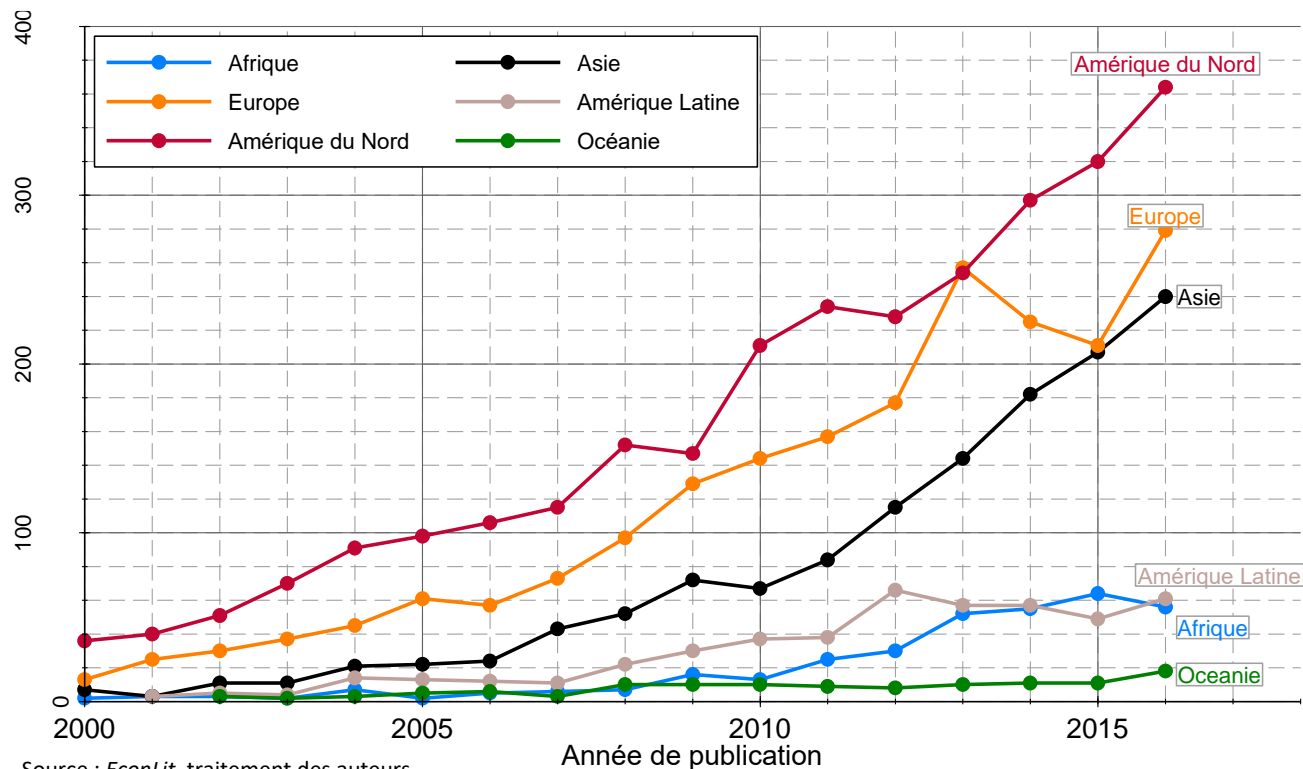


Hausse du nombre d'articles depuis les années 1980, avec un infléchissement à la hausse à partir des années 2000

Source : EconLit, traitement des auteurs.

Note de lecture : Il n'existe que très peu d'articles avant 1986.

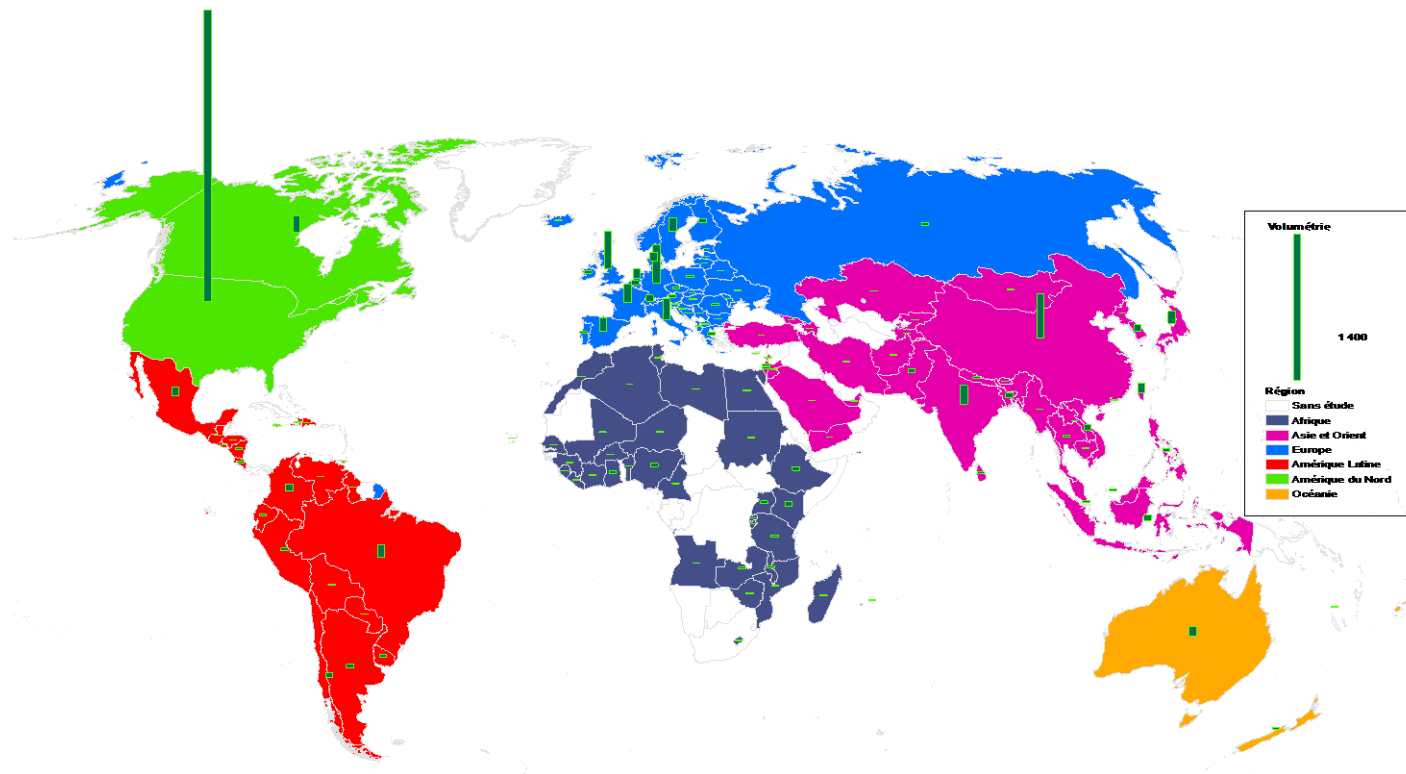
EVOLUTION PAR GRANDES RÉGIONS



Source : EconLit, traitement des auteurs.

Note de lecture : Chaque point des courbes représente le nombre d'articles publiés qui concernent un pays appartenant à la grande région. L'Amérique du Nord s'entend seulement avec le Canada et les États-Unis.

UNE PRODUCTION MONDIALE DOMINÉE PAR LES ÉTATS-UNIS



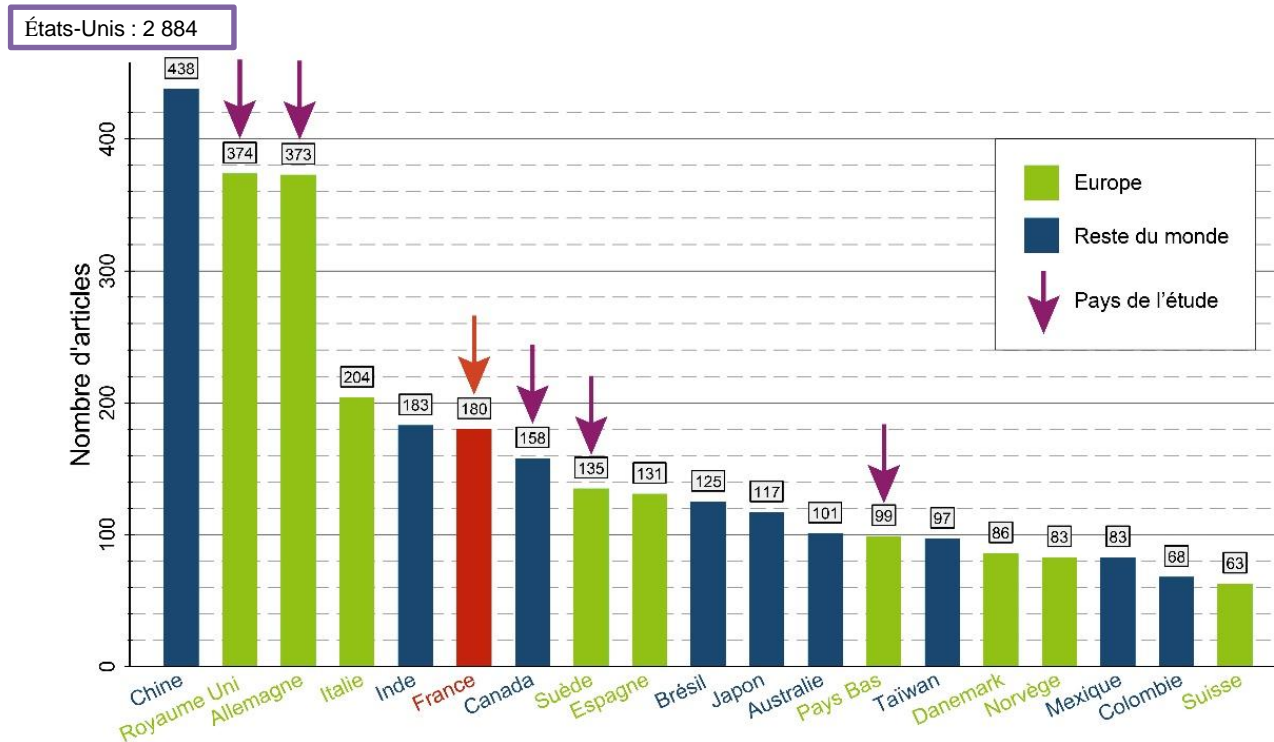
Les Etats-Unis sont de loin le pays pour lequel on dispose le plus d'évaluations d'impact.

NB : les Etats-Unis ont été le premier pays (historiquement) à produire des évaluations d'impact

Source : Traitement des auteurs, zonage défini par *EconLit*.

Note de lecture : La hauteur des histogrammes correspond au nombre total d'articles par pays dans la base de travail.

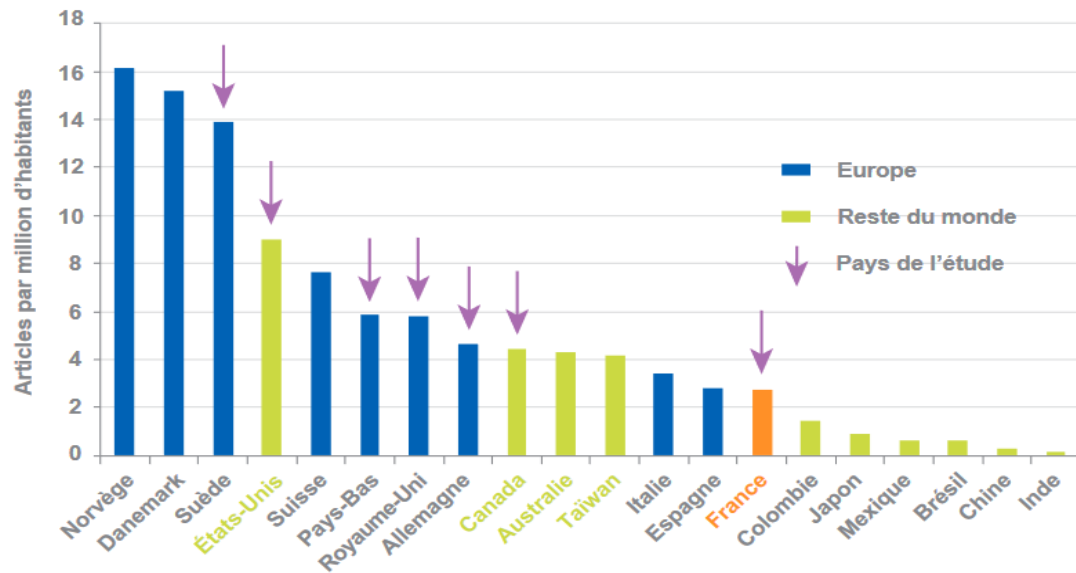
LES PAYS LES PLUS PERFORMANTS SUR L'ENSEMBLE DE LA PÉRIODE



Source : EconLit, traitement des auteurs.

Note de lecture : On reporte les 20 premiers pays en fonction du nombre d'études qui leur est associé jusqu'en 2016. La valeur pour les États-Unis est reportée en encadré pour une meilleure lecture du graphique.

Nombre d'articles d'évaluation d'impact par million d'habitants pour les 20 pays en produisant le plus

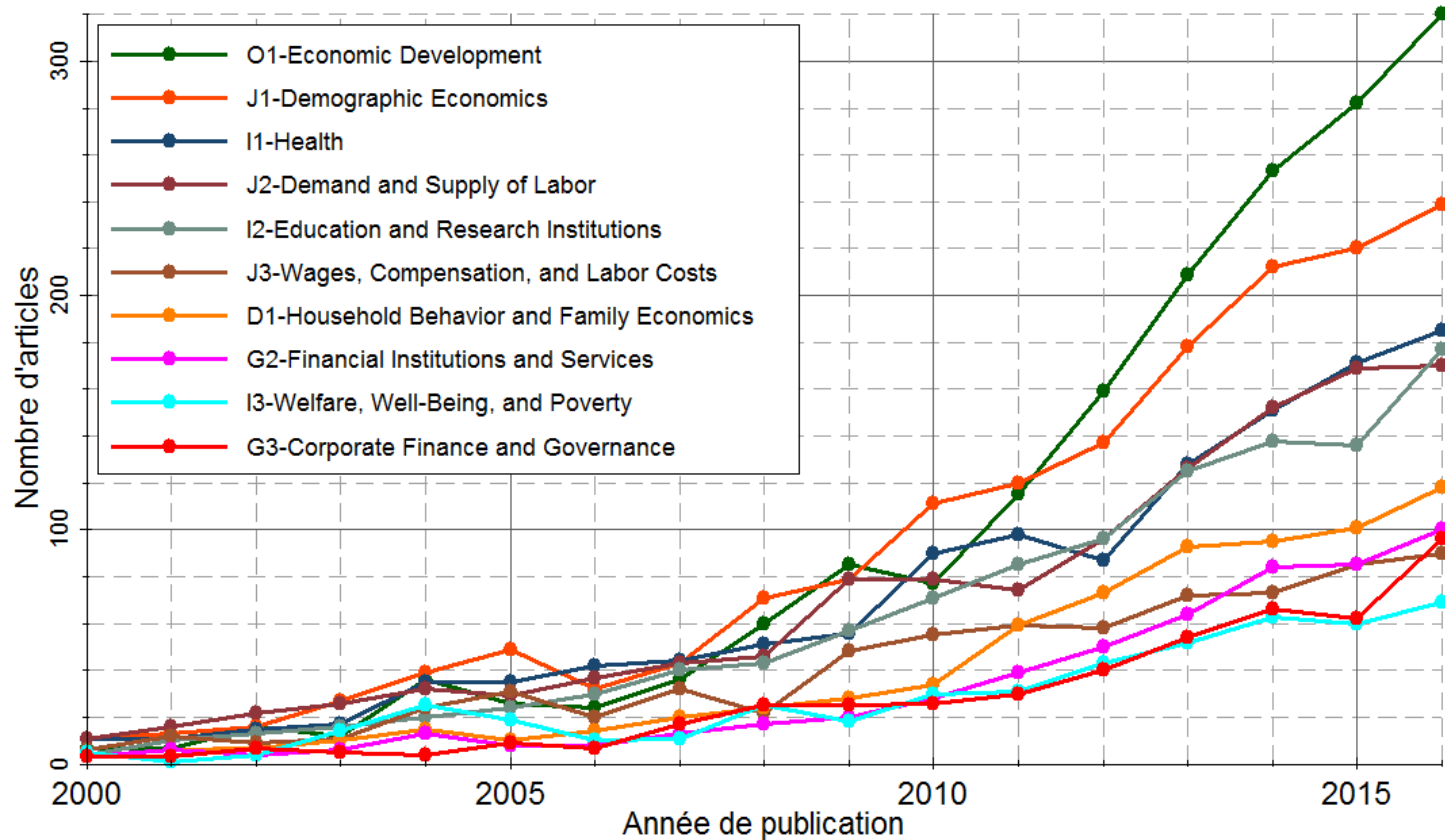


Lecture : Ce graphique exprime, pour les vingt pays ayant produit le plus d'évaluations d'impact jusqu'en 2016, le nombre d'articles divisé par la population du pays en 2014 (en million d'habitants).

Source : EconLit, Penn World Table 9.0 pour la population en 2014, traitement des auteurs

Source : https://www.nouvelle-aquitaine.fr/sites/default/files/2020-08/note_de_synthese_politique_publique.pdf (page 1)

Évolution des 10 premiers codes JEL entre 2000 et 2016



J6: Mobility, Unemployment, Vacancies, and Immigrant Workers

J3: Wages, Compensation, and Labor Costs

J2: Demand and Supply of Labor

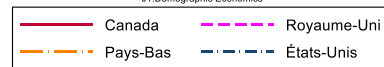
D1: Household Behavior and Family Economics

I1: Health

I2: Education and Research Institutions

I3: Welfare, Well-Being, and Poverty

J1: Demographic Economics



J6: Mobility, Unemployment, Vacancies, and Immigrant Workers

J3: Wages, Compensation, and Labor Costs

J2: Demand and Supply of Labor

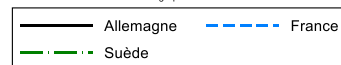
D1: Household Behavior and Family Economics

I1: Health

I2: Education and Research Institutions

I3: Welfare, Well-Being, and Poverty

J1: Demographic Economics



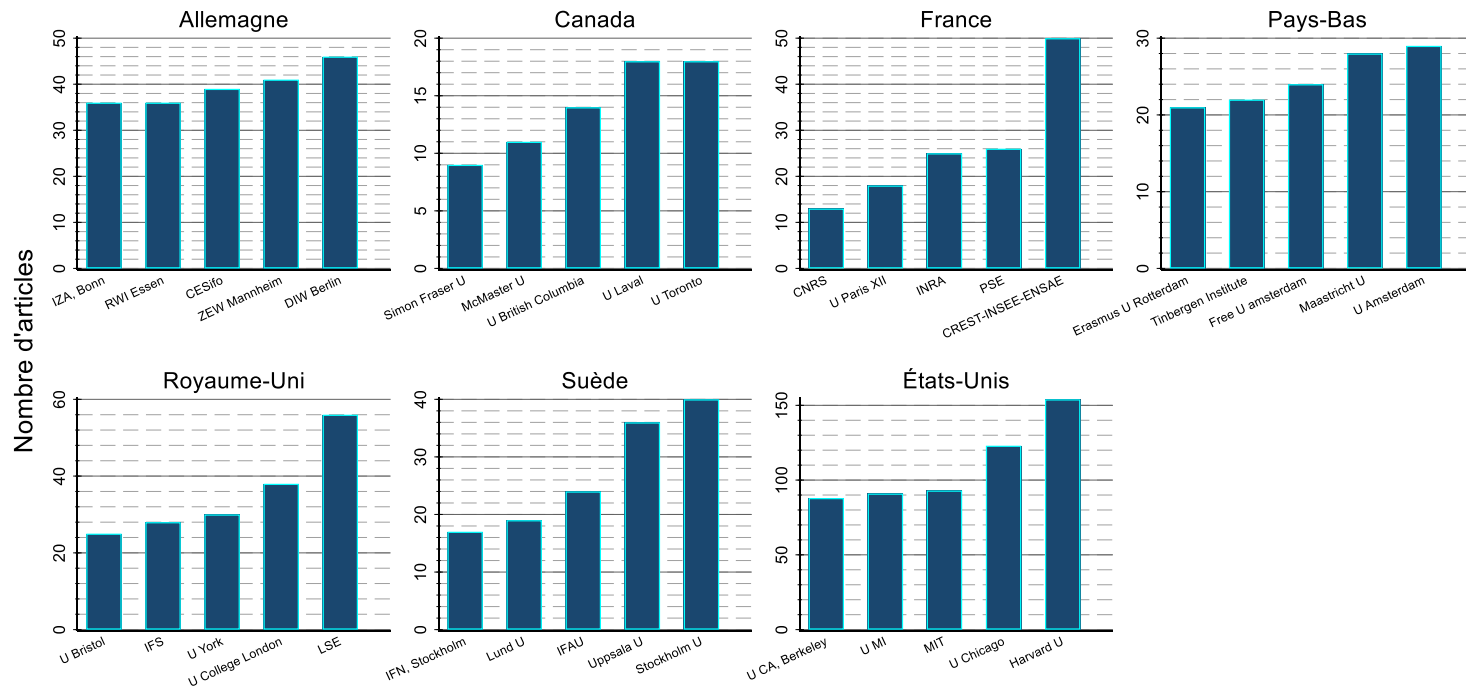
- **Canada, Royaume-Uni, Pays-Bas, États-Unis** orientés vers les thématiques de santé et d'éducation

- **Allemagne, France, Suède** nettement plus préoccupées par les questions relatives au marché du travail

Note de lecture : est calculé pour chaque pays le pourcentage d'occurrence des codes JEL les plus représentés parmi le nombre total de codes JEL cités dans les études portant sur le pays

LES GRANDS ACTEURS NATIONAUX DE L'ÉVALUATION

Top 5 des institutions par pays, métrique simple



RÉSUMÉ DES RÉSULTATS POUR LA FRANCE

La France a un classement relativement défavorable

Si on regarde le **nombre de publications**, on voit que la France comble peu à peu son retard, grâce à une accélération, dans la période récente, de la publication d'articles d'évaluation d'impact
Cependant, si on regarde le **nombre de publications par habitant**, la France reste à la traîne

A L'ÉTRANGER

Le modèle britannique : IFS

- Institute for Fiscal Studies (IFS)
- Privé et indépendant
- Lié au monde universitaire, financement par l'ESRC (équivalent ANR)

Le modèle des Pays-Bas : CPB

- Centraal planbureau (CPB)
- Public, fonctionnaires, financement par le ministère de l'économie
- Indépendance très forte dans les faits

Le modèle américain : CBO

- Congress Budget Office (CBO)
- Public, lié au législatif

HYPOTHÈSES POUR EXPLIQUER LE RETARD EN FRANCE

- Moindre culture de l'évaluation des PP
- Formation insuffisante des élites administratives et politiques à l'évaluation d'impact
- Défaut d'impulsion politique et faible rôle du Parlement dans l'évaluation des politiques publiques en France
- Accès des chercheurs aux données individuelles plus tardif et plus complexe que dans d'autres pays
- Une offre d'évaluateurs plus restreinte, sans doute en lien avec les facteurs précédents

SOURCES

Baïz A, Bono P-H, Debu S, Desplatz R, Lacouette-Fougère C, Trannoy A (2018). Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique. Document de travail de France stratégie.

- Lien : <https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/fs-dt-impact-politiques-publiques-decembre-2018.pdf>

Bono P-H, Debu S, Desplatz R, Hayet M, Lacouette-Fougère C, Trannoy A (2020). La France évalue encore peu l'impact de ses politiques publiques. Note de synthèse de France stratégie.

- Il s'agit de la note de synthèse du document de travail ci-dessus
- Lien : https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/note_de_synthese_politique_publique_web.pdf