

SÉRIE DE DOCUMENTS DE TRAVAIL DU NBER

OÙ LA DÉCOUVERTE SE PRODUIT :
INSTITUTIONS DE RECHERCHE ET CONNAISSANCES FONDAMENTALES EN SCIENCES DE LA VIE

Amitabh Chandra
Connie Xu

Document de travail 33996
<http://www.nber.org/papers/w33996>

BUREAU NATIONAL DE LA RECHERCHE ÉCONOMIQUE
1050 Massachusetts Avenue
Cambridge, MA 02138
Juillet 2025

Nous remercions, sans impliquer quiconque, Alan Garber, George Daley, Kyle Myers, Bitsy Perlman, Bhaven Sampat et Amy Wagers pour leurs précieux commentaires et discussions. Cette recherche a été financée par la Harvard Business School. Nous remercions également les participants au séminaire de la Harvard Medical School en avril 2024, à la réunion du NBER sur l'entrepreneuriat en décembre 2024 et à l'USC en février 2025 pour leurs précieux commentaires. Les opinions exprimées ici sont celles des auteurs et ne reflètent pas nécessairement celles du National Bureau of Economic Research.

Au moins un co-auteur a révélé des relations supplémentaires potentiellement pertinentes pour cette recherche. De plus amples informations sont disponibles en ligne à l'adresse <http://www.nber.org/papers/w33996>

Les documents de travail du NBER sont diffusés à des fins de discussion et de commentaires. Ils n'ont pas été évalués par des pairs ni soumis à l'examen du conseil d'administration du NBER, contrairement aux publications officielles du NBER.

© 2025 par Amitabh Chandra et Connie Xu. Tous droits réservés. De courts extraits de texte, ne dépassant pas deux paragraphes, peuvent être cités sans autorisation explicite, à condition que la source soit citée en entier, y compris la mention ©.

Là où la découverte se produit : les institutions de recherche et les connaissances fondamentales dans les sciences de la vie
Amitabh Chandra et Connie Xu

Document de travail du NBER n° 33996

Juillet 2025

JEL n° H5, I2, O3

ABSTRAIT

Les connaissances fondamentales en sciences de la vie ont des implications considérables pour la médecine et les innovations médicales qui en découlent. En utilisant des publications dans des revues de renom en sciences de la vie pour mesurer les connaissances fondamentales, nous documentons de vastes concentrations au sein des institutions où elles sont découvertes, ainsi qu'une forte corrélation entre les connaissances et les citations ultérieures dans les brevets. Nous évaluons si l'institution où la recherche est produite influence la production des scientifiques en utilisant un modèle de transfert de scientifiques, qui compare la production annuelle de recherche avant et après le transfert d'un même scientifique. Entre 50 et 60 % de la production de recherche d'un scientifique est attribuable à l'institution où il travaille, et les deux tiers de cet effet sont dus à la présence de chercheurs vedettes. L'ampleur de ces effets n'a pas diminué ces dernières années, malgré les technologies facilitant les collaborations interinstitutionnelles, ni augmenté lors des transferts vers de plus grandes agglomérations, ni n'est concentrée dans des domaines scientifiques particuliers. Nous analysons les implications de ces résultats pour les allocations de recherche en sciences et le départ des scientifiques d'une institution vers une autre.

Amitabh Chandra

Université de Harvard

Département d'économie
et le NBER

amitabh_chandra@harvard.edu

Connie Xu

Université de Harvard

conniexu@g.harvard.edu

Les avancées fondamentales dans les sciences de la vie, telles que la découverte de l'ADN, des gènes, de l'ARNm, des anticorps et de CRISPR, ont permis de développer de nouveaux médicaments, d'améliorer les pratiques cliniques et d'approfondir notre compréhension de la biologie et de la physiologie, permettant ainsi d'approfondir encore davantage les connaissances (Zerhouni, 2005). Ces découvertes proviennent principalement des universités et des hôpitaux de recherche, ce qui soulève la question de savoir si certaines institutions sont plus productives que d'autres en matière de découverte de connaissances scientifiques fondamentales. Cette réponse est d'autant plus pertinente que le marché est susceptible de sous-financer les connaissances fondamentales, malgré leur nature fondamentale, et que le financement de ces sciences peut être insuffisant ou capricieux. Comprendre les forces à l'origine de ces découvertes a des implications pour le bien-être, la compréhension de la répartition des ressources entre les institutions et les scientifiques individuels, et la manière dont les scientifiques devraient évaluer leur mobilité vers d'autres institutions.

L'institution où la recherche est menée peut accroître la production scientifique en donnant accès à des ressources telles que des étudiants talentueux, de meilleurs équipements, des mentors expérimentés et des réseaux de collaboration plus riches (Furman et Stern, 2011). De plus, si la proximité avec d'autres chercheurs facilite la diffusion des connaissances et les collaborations, des écosystèmes de recherche plus vastes peuvent présenter des avantages, ce qui implique également une concentration des découvertes (Zucker et al., 1998).

Évaluer l'effet causal d'une institution sur la production de recherche est complexe, car les scientifiques les plus productifs peuvent choisir eux-mêmes leur lieu d'implantation en fonction de préférences personnelles ou professionnelles – comme le lieu de formation, les possibilités de double carrière ou les commodités – ce qui fait que certains lieux apparaissent plus productifs en raison des personnes qui choisissent d'y travailler. Un tel tri signifie que les concentrations de découvertes pourraient refléter des différences de production institutionnelle ou une concentration de scientifiques productifs. Notre article surmonte ces difficultés d'identification en distinguant les effets institutionnels de la sélection de la recherche et contribue à la vaste littérature documentant le rôle central des institutions dans les découvertes scientifiques fondamentales (Nelson, 1959 ; Allison et Long, 1990 ; Zucker et al., 1998 ; Belenzon et Schankerman, 2013 ; Azoulay et al., 2011 ; Hvide et Jones, 2018).

Nous présentons de nouvelles données sur la concentration des découvertes fondamentales en sciences de la vie. Nous proposons une méthode pour définir ces découvertes à l'aide de données de publications transparentes et reproductibles, et nous nous appuyons sur les données de publication d'OpenAlex, un catalogue ouvert et complet de la recherche mondiale (Priem et al., 2022). Pour cet exercice descriptif, notre échantillon comprend plus de 560 000 articles en sciences de la vie (publiés entre 1945 et 2023) dans 15 revues, cités collectivement plus de 65 millions de fois.¹ Nous mesurons la production de recherche comme la fréquence à laquelle un scientifique est cité comme premier ou deux derniers auteurs d'un article, en tenant compte du nombre de citations d'articles et des facteurs d'impact des revues (nous démontrons que ces choix de pondération n'ont pas de conséquences sur nos résultats).

Pour déterminer si l'association entre le lieu de travail des scientifiques et leur production est causale, nous mettons en œuvre un modèle de migration, où nous estimons une spécification d'étude d'événement de la production de recherche d'un individu autour des changements d'emploi, en exploitant la variation de la « taille du changement », définie comme la différence dans la production de recherche moyenne (par scientifique) des institutions d'origine et de destination.² Si les facteurs institutionnels expliquent 100 % de la production d'un scientifique, alors le passage à une

Ces données soulignent une autre raison d'étudier la création de connaissances dans les sciences de la vie : il s'agit d'un secteur où l'on observe l'intégralité du cycle de vie de la R&D – de la recherche universitaire aux brevets, en passant par les essais cliniques et l'approbation de nouveaux médicaments. Ce ne serait pas le cas pour les secteurs qui s'appuient sur des secrets commerciaux plutôt que sur des brevets ou dont les connaissances fondamentales ne sont pas publiées dans des revues scientifiques.

²Nous désignons les articles par scientifique par le terme « production » et non par le terme « productivité » car nous ne mesurons pas d'autres intrants qui affectent la productivité, comme les installations, les ressources, le soutien administratif, les collègues et les étudiants.

une institution avec (disons) un rendement moyen supérieur de 20 % devrait augmenter le rendement d'un scientifique de 20 %. À l'inverse, si la production de recherche est un phénomène entièrement individuel, la relocalisation vers une institution plus ou moins productive n'aura aucun effet sur la production d'un scientifique. Puisque le modèle des déménageurs relie les variations de la production d'un scientifique à celles de son lieu d'implantation, nous pouvons contrôler les caractéristiques individuelles invariantes dans le temps et les tendances temporelles globales, et tester si les scientifiques plus productifs ont tendance à migrer vers des institutions plus productives, ce qui perturberait l'identification des véritables effets de lieu. Pour ce travail, nous suivons la localisation de plus de 180 000 auteurs et leur production de recherche, y compris l'évolution de la production de ceux qui ont déménagé (plus de 19 000 auteurs), avant et après leur déménagement entre institutions (nous validons cette approche en examinant les schémas de co-auteurs des scientifiques et en déterminant si un déménagement prédit un changement de co-auteurs).

Comme nous l'expliquons ci-dessous, le déplacement moyen dans notre échantillon se fait vers un établissement de destination environ 9 % moins productif que l'établissement d'origine, avec de fortes variations dans les deux sens autour de cette moyenne. Empiriquement, nous constatons que l'approche par études d'événements produit des résultats très similaires à ceux d'un modèle bidirectionnel à effets fixes plus simple, qui utilise également les migrants pour estimer les effets institutionnels et individuels.

Notre approche pour le modèle des « movers » est étroitement liée à celle de Lerner et al. (2024), pionniers de l'utilisation d'un modèle de « mover » de « savants itinérants » pour estimer que les facteurs universitaires expliquent un cinquième de la variation entre les universités dans la production de connaissances commercialement pertinentes, mesurée par le degré de citation de la recherche universitaire dans les brevets. Bien que nous partagions leur intérêt pour la recherche en sciences de la vie, notre champ d'action s'étend au-delà des applications commerciales pour examiner la création de connaissances plus largement. Nous reconnaissons que la recherche fondamentale possède une valeur intrinsèque, quel que soit son potentiel commercial. De plus, même lorsque la recherche influence finalement les brevets, la pratique clinique ou les thérapies, ces impacts se manifestent souvent sur plusieurs décennies par des voies non linéaires complexes (Myers et Lanahan, 2022). Si cela est vrai, les estimations de Lerner et al., qui capturent principalement les liens directs et immédiats, pourraient être prudentes, car elles ne mesurent pas l'influence complète des environnements universitaires sur la production scientifique au sens large. Nos principaux résultats sont également similaires à ceux de (Allison et Long, 1990) qui utilisent une approche empirique différente pour étudier un échantillon plus restreint de personnes en mouvement provenant d'un éventail plus large de domaines.

Pour notre analyse, nous devons définir la recherche scientifique fondamentale, la mesurer de manière systématique et l'attribuer à l'institution où elle a été réalisée. Nous adoptons la définition de la recherche fondamentale de la National Academy of Medicine.³ Cette définition inclut la recherche traditionnelle en laboratoire (comme la division cellulaire et la dégradation des protéines) ainsi que la recherche translationnelle précoce (comme les premiers tests de nouvelles thérapies chez l'homme). Cette définition met en évidence les raisons pour lesquelles la recherche fondamentale peut être sous-estimée par rapport à l'optimum social : elle révèle souvent des phénomènes naturels tels que la transcription et les gènes pour lesquels les brevets sont inadmissibles (Kesselheim et al., 2013) ou elle peut être trop éloignée des applications commerciales immédiates pour satisfaire au critère d'« utilité » (également appelé critère de l'article 101) de brevetabilité (Budish et al., 2015). De plus, la divulgation de ces découvertes dans des revues scientifiques crée des externalités positives, car d'autres chercheurs peuvent librement les exploiter. La combinaison d'une appropriabilité limitée et d'externalités positives implique que les entreprises à but lucratif sous-investiront dans la recherche scientifique fondamentale, soulignant l'importance de comprendre les institutions publiques et à but non lucratif qui produisent cette recherche.

3L'Académie nationale de médecine (NAM) considère que « les nouvelles connaissances sur les mécanismes des maladies acquises en laboratoire ont permis de développer de nouvelles méthodes de diagnostic, de thérapie et de prévention, ainsi que leurs premiers tests sur l'homme » (Sung et al., 2003).

(Nelson, 1959 ; Akcigit et al., 2021 ; Azoulay et Li, 2022 ; Myers et Lanahan, 2022).

Ces données révèlent une forte concentration de la recherche fondamentale en sciences de la vie (Carlino et Kerr, 2015). Globalement, les États-Unis dominent le secteur, produisant plus de la moitié de la production mondiale. L'écart de production entre les États-Unis et leur concurrent le plus proche, la Chine, est notamment plus de cinq fois supérieur ces derniers temps. Trois pays (les États-Unis, la Chine et le Royaume-Uni) représentent 70 % de la production mondiale de recherche. Aux États-Unis, Boston et la Baie de San Francisco produisent une part disproportionnée – plus de 15 % – de la recherche fondamentale mondiale en sciences de la vie. Une poignée d'institutions, principalement américaines, représentent une part considérable des découvertes mondiales, Harvard et Stanford représentant plus de 8 % de la production mondiale, ce qui souligne à quel point une poignée d'institutions, principalement américaines, produisent plus de science fondamentale que l'ensemble de nombreux pays. De plus, nous constatons que les villes produisant la recherche la plus fondamentale sont également plus citées dans les brevets par la suite, ce qui indique que cette recherche est non seulement de meilleure qualité scientifique (mesurée par le nombre de citations), mais aussi d'une plus grande pertinence commerciale (mesurée par le nombre de citations d'articles scientifiques vers les brevets), ce qui renforce le rôle de la recherche fondamentale dans l'innovation ultérieure dans les sciences de la vie. Un corollaire de ces faits est que perturber cette répartition ralentira la connaissance à travers le monde (la production d'un scientifique est la contribution d'un autre scientifique), avec des effets néfastes sur les pers

Il est important de reconnaître que les agglomérations géographiques reflètent deux forces : le nombre de chercheurs multiplié par leur production moyenne, et pas seulement cette dernière. Bien que les villes et les institutions comptant de nombreux chercheurs abritent certains des chercheurs les plus productifs, notre analyse de la variation géographique de la production par auteur montre que les scientifiques très productifs sont répartis dans un ensemble diversifié d'institutions, notamment des universités, des laboratoires gouvernementaux et des entreprises de biotechnologie. Ceci incite à comprendre si les institutions affichant une production moyenne élevée augmentent causalement la production des scientifiques qui y rejoignent, ou si un déménagement vers une institution affichant une production par scientifique relativement plus faible réduit la production ultérieure de la personne qui y migre.

Les estimations causales issues du modèle des déménageurs montrent que la localisation a une incidence considérable sur la production scientifique. Environ 50 à 60 % de la variation géographique de la production des chercheurs est imputable à des facteurs institutionnels spécifiques. Autrement dit, le lieu de recherche d'un scientifique a une influence indépendante importante sur sa production, même après prise en compte de facteurs individuels tels que l'expérience, le talent et les relations. L'effet est asymétrique : les chercheurs qui déménagent vers des environnements plus productifs connaissent une augmentation légèrement plus importante de leur production que la baisse observée lorsqu'un chercheur s'installe dans un établissement moins productif.

Nous analysons les implications de ces résultats pour la politique scientifique. Les bailleurs de fonds, comme les National Institutes of Health (NIH) et les philanthropes, doivent choisir entre répartir largement les fonds entre de nombreuses institutions ou concentrer les ressources dans quelques institutions où le scientifique moyen est très productif. Nos résultats suggèrent que traiter toutes les institutions comme des bénéficiaires égaux serait inefficace si l'objectif est de maximiser la production de connaissances (cela peut, bien sûr, répondre à d'autres objectifs de répartition). Même au sein du groupe restreint des 50 institutions comptant les scientifiques les plus productifs, on observe des différences d'un ordre de grandeur quadruple dans la production de recherche, et nos estimations impliquent que plus de la moitié de ces différences sont causales. Autrement dit, si les fonds de recherche contribuent à la découverte, comme le souligne une littérature empirique abondante et solide (Hall et Lerner, 2010 ; Howell, 2017 ; Babina et al., 2023 ; Tham et al., 2024), et que les bailleurs de fonds doivent choisir entre deux scientifiques tout aussi productifs, l'un travaillant dans une institution dont la production de recherche moyenne est deux fois supérieure à celle de l'a

Français les bailleurs de fonds pourraient alors obtenir plus de 50 % de recherche en plus en donnant la priorité à un scientifique de l'institution la plus productive. Dans le même temps, nous constatons que la production de recherche est déterminée par la production d'une institution et non par la taille de son pôle de recherche local, ce qui suggère que les effets d'agglomération sont plus faibles pour la découverte scientifique, même s'ils peuvent être importants pour l'inventivité (par exemple Moretti (2021) ; Azoulay et al. (2012)). Cela se produirait si la découverte scientifique fondamentale dépendait de forces très localisées, telles que les installations de recherche, l'organisation, le leadership institutionnel, la culture collaborative et le mentorat, plus que des retombées de connaissances d'une institution à une autre. Par conséquent, les institutions très productives dans des localités plus petites qui ne sont pas colocalisées avec d'autres grandes institutions (par exemple, Dartmouth à Hanover, NH, ou l'Université du Michigan à Ann Arbor, MI) sont tout aussi efficaces pour produire de la science fondamentale que les institutions tout aussi productives dans des pôles de recherche plus grands comme New York. Ces résultats concordent étroitement avec les recherches antérieures de Gruber et al. (2022), qui documentent les rendements décroissants de l'ajout de chercheurs supplémentaires aux pôles scientifiques denses, et Guzman et al. (2023), qui soulignent que les politiques locales réussissent mieux lorsqu'elles exploitent les forces institutionnelles et les viviers de talents existants.

Ces résultats soulèvent la question de savoir comment une institution peut améliorer sa productivité. Nous constatons que plus des deux tiers de l'effet institutionnel estimé sont dus à la présence de chercheurs vedettes, définis comme des scientifiques se situant dans les 5 % supérieurs de la distribution de la production de leur institution (nous fournissons également des preuves suggérant que la production des co-auteurs d'un chercheur vedette dans l'institution d'origine diminue après leur déménagement). Cette importance des « stars » concorde avec des recherches antérieures selon lesquelles les scientifiques de haut niveau stimulent de manière disproportionnée la création de connaissances, ce qui identifie une opportunité unique de développer l'excellence en recherche. Par exemple, Waldinger (2016) constate que le licenciement de scientifiques vedettes pendant l'Allemagne nazie a entraîné des baisses particulièrement sévères et persistantes de la production de recherche. De même, Azoulay et al. (2019) constatent une baisse durable des taux de publication ajustés en fonction de la qualité pour les collaborateurs de scientifiques vedettes décédés, un résultat que nous reproduisons avec nos données. Enfin, nous observons que nos effets fixes institutionnels estimés sont positivement corrélés à des facteurs tels que les dépenses de R&D d'une institution, sa capacité à attirer des subventions de recherche et l'importance de son instrumentation ; Nous soulignons que ces corrélations ne sont pas causales, mais contribuent à valider les effets institutionnels que nous estimons. La compréhension du rôle des intrants financiers et non financiers demeure un domaine de recherche important pour l'avenir.

I. Données et mesures

A. Mesurer la recherche scientifique fondamentale

La définition de la recherche scientifique fondamentale du NAM englobe deux catégories distinctes : (1) la recherche scientifique fondamentale, qui examine les phénomènes biologiques et chimiques sans objectifs thérapeutiques ou pathologiques spécifiques — couvrant les études de l'anatomie, des organismes et des processus biologiques tels que l'ARN, l'ADN, les gènes, la signalisation cellulaire et la réplication virale — et (2) la recherche scientifique translationnelle, qui applique ces connaissances fondamentales au développement médical et aux applications thérapeutiques.⁴

⁴La définition du NAM est centrée sur les sciences de la vie et exclut la recherche en sciences sociales, comme la recherche sur les services de santé, la recherche en santé publique, la recherche méthodologique, notamment l'économétrie, la biostatistique et l'épidémiologie, ainsi que la recherche économique, comme cet article et la plupart de ses références. De nombreux articles exclus présentent des caractéristiques d'intérêt général similaires à la définition de la recherche fondamentale utilisée par le NAM.

Pour opérationnaliser cette définition de la recherche en sciences fondamentales, nous interrogeons les Medical Subject Headings (MeSH) attribués à chaque publication dans PubMed. Ces classifications hiérarchisées, créées par la Bibliothèque nationale de médecine des NIH, servent de taxonomie standardisée pour l'indexation de la littérature biomédicale. Nous interrogeons spécifiquement les termes MeSH suivants : anatomie, produits chimiques et médicaments, organismes, phénomènes et processus, maladies, anesthésie et analgésie, et thérapeutiques. Les termes MeSH prédominants dans notre échantillon sont les récepteurs, l'ADN, l'ARN, les facteurs de transcription et les neurones, ce qui valide notre approche de mesure systématique de la recherche en sciences fondamentales.

Notre base de données comprend des articles de revues publiés entre 1945 et 2023, à l'exclusion des publications classées comme commentaires, études de cas, notes techniques, lettres et revues. Nous affinons l'échantillon afin de garantir une recherche de haute qualité, de manière simple et transparente, en limitant notre requête aux publications de quinze revues spécialisées en sciences de la vie : Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et FASEB Journal.

Français Les auteurs qui ont publié dans ces revues sont également ceux dont nous étudions les résultats de recherche dans le modèle des moteurs, donc avoir une liste plus large de revues est plus souhaitable car cela génère plus de moteurs qu'une liste plus sélectionnée.⁶ Ce processus de sélection produit un échantillon de base de 563 620 publications, rédigées par 1 433 962 scientifiques. La proportion d'articles de science fondamentale dans ces revues, mesurée selon notre définition, varie de 37 % (Science) à 98 % (Oncogene), les articles restants étant liés aux sciences physiques et sociales (voir le tableau A1 de l'annexe).

Après avoir obtenu l'ensemble pertinent d'articles scientifiques fondamentaux auprès de PubMed, nous collectons des métadonnées détaillées sur leurs auteurs à l'aide d'OpenAlex, un catalogue complet et en libre accès du système de recherche mondial développé par l'association à but non lucratif OurResearch (Priem et al., 2022).⁷ OpenAlex répertorie les entités scientifiques, notamment les travaux universitaires, les auteurs, les institutions, les revues et les concepts de recherche, tous interconnectés pour faciliter l'analyse. Ces données proviennent de l'agrégation et de la normalisation d'informations provenant de plus de 249 000 sources, telles que des bases de données de publications comme PubMed et Microsoft Academic Graph (MAG), ainsi que des organisateurs de bases de données de métadonnées comme Crossref.⁸ Selon la OpenAlex, la publication de données, OpenAlex indexe plus de 250 millions d'œuvres de 95 millions d'auteurs affiliés à 109 000 institutions différentes. Ces données sont mises à jour fréquemment, souvent toutes les quelques heures, via son API.

OpenAlex offre plusieurs avantages par rapport à des bases de données comparables comme PubMed ou Web of Science pour identifier la géographie de la recherche. Il s'efforce délibérément de lever l'ambiguïté entre les auteurs et de normaliser les affiliations institutionnelles. En utilisant les informations sur les auteurs de MAG, Crossref,

⁶ Il s'agit de la première année au cours de laquelle les termes MeSH sont attribués de manière fiable aux articles dans PubMed.

Nous avons établi cette liste en demandant à cinq scientifiques de renom de la faculté de médecine de Harvard, dont le doyen, de lister les dix principales sciences fondamentales des sciences de la vie. L'inclusion de neuf revues, dont Cell, Nature et Science, a fait l'unanimité, tandis que l'inclusion de PLOS One et d'Oncogene, approuvées par un scientifique, a suscité la plupart des désaccords. Les articles de revues cliniques telles que le Journal of the American Medical Association (JAMA) ou le New England Journal of Medicine sont exclus de nos analyses, car la définition de la science fondamentale selon la NAM exclut les études cliniques telles que les essais de phase III avancés.

Nous remercions Bhaven Sampat pour avoir suggéré cette ressource ; dans les versions précédentes, nous utilisions PubMed. Toute erreur de mise en œuvre est de notre responsabilité.

⁸ Crossref est une organisation à but non lucratif qui fournit des identifiants d'objets numériques (DOI) pour indexer le contenu scientifique qui permet de relier les métadonnées, tandis qu'ORCID (Open Researcher and Contributor ID) est un système d'identification alphanumérique unique qui aide les chercheurs à gérer leurs informations professionnelles et leurs dossiers de publication.

Pour les sites web de PubMed, ORCID et des éditeurs, OpenAlex utilise un algorithme basé sur le nom de l'auteur, son historique de publication, ses modèles de citation et (le cas échéant) son ORCID pour identifier toutes les publications appartenant à un auteur spécifique. Par exemple, J. Smith et John Smith, écrivant sur la biologie structurale, seraient considérés comme le même auteur, tandis que John J. Smith, écrivant sur les mathématiques, serait considéré comme un auteur différent. OpenAlex analyse également les chaînes d'affiliation brutes des métadonnées et utilise un algorithme d'apprentissage automatique pour extraire les informations géographiques et institutionnelles. Cela permet un traitement cohérent des affiliations telles que « MIT, Boston, États-Unis » et « Massachusetts Institute of Technology ». Ces fonctionnalités d'OpenAlex fournissent un ensemble plus complet d'affiliations d'auteurs sur la période analysée que d'autres sources.

Enfin, nous utilisons OpenAlex pour déterminer l'âge des chercheurs en sciences fondamentales afin de comprendre l'impact des institutions sur les résultats du cycle de vie. Pour ce faire, nous identifions la première apparition d'un chercheur dans un article indexé par OpenAlex. Nous attribuons à chaque chercheur l'âge de 25 ans l'année de sa première publication.

B. Mesurer la production de recherche des scientifiques

Nous passons des publications aux résultats de recherche individuels des scientifiques en quantifiant une mesure de contribution à la recherche « efficace » pour chaque auteur d'un article donné. Pour ce faire, nous attribuons d'abord le crédit en divisant chaque article équitablement entre le nombre d'auteurs. Nous ajustons ensuite l'impact de la recherche à l'aide de pondérations spécifiques à chaque article, qui tiennent compte à la fois de la qualité de la revue et de l'influence des citations. Plus précisément, nous attribuons des pondérations en fonction du facteur d'impact quinquennal le plus récent de la revue, basé sur Clarivate (2023), et du nombre moyen de citations annuelles depuis la publication. Ces ajustements de pondération garantissent que les articles publiés dans des revues à fort impact reçoivent une pondération plus élevée, tout comme ceux qui accumulent davantage de citations au fil du temps. Il est important de noter que le système de pondération est conçu de manière à ce que la somme totale des contributions à la recherche de tous les auteurs soit toujours égale au nombre total d'articles dans notre ensemble de données. Les détails mathématiques de ce système de pondération sont fournis dans la section B de l'annexe. De plus, la partie supérieure du tableau A2 de l'annexe compare différentes méthodes de pondération, démontrant que toutes les approches produisent des résultats fortement corrélés aux niveaux du pays, de la ville et de l'établissement.

Pour notre analyse principale, nous nous concentrons sur les chercheurs figurant parmi les deux premiers ou les deux derniers auteurs d'une publication. Dans le domaine des sciences de la vie, les auteurs sont généralement classés par ordre de contribution majeure au travail et à la rédaction, tandis que ceux figurant à la fin représentent généralement le ou les chercheurs principaux. Cette restriction permet de contrôler les différences de normes en matière de co-auteurs selon les lieux, limitant ainsi le nombre d'auteurs crédités à quatre. Dans la section « robustesse », nous réestimons notre spécification principale à l'aide de critères d'inclusion alternatifs, en considérant à la fois un échantillon restreint de premier et dernier auteurs uniquement et un échantillon plus large incluant tous les auteurs. En fait, notre concentration sur les deux premiers et les deux derniers auteurs fournit probablement une estimation prudente des effets liés au lieu. En étendant l'analyse à tous les auteurs, nous constatons que les influences liées au lieu augmentent d'environ 10 %. Cela suggère que les auteurs débutants et finaux ont tendance à

Nous validons notre hypothèse de 25 ans à la première publication par quelques vérifications. En parcourant OpenAlex pour trouver les articles répertoriés comme des thèses, nous constatons une relation très linéaire entre notre définition de la date de naissance de la publication et l'année d'obtention du doctorat du chercheur. De plus, le profil d'âge de sortie basé sur notre définition correspond à celui observé dans des recherches antérieures (Jones, 2009).

être des chercheurs plus établis et sont relativement moins affectés par les facteurs institutionnels locaux, tandis que les auteurs intermédiaires peuvent être plus affectés par les influences locales.

C. Mesurer la production institutionnelle

Identifier l'institution d'origine où les scientifiques mènent physiquement leurs recherches est difficile en raison de la prévalence de multiples affiliations simultanées. Pour résoudre ce problème, nous déduisons l'institution principale de chaque auteur en analysant l'historique complet de ses publications et en identifiant l'affiliation la plus cohérente pour une année donnée.¹⁰ Si une affiliation principale unique existe pour cette année, nous l'attribuons. En cas d'égalité ou de tendances alternées, nous supposons une continuité institutionnelle. Par exemple, si un auteur est affilié à l'institution A en 2008 et 2010, mais pas en 2009, nous désignons l'institution A comme affiliation principale pour toutes les années. Si un auteur mentionne deux affiliations (A et B) pour la même année, nous choisissons A si A était l'affiliation principale l'année précédente ; sinon, nous sélectionnons celle qui apparaît comme affiliation principale l'année suivante. Enfin, si l'égalité persiste, nous attribuons aléatoirement l'une des institutions à l'auteur, ce qui affecte environ 5 % des années-auteurs.

Après avoir attribué chaque auteur à une institution unique pour chaque année, nous agrégeons les résultats à des niveaux géographiques plus larges (par exemple, villes et pays). Pour les régions des États-Unis, nous utilisons les zones statistiques métropolitaines (MSA) et, le cas échéant, nous regroupons les principales MSA en marchés du travail intégrés. Par exemple, nous définissons la ville de la baie de San Francisco, en Californie, comme la combinaison de San Francisco- Oakland-Hayward et de San Jose-Sunnyvale-Santa Clara. De même, nous définissons Research Triangle Park, en Caroline du Nord, comme la combinaison de Durham-Chapel Hill et de Raleigh-Cary. Dans la suite de l'article, nous désignons ces classifications par le terme « villes ».

Enfin, pour tenir compte des différences de tendances de production entre les sous-domaines des sciences de la vie, nous classons chaque auteur-année dans l'un des quinze domaines de recherche en sciences de la vie pour chaque période de 20 ans de notre échantillon (par exemple, la signalisation cellulaire et les cellules souches). Cette approche nous permet de contrôler les scientifiques sélectionnant des domaines qui progressent à des rythmes différents et qui peuvent affecter la probabilité d'acceptation dans une revue scientifique (par exemple, la génétique par rapport à la biologie structurale ou cancérologique après la cartographie du génome humain). Pour définir ces domaines, nous utilisons un algorithme de clustering textuel k-means utilisant les titres des articles, les termes MeSH et les résumés des publications de chaque auteur comme données d'entrée. Nous appliquons cet algorithme de clustering au niveau auteur-année séparément pour chaque période de 20 ans, permettant ainsi aux définitions du domaine d'évoluer au fil du temps. La section C de l'annexe détaille les caractéristiques tokenisées qui définissent chaque domaine de recherche. Nous intégrons ces effets de sous-domaine comme spécification de régression alternative dans nos exercices de robustesse et démontrons que nos principaux résultats ne changent pas.

Pour illustrer notre approche, considérons les deux articles suivants qui ont contribué au développement de la technique d'édition du génome CRISPR-Cas9, une découverte révolutionnaire qui a valu à Emmanuelle Charpentier et Jennifer A. Doudna le prix Nobel de chimie 2020. Le premier article, « Une endonucléase d'ADN programmable guidée par double ARN dans l'immunité bactérienne adaptative »,

¹⁰Nous omettons les affiliations identifiant un bailleur de fonds (par exemple, le Wellcome Trust, le Howard Hughes Medical Institute (HHMI) ou le Broad Institute). Nous comptabilisons le HHMI uniquement lorsque l'entrée mentionne le Janelia Research Campus, et le Broad Institute uniquement lorsqu'il s'agit de l'employeur principal du scientifique plutôt que d'un membre du corps enseignant affilié. Les affiliations hospitalières dont le nom inclut une université (par exemple, NYU Langone, Stanford Hospital) sont enregistrées sous l'université, tandis que les hôpitaux affiliés à un établissement universitaire mais portant une marque distincte (par exemple, Mass General Brigham) restent distincts.

L'article a été publié dans la revue Science en 2012 et a été cité 13 832 fois. Pour construire notre mesure de la production de recherche, nous attribuons d'abord une contribution effective de 0,25 à Martin Jinek (premier auteur) de l'UC Berkeley (dans la baie de San Francisco), Krzysztof Chylinski (deuxième auteur) de l'Université d'Umeå, Jennifer A. Doudna (avant-dernier auteur) à l'UC Berkeley et Emmanuelle Charpentier (dernier auteur) à l'Université d'Umeå. En 2012, les quatre auteurs ont été classés comme travaillant dans le sous-domaine de l'ADN et de la stabilité du génome.¹¹ En utilisant le dernier facteur d'impact sur 5 ans de Science (54,5) et le nombre moyen de citations par an depuis la publication ($13832/12 \approx 1152$), nous repondérons notre mesure de production pour représenter l'influence relative de cet article par rapport aux autres travaux de notre échantillon. Par conséquent, ces quatre auteurs (et leurs institutions affiliées) reçoivent une contribution recalibrée de 28,62 articles pour cette publication.

À titre d'exemple, l'article « Interrogation de l'ADN par l'endonucléase Cas9 guidée par l'ARN CRISPR » a été publié dans Nature en 2014 et a été cité 1 647 fois. Nous nous concentrons sur Samuel Sternberg (premier auteur) de l'UC Berkeley, Sy Redding (second auteur) de l'Université Columbia, Eric C. François Greene (avant-dernier auteur) à l'Université Columbia et Jennifer A. Doudna (dernier auteur) à l'UC Berkeley. En 2014, les quatre auteurs ont été classés comme travaillant dans le sous-domaine de l'ADN et de la stabilité du génome.¹² Après ajustement pour le facteur d'impact de Nature (60,9) et le nombre moyen de citations par an depuis la publication ($1647/12 \approx 137$), Samuel Sternberg et Jennifer A. Doudna (et leurs institutions affiliées) se voient attribuer une contribution effective de 3,30 articles pour cette publication. Pour mettre ces ordres de grandeur de la production de recherche en contexte, le tableau 3 montre que l'auteur moyen dans un ensemble d'articles scientifiques d'élite a une production annuelle de 0,39 article et une production à vie de 1,48 article (mesurée jusqu'à la fin de notre période en 2023).

II. Découverte scientifique fondamentale

A. Variation géographique des résultats de la recherche

La figure 1 montre la part de la production de recherche produite par chaque pays depuis 1945. Les données révèlent que les États-Unis ont constamment dominé la production scientifique fondamentale, représentant plus de la moitié de la production mondiale et dépassant les autres pays par une marge substantielle.¹³ Comme l'illustre la figure 1, le déclin de la part de production des États-Unis, de 70 % au début des années 1990 à 50 % aujourd'hui, est en partie la conséquence de la présence croissante de la Chine dans la production scientifique fondamentale. Depuis le début des années 2000, lorsque la Chine est entrée pour la première fois dans l'ensemble des grands producteurs scientifiques, la Chine a fait

Selon notre classification par sous-domaine, Martin Jinek a débuté ses recherches en biologie structurale et se concentre actuellement sur la génomique. Krzysztof Chylinski a débuté ses recherches en ADN et stabilité du génome et travaille désormais en écologie et biologie environnementale. Jennifer Doudna a débuté ses recherches en métabolisme cellulaire et bioénergétique et se concentre désormais sur la biologie cellulaire. Emmanuelle Charpentier a débuté sa carrière en neurosciences et neurobiologie avant d'élargir ses recherches à la biologie moléculaire et cellulaire, à la génétique et à la génomique.

Tout au long de leur carrière, Samuel Sternberg, Sy Redding et Eric Greene restent principalement dans le domaine de l'ADN et du génome stabilité.

¹³Le pic important autour des années 1970 dans la part de la production de recherche des pays « restants » est la publication de Cleavage of structural proteins during the assembly the head of bacteriophage T4 dans Nature dont l'unique auteur était le scientifique suisse Ulrich K. Laemmli en 1970. L'article historique de Laemmli, avec plus de 300 000 citations, documente sa découverte de l'électrophorèse sur gel de polyacrylamide SDS, l'une des techniques les plus largement utilisées et les plus importantes en biologie moderne.

Des progrès remarquables ont été réalisés dans la production de recherche scientifique fondamentale, dépassant le Royaume-Uni comme deuxième producteur mondial ces dernières années. Il convient également de noter le recul du rôle du Royaume-Uni depuis l'après-Seconde Guerre mondiale, où il produisait environ 25 % de la recherche scientifique mondiale, jusqu'à la période la plus récente, où sa part est inférieure à 10 %.

Aux États-Unis, la production globale de recherche est fortement concentrée dans quelques zones métropolitaines et institutions clés (en clair, la production globale reflète le nombre de scientifiques et leur production individuelle). Boston et la Baie de San Francisco, en particulier, représentent une part disproportionnée de la production totale, et au sein de ces régions, Harvard et Stanford se distinguent comme des contributeurs majeurs à la science mondiale. Les tableaux A4 et A5 répertorient les 50 premières institutions selon la production annuelle moyenne pour les institutions basées aux États-Unis et les institutions internationales, respectivement. Les figures 2 et 3 illustrent la persistance de ces lieux à produire une part importante de la production de recherche fondamentale. Ces tendances soulignent qu'au niveau agrégé, un nombre relativement restreint de villes et d'institutions sont responsables d'une grande partie de la production totale de recherche.

Français Cependant, lorsque l'on passe de la production globale à la production individuelle des chercheurs, nous observons une tendance plus distribuée. Le tableau 1 présente les 20 premières villes en termes de production moyenne des chercheurs entre 2015 et 2023. Notre mesure de la production de recherche est interprétée comme le nombre « effectif » de publications pour un auteur. Ainsi, le chercheur moyen de la région métropolitaine de Boston publie effectivement deux fois plus que celui de la région métropolitaine d'Austin.¹⁴ Le tableau 2 met également en évidence les meilleures institutions dans les dix villes les plus productives. Ces résultats montrent que le chercheur moyen du Dana-Farber Cancer Institute a effectivement produit quatre fois plus de publications par an que le chercheur moyen du Lawrence Berkeley National Laboratory. Alors que des institutions d'élite comme Harvard et Stanford dominent la production scientifique mondiale en raison de leur taille, nous observons que de nombreuses institutions plus petites abritent des chercheurs très productifs, notamment des universités, des organismes de recherche, des agences gouvernementales et des entreprises de biotechnologie.

Les tableaux A6 et A7 de l'annexe présentent la liste des 50 premières institutions selon la production moyenne de leurs chercheurs, respectivement aux États-Unis et à l'étranger. Ces tableaux renforcent la distinction entre la production de recherche globale, dominée par un petit nombre d'institutions, et la production au niveau des chercheurs, répartie sur un ensemble diversifié d'institutions. Plusieurs institutions internationales, telles que le Wellcome Sanger Institute, plusieurs hôpitaux et universités néerlandais et l'Institut Weizmann des sciences en Israël, affichent une productivité par habitant comparable à celle des principales institutions américaines. Ces institutions sont relativement plus petites que les institutions américaines, ce qui explique leur faible part de la production mondiale, mais leurs scientifiques sont en moyenne aussi productifs que ceux de Dana-Farber, de l'Allen Institute, du MIT, de Sloan Kettering et de Stanford.

B. Science fondamentale et pertinence commerciale

Jusqu'à présent, nous n'avons pas établi si la recherche scientifique fondamentale a des implications directes pour l'activité commerciale, ou si elle constitue une connaissance fondamentale, mais non corrélée à des inventions. En nous appuyant sur les travaux de (Bryan et al., 2020), (Marx et Fuegi, 2020), nous mesurons la propension de la recherche scientifique fondamentale à être citée dans les brevets. Cette mesure des « articles »

¹⁴Lors du calcul de la production moyenne par année-auteur dans une zone géographique donnée, nous excluons les établissements comptant moins de 100 auteurs. Pour les établissements américains, nous excluons également les établissements comptant moins de 10 auteurs. Ces ajustements permettent d'éliminer les établissements potentiellement mal codés ou insuffisamment représentés.

La production « dans les brevets » est similaire à notre mesure de la production de recherche, sauf que les pondérations spécifiques à l'article s'ajustent en fonction du nombre de fois où l'article a été cité dans le corps du texte des brevets.

Les brevets citent les articles de recherche sur la première page de la demande (« citations de première page ») et dans le corps du texte (« citations de corps »). Nous nous concentrons sur les citations de corps, car des recherches antérieures ont montré qu'elles sont plus susceptibles de refléter les connaissances d'un inventeur et de mieux représenter la diffusion des connaissances, car elles sont incluses par les inventeurs eux-mêmes (à l'inverse, les citations de première page sont généralement ajoutées par les avocats spécialisés en brevets (Bryan et al., 2020)). Comme précédemment, nous recalculons ces pondérations pour que la somme soit égale à 1 afin de garantir que la somme de notre mesure « articles dans les brevets » soit égale au nombre total d'articles de notre échantillon.

Nous traçons la relation, à l'échelle de la ville, entre la production de recherche scientifique fondamentale et la production d'articles brevetés dans la figure 4. La pente est de 1, ce qui nous permet de déduire que les villes qui produisent davantage de recherche scientifique fondamentale sont également plus susceptibles de produire des articles de recherche cités dans des brevets. Cela souligne les implications commerciales et sociales de la recherche scientifique fondamentale. Cependant, notre étude ne s'appuie pas sur ce fait, car la recherche scientifique fondamentale est suffisamment fondamentale pour ne présenter aucune pertinence commerciale.¹⁵

III. Résultats de recherche des institutions et des individus

A. Stratégie empirique

Nous présentons un cadre empirique utilisant un modèle « movers » pour décomposer la variation observée dans la production logarithmique de la recherche fondamentale en composantes locales et individuelles. Cette approche suit de près les travaux antérieurs utilisant la même stratégie d'identification (Finkelstein et al., 2016 ; Chetty et Hendren, 2018 ; Lerner et al., 2024 ; Keys et al., 2023). Les composantes locales capturent les facteurs institutionnels locaux tels que les dépenses de R&D au niveau de l'entreprise, tandis que les composantes individuelles capturent les caractéristiques individuelles telles que le niveau d'éducation et la motivation intrinsèque.

Nous modélisons la sortie logarithmique d'un scientifique individuel, y_{it} , dans l'établissement en année avec une méthode bidirectionnelle spécification des effets fixes :

$$y_{it} = \alpha_i + \gamma_t + \beta_1 x_{it} + \epsilon_{it} \quad (1)$$

où α_i est un effet fixe individuel, un effet fixe institutionnel et un effet fixe annuel. Ce modèle peut être étendu pour prendre en compte les caractéristiques observables des chercheurs en incluant les cas où les moyennes au niveau des chercheurs seraient supérieures à α_i plutôt que α_i . Pour des raisons de commodité d'exposition, nous décrivons la théorie motivante avec le modèle simplifié, mais dans certaines spécifications de régression, nous intégrons des effets fixes de sous-domaines qui seraient compris sous α_i .

Un défi majeur en matière d'identification réside dans le fait que les effets fixes pour l'institution, γ_t , ne sont identifiés que si l'échantillon inclut les scientifiques qui changent d'institution. Sans ces changements, il est impossible de distinguer les effets de lieu de l'hétérogénéité individuelle. Une interprétation causale de ces effets de lieu nécessite plusieurs hypothèses. Premièrement, les variations non observées du potentiel de production logarithmique d'un scientifique...

¹⁵Nous avons également examiné la sensibilité de nos résultats à l'utilisation de citations en première page par rapport aux citations dans les organismes de brevets – quelle que soit la mesure choisie, nous voyons dans l'annexe A3 qu'il existe une forte corrélation entre la production de recherches scientifiques fondamentales et leur citation dans les brevets au niveau du pays, de la ville et de l'institution.

tial ne doit pas être corrélé avec la différence de sortie logarithmique entre leur origine et leur destination institutions. Cette hypothèse serait violée si les « étoiles montantes » se relocalisaient systématiquement des institutions à faible rendement vers des institutions à haut rendement. De telles violations apparaîtront dans les pré-tendances de notre événement. l'étude, il est donc essentiel de les examiner. Deuxièmement, nous supposons une séparabilité additive de et en logarithme résultat, ce qui implique que les caractéristiques des personnes et des lieux sont multiplicatives à plusieurs niveaux. Autrement dit, un passage vers un établissement à plus haut rendement a un impact absolu plus important sur les chercheurs à plus faible rendement. Enfin, nous supposons que les effets de lieu estimés () pour les personnes ayant déménagé se généralisent aux personnes n'ayant pas déménagé. Si cette hypothèse est violée, alors nos estimations ne s'appliqueraient qu'à l'ensemble des chercheurs qui relocaliser plutôt que l'ensemble de la population de chercheurs.

Si ces hypothèses sont satisfaites, il est alors possible de décomposer la variation en logarithme moyen la production des chercheurs à travers les institutions en composantes basées sur le lieu et la personne, notons moyenne de tous les chercheurs travaillant dans l'établissement au cours de l'année, $\bar{y}_{it}$ et comme moyenne $\bar{y}_{it}$ de au fil du temps. De même, $\bar{y}_{it}$ et $\bar{y}_{it}$ représentent les moyennes au niveau des chercheurs. La différence de sortie logarithmique moyenne entre deux institutions et est donné par :

$$\bar{y}_{it} - \bar{y}_{it} = (\bar{y}_{it} - \bar{y}_{it}) + (\bar{y}_{it} - \bar{y}_{it})$$

Cela nous permet de déclarer les parts attribuables à l'institution et aux chercheurs comme suit :

$$(\bar{y}_{it} - \bar{y}_{it}) = \frac{\bar{y}_{it} - \bar{y}_{it}}{\bar{y}_{it} - \bar{y}_{it}},$$

$$(\bar{y}_{it} - \bar{y}_{it}) = \frac{\bar{y}_{it} - \bar{y}_{it}}{\bar{y}_{it} - \bar{y}_{it}}.$$

Nous rapportons ces résultats et partageons ces résultats dans les deux panneaux du tableau 4.

B. Estimation de l'étude d'événement

Nous mettons en œuvre une approche d'étude d'événements pour observer visuellement l'évolution dynamique des effets de lieu sur la sortie du journal scientifique et évaluer la robustesse de nos hypothèses à travers les pré-tendances.

Nous mettons d'abord à l'échelle la sortie du journal de chaque moteur pour tenir compte des différents niveaux de sortie d'origine et institutions de destination. Ceci est crucial, car tous les déménageurs n'ont pas la même origine et la même destination. $\bar{y}_{it}$ à emplacement et, sans mise à l'échelle, se déplace de à Plus $\bar{y}_{it}$ serait annulé par les mouvements de $\bar{y}_{it}$ à . précisément, nous définissons la « taille de déplacement » du moteur comme la différence de sortie logarithmique moyenne entre institution () (destination) et institution () (origine), à l'exclusion du déménageur lui-même :

$$= \bar{y}_{it} - \bar{y}_{it} - (\bar{y}_{it} - \bar{y}_{it})$$

Ici, il est construit en faisant la moyenne de toutes les années-auteurs (y compris les non-déménageurs mais à l'exclusion le déménageur lui-même) à l'établissement de destination et est calculé de la même manière pour l'institution d'origine. Puisque nous nous limitons aux chercheurs ayant un seul déplacement, chaque déplacement est associé avec une valeur unique de $\bar{y}_{it}$. De plus, les moyennes institutionnelles sont calculées sur toutes les années-auteurs,

ce qui signifie que les chercheurs ayant le même appariement institution de destination-origine auront le même ¹⁶.

Suivant (Bronnenberg et al., 2012), nous mettons à l'échelle la sortie du journal pour le moteur comme

$$= \frac{\bar{y}_{it} - \bar{y}_{it}^0}{\bar{y}_{it}^0}.$$

Selon cette échelle, $\bar{y}_{it} = 0$ lorsque la sortie logarithmique du chercheur est égale à celle de son institution d'origine sortie logarithmique moyenne et $\bar{y}_{it} = 1$ lorsqu'elle est égale à la sortie logarithmique moyenne de leur institution de destination.

Ainsi, dans une étude d'événement simple, la régression

$$\bar{y}_{it} = \alpha + \beta_t + \epsilon_{it},$$

les coefficients représenteraient la fraction moyenne de la taille du mouvement clôturé par rapport année (t, i) . Ici, nous définissons l'année relative (t, i) comme $t - i$ si le déménageur se déplace dans l'année i . Ainsi,

$(t, i) = -1$ représente la dernière année dans l'établissement d'origine et $(t, i) = 0$ représente la première année à l'emplacement de destination.

En pratique, comme il est souvent proche de 0 comme le montre la figure A1, nous n'estimons pas le simple. Nous avons donc réorganisé notre équation pour éviter de diviser par :

$$\bar{y}_{it} = \alpha + \beta_t + \epsilon_{it}$$

Lorsque nous appliquons cette équation à nos données, nous incluons un ensemble de contrôles à effets fixes et l'échantillon analogue de $\bar{y}_{it}$ pour obtenir notre spécification d'étude d'événement de base de :

$$\bar{y}_{it} = \alpha + \beta_t + \gamma_i + \epsilon_{it} \quad (2)$$

Où sont les effets fixes champ-par-année et (t, i) les effets fixes relatifs par année. Nous combinons α et γ_i en un seul effet fixe individuel, $\bar{y}_{it}$ la tendance de production indépendante de l'ampleur du mouvement. Cette spécification permet également une (capturée par (t, i)) pour tous les acteurs mobiles, quelle que soit leur taille. et la direction du mouvement. Ces effets reflètent les tendances de la production de recherche qui sont communes à tous déménageurs – par exemple, une baisse de la production dans les années suivant un déménagement, ou une explosion de la production avant un.

Le paramètre clé d'intérêt dans l'équation 2 sont les coefficients relatifs des années qui (t, i) , nous normalisons à zéro l'année précédant le déménagement $(t, i) = -1$. Ces coefficients mesurent le variation en pourcentage de la production logarithmique au cours des années entourant le mouvement, mise à l'échelle par la taille et la direction du déplacement. Nous nous attendons à ce que la sortie logarithmique de la destination ne soit pas prédictive de la sortie logarithmique avant un déplacement, mais pour prédire la sortie logarithmique après un déplacement. Ainsi, (t, i) devrait être statistiquement indiscernable de zéro pour $(t, i) < 0$ (pas de pré-tendances) et non nul après un mouvement.

¹⁶ Notez qu'il s'agit d'une mesure invariante dans le temps, calculée comme la différence du résultat moyen sur toutes les périodes entre les lieux de destination et d'origine. Si la moyenne du lieu est $\bar{y}_{it}^0$ et $\bar{y}_{it}^1$, ont été calculés pour chaque fois période pour créer un variable dans le temps, cela introduirait probablement une endogénéité dans le modèle. Une telle variable dans le temps La mesure pourrait saisir les chocs ou les tendances contemporains dans les lieux d'origine et de destination qui sont corrélés avec à la fois le moment du mouvement et le résultat de l'individu, plutôt que de refléter uniquement la différence stable dans effets de lieu.

L'ampleur du saut de (β, γ) après un déplacement reflète l'influence moyenne des facteurs spécifiques au lieu sur la sortie logarithmique. En d'autres termes, $(\beta, \gamma) = 1$ indique que la sortie logarithmique du mobile a convergé complètement vers la moyenne à destination et impliquerait que la sortie logarithmique est entièrement influencée par les facteurs spécifiques au lieu. En revanche, $(\beta, \gamma) = 0$ indique l'absence de convergence de la sortie logarithmique du mobile vers celle de destination et impliquerait l'absence de rôle des facteurs spécifiques au lieu.

Nous rapportons des erreurs standard robustes regroupées au niveau de l'institution.¹⁷

IV. Estimation des effets de l'institution et du chercheur

A. Caractéristiques des déménageurs scientifiques

En utilisant l'échantillon d'auteurs dont les recherches ont été décrites ci-dessus, nous nous concentrons maintenant sur l'échantillon d'auteurs ayant changé d'institution (c'est la principale raison pour laquelle nous avons pris en compte un large éventail de revues, car cela élargit le bassin de chercheurs ayant changé d'institution). Nous définissons un scientifique comme ayant changé d'institution s'il est affilié à différentes institutions pendant des années consécutives. Ce panel géographique étant construit à partir de l'historique des publications, il est à craindre que les décalages de publication n'affectent la précision de l'identification des dates de changement d'auteur. À titre de validation, nous traçons la part des coauteurs en science fondamentale d'un chercheur ayant changé d'institution, issus de ses institutions d'origine et de destination, au fil du temps. En théorie, le nombre de coauteurs issus de l'institution de destination ne devrait pas augmenter avant le changement d'institution. En réalité, nous constatons que les chercheurs ayant changé d'institution commencent à accumuler des coauteurs issus de leur institution de destination environ trois ans avant leur première publication officiellement affiliée à la nouvelle institution. Par conséquent, nous définissons l'année de changement d'institution comme étant trois ans avant la première publication observée dans l'institution de destination. La figure 5 illustre l'effet de cet ajustement, montrant que la part des coauteurs issus de l'institution de destination n'augmente qu'après notre nouvelle année de changement d'institution.

Le tableau 3 présente les statistiques récapitulatives concernant les chercheurs ayant déménagé et ceux n'ayant pas déménagé dans notre échantillon. Parmi les chercheurs basés aux États-Unis, nous avons pu identifier 37 809 scientifiques ayant déménagé et, compte tenu du rôle important des États-Unis dans la recherche scientifique fondamentale, nous ne considérons pas notre restriction aux scientifiques basés aux États-Unis comme une limitation majeure (dans notre échantillon initial de 563 620 articles, 294 782 étaient rédigés par 331 846 chercheurs basés aux États-Unis). En moyenne, ces chercheurs ayant déménagé ont déménagé 1,4 fois au cours de notre période. En raison du bruit potentiel dans la précision de l'algorithme de désambiguïsation des auteurs d'OpenAlex et de notre méthode d'imputation des affiliations, nous nous concentrons sur un ensemble de 25 845 chercheurs n'ayant déménagé qu'une seule fois. En moyenne, les chercheurs ayant déménagé ont tendance à venir de villes légèrement plus petites, mais produisent en moyenne plus de résultats que les chercheurs n'ayant pas déménagé. Les chercheurs ayant déménagé pour la première fois représentent environ 60 % de l'échantillon ayant déménagé et sont représentatifs de la population plus large qui se déplace.

¹⁷Notez que s'il y a une erreur de mesure classique dans β , Comme c'est probablement le cas pour les petites institutions, cela biaiserait nos estimations vers zéro. À titre de test de robustesse, nous estimons notre spécification principale sur un ensemble de « grandes institutions », définies comme celles comptant au moins 100 auteurs et 10 contributeurs, et constatons que les effets de lieu estimés augmentent de 60 à 70 %.

B. Estimations à partir d'effets fixes bidirectionnels

Le modèle à effets fixes à deux facteurs en 1 fournit l'ensemble d'estimations le plus simple et constitue un moyen simple pour effectuer un exercice de décomposition où nous estimons les effets spécifiques à l'institution (échantillon de personnes ayant) en utilisant notre démenagé et de personnes n'ayant pas démenagé). Ces estimations sont cohérentes tant que l'identification

les hypothèses mentionnées ci-dessus sont respectées.

Le panneau A du tableau 4 examine la proportion de variance institutionnelle transversale dans la recherche résultats qui peuvent être attribués aux effets du chercheur par rapport aux effets institutionnels, et nous quantifions ces deux composantes telles que (1) la part de la variance interinstitutionnelle qui serait éliminée si toutes les institutions

les effets ont été égalisés : $(\quad) = 1 - \frac{(\quad)}{(\quad)}$; (2) La part de variance qui serait éliminée si

tous les effets du chercheur ont été égalisés : $(\quad)$. Cette analyse révèle que l'égalisation

Les effets institutionnels élimineraient plus de 73 % de la variance observée, tout en égalisant les chercheurs les effets élimineraient 49 % de la variance.

Grâce à ces estimations, nous pouvons également examiner si les rôles relatifs des institutions et les individus varient selon la distribution de la production institutionnelle (panneau B du tableau 4). Cela se produire si les institutions les plus ou les moins productives accordent un rôle disproportionné aux individus ou des effets institutionnels. Les colonnes de ce tableau proposent différentes manières de définir les effets supérieurs et inférieurs. institutions : les 5 %, 10 %, 25 % et 50 % supérieurs et inférieurs (nous devrions observer les plus grandes différences entre les 5 % supérieurs et inférieurs. La première ligne représente la différence de production moyenne des chercheurs. entre les deux groupes d'institutions : '. La deuxième ligne représente la différence due à la caractéristiques du chercheur, et la troisième ligne est la différence restante due à la effets de l'institution '. Les deux dernières lignes convertissent les composantes du chercheur et l'institution en parts relatives de la différence globale. Nos résultats montrent qu'environ 54 % de la variation observée dans la production des chercheurs entre les groupes institutionnels provient de caractéristiques différentielles des lieux, les 46 % restants étant attribuables à des différences dans les facteurs propres aux chercheurs. Ces valeurs concordent avec l'analyse de l'étude événementielle, qui Une part locale de 50 à 60 % a été suggérée. Notez que les estimations peuvent différer en raison de l'additif. L'exercice de décomposition consiste à analyser la différence entre deux groupes d'institutions plutôt que en faisant la moyenne de la part basée sur le lieu parmi tous les déménageurs, et en incluant également les déménageurs et les non-déménageurs dans l'estimation. De plus, la stabilité de la part de l'institution à travers différentes partitions est conforme à la relation linéaire illustrée dans la figure 6.

Ensuite, nous nous attaquons à la question de savoir si les hypothèses sous-jacentes au modèle à deux voies sont appuyé par une étude d'événement plus robuste.

C. Estimations de l'étude des événements de base

Nous motivons l'analyse de l'étude d'événement en montrant d'abord la variation d'identification derrière nos moteurs spécification. La figure A1 de l'annexe présente la distribution de notre variable indépendante clé, la Différence entre la production logarithmique moyenne dans l'établissement de destination et d'origine d'un déménageur. Par construction,

$= 0$ indique une sortie logarithmique moyenne égale à l'origine et à la destination. L'écart type est 0,77, avec un nombre significatif de mouvements importants dans les deux sens. Cette distribution est centrée.

environ -0,091 (ce qui signifie que le mouvement typique se fait vers une institution dont la production est d'environ 9

pour cent inférieur à l'institution d'origine) et relativement symétrique, indiquant un nombre à peu près égal de mouvements des institutions à faible rendement vers les institutions à rendement élevé (> 0) à celui des mouvements des institutions à rendement élevé vers les institutions à faible rendement (< 0).

Il est naturel de se demander pourquoi les scientifiques migrent. La preuve la plus directe est fournie par (Lerner et al., 2024), qui ont mené une enquête auprès de scientifiques en mobilité, qui révèle que la raison principale de la migration était un meilleur salaire, suivi de meilleures ressources et d'une meilleure qualité de vie. Il est possible que des établissements ayant un rendement par habitant plus faible offrent des salaires, des ressources et une qualité de vie plus élevés, et nous avons été rassurés par le fait que la figure A1 de l'annexe montre des migrations dans les deux sens (donc pas seulement vers des établissements plus productifs), et vers des établissements ayant un rendement de recherche similaire à celui de l'institut d'origine.

Pour contextualiser une taille de transfert de -0,091, le « chercheur itinérant » moyen de notre échantillon passe de l'Université de Chicago au Baylor College of Medicine. Le tableau A6 présente les 50 meilleures institutions américaines en fonction de la production moyenne des chercheurs (en niveaux, et non en logarithmes). À titre d'illustration, ce tableau a été limité aux institutions comptant plus de 100 chercheurs et 10 chercheurs itinérants entre 2015 et 2023 afin de mieux comprendre la variation à l'origine de nos estimations.

Cependant, lors de l'estimation des régressions, nous incluons toutes les institutions.

Français Dans le cadre d'une exploration initiale de l'impact du déménagement vers une nouvelle institution sur la production logarithmique d'un scientifique, nous traçons le changement après le déménagement dans la production logarithmique du scientifique (sur l'axe des Y) par rapport à la taille du déménagement (sur l'axe des X) dans la Figure 6. Nous définissons le changement dans la production logarithmique d'un scientifique comme sa production logarithmique moyenne après le déménagement moins sa production logarithmique moyenne avant le déménagement. La pente de la droite ajustée indique dans quelle mesure le déménagement vers une institution à plus haut rendement influence la production logarithmique d'un scientifique. Si le lieu expliquait entièrement les différences de production logarithmique, la pente serait de 1. Inversement, si des facteurs individuels expliquaient entièrement la variation, la pente serait de 0. La Figure 6 montre une pente de 0,73 suggérant que 73 % de la variation de la production logarithmique du scientifique peut être attribuée à des facteurs institutionnels, tandis que les 27 % restants sont dus à des différences individuelles. De plus, la nature linéaire et symétrique de la relation au-dessus et en dessous d'une taille de mouvement de zéro soutient notre

hypothèse de séparabilité additive entre et dans l'équation 1, ce qui signifie que les mouvements vers des institutions relativement plus productives sont quantitativement similaires, mais avec le signe opposé, aux mouvements vers des institutions relativement moins productives.

Nous présentons les résultats de l'étude d'événement de l'équation 2 de la figure 7a pour la période 1945-1923. Français La figure représente les (β, σ) avec des intervalles de confiance à 95 %, en normalisant $(\beta, \sigma) = -1$ à zéro. Les estimations des erreurs types regroupées au niveau de l'établissement afin de tenir compte de la corrélation intra-établissement dans la production logarithmique de la recherche. La figure ne montre aucune pré-tendance significative dans (β, σ) au cours des dix années précédant la relocalisation, ce qui confirme notre hypothèse selon laquelle les facteurs de production non observés basés sur la personne ne déterminent pas la décision de déménager. La production logarithmique reste stable jusqu'à environ deux ans après le déménagement, après quoi elle présente une augmentation discontinue jusqu'à environ 0,5. Cet effet reste relativement stable l'année suivante, culminant à environ 0,6. Ces résultats suggèrent que les facteurs institutionnels représentent environ 50 à 60 % de la variation de la production de recherche au fil du temps¹⁹. Cette estimation est légèrement inférieure à l'estimation statique de 72 % de la figure 6, reflétant la nature dynamique des effets spécifiques au lieu et contrôlant les attributs non observables des chercheurs par le biais d'effets fixes individuels.

¹⁸Nous définissons la taille du mouvement comme la différence entre la production logarithmique moyenne de deux institutions, ce qui est différent des différences entre les différences logarithmiques moyennes. Le tableau A6 présente la production de recherche moyenne en niveaux pour une institution ; l'enregistrement de ces valeurs produit donc le logarithme d'une moyenne et ne correspond pas à la taille du mouvement.

¹⁹Nous exécutons la même spécification en utilisant les résultats de recherche sans aucune pondération de citation comme résultat alternatif et obtenir des résultats similaires.

La figure 7b présente la même analyse d'étude d'événements pour les migrants de la période plus récente, de 1995 à 2023. L'effet de lieu estimé est d'une ampleur similaire, indiquant que l'essor de la communication numérique et la simplification de la collaboration à distance n'ont pas sensiblement réduit l'influence de l'environnement institutionnel local sur la production. De plus, afin d'examiner l'effet potentiel de la mesure, nous limitons l'échantillon aux institutions comptant au moins 100 auteurs.

10 mouvements entre 2015 et 2023, alors nous obtenons essentiellement le même résultat.

D. Hétérogénéité selon la taille du mouvement et les caractéristiques du chercheur

Pour comprendre les mécanismes sous-jacents de la façon dont le lieu influence les résultats de la recherche, nous examinons si nos effets varient en fonction du type de déplacement et des caractéristiques du chercheur.

Le premier mécanisme que nous explorons est le rôle du sens du transfert. Nous classons les transferts comme positifs (> 0) si un chercheur rejoint un établissement à haut rendement et négatifs (< 0) s'il rejoint un établissement à faible rendement. Les établissements à haut rendement disposent probablement de davantage de ressources et de meilleures installations, ce qui permet aux scientifiques de travailler à la pointe de leur domaine avec des techniques de recherche plus avancées. En supposant que les chercheurs intègrent ces avantages à leurs propres recherches lorsqu'ils rejoignent des établissements à haut rendement, mais qu'ils ne désapprennent pas ces compétences lorsqu'ils rejoignent des établissements à faible rendement, nous nous attendons à observer des effets de lieu plus importants pour les transferts positifs. Nous distinguons l'effet estimé du lieu par le sens du transfert dans la figure A2. Ces graphiques confirment notre théorie, car nous observons une différence d'effets de lieu entre les scientifiques qui rejoignent des établissements à haut rendement et ceux qui rejoignent des établissements à faible rendement. L'effet de lieu estimé pour les transferts positifs est plus important et varie de 0,6 à 0,8, tandis que l'effet de lieu pour les transferts négatifs se situe entre 0,3 et 0,5.

Compte tenu du volet complémentaire de recherche sur le rôle des « stars » dans la découverte, nous examinons l'influence des chercheurs vedettes dans l'explication de nos estimations de l'effet de lieu (Azoulay et al., 2012, 2019 ; Waldinger, 2016). Nous définissons les « stars » comme les chercheurs se situant dans le 95e percentile supérieur de la distribution de la production de leur institution et modifions notre définition pour représenter la différence entre la production moyenne des stars dans les institutions de destination et d'origine. La figure A1 de l'annexe illustre la distribution de cette taille de déplacement révisée, tandis que la figure 8a présente l'effet de lieu estimé à l'aide de cette nouvelle définition. Les résultats indiquent que les stars contribuent significativement à la production des chercheurs en mouvement, avec un effet de lieu estimé d'environ 0,4. En d'autres termes, les scientifiques vedettes et tout ce qui leur est associé (qui peut inclure les laboratoires de recherche et les étudiants) expliquent environ les deux tiers de l'effet de lieu que nous observons dans nos résultats de base. À l'instar de (Azoulay et al., 2019), nous explorons également l'impact de la production des co-auteurs lorsqu'une star déménage dans la figure 8b. Bien que les estimations soient bruyantes, nous voyons que les estimations ponctuelles varient de -0,3 à -0,8, ce qui implique que la production d'un coauteur diminue de 30 à 80 % lorsque son coauteur part et est considéré comme un chercheur vedette.

Français Motivés par des recherches antérieures sur la production du cycle de vie des chercheurs et les effets d'agglomération (Dietz et Bozeman, 2005 ; Jones, 2010, 2009 ; Levin et Stephan, 1991 ; Myers, 2020 ; Carlino et Kerr, 2015 ; Kerr et Robert-Nicoud, 2020 ; Moretti, 2021), nous examinons également l'hétérogénéité des effets de lieu selon l'âge du migrant et la taille de la ville. Les figures A3 et A4 montrent des amplitudes similaires des effets de lieu, que les chercheurs déménagent avant ou après l'âge de 40 ans, ou qu'ils déménagent dans des villes plus grandes ou plus petites, mesurées par le nombre de scientifiques dans ces villes, à l'exclusion de leur propre institution. Cela suggère que ni l'âge ni la taille de l'agglomération scientifique n'ont d'impact significatif sur le rôle des effets de lieu dans l'évolution de la production des chercheurs.

Pour mieux comprendre l'hétérogénéité des effets de lieu, nous suivons la pratique courante des modèles de migration et examinons les corrélats de ces effets de lieu. Nous projetons nos effets fixes estimés au niveau des institutions sur diverses caractéristiques obtenues auprès de la National Science Foundation.

Français L'enquête annuelle de l'Organisation sur la recherche et le développement dans l'enseignement supérieur (HERD). La figure A5 de l'annexe présente les résultats de ces régressions bivariées, révélant des corrélations positives significatives entre les effets de production spécifiques au lieu et plusieurs facteurs institutionnels, notamment les dépenses totales de R&D, le succès à attirer des subventions de recherche compétitives et l'investissement dans l'instrumentation scientifique à grande échelle. Ces corrélations suggèrent que les aspects gourmands en ressources de l'environnement institutionnel peuvent jouer un rôle important dans la formation de la production scientifique.

E. Robustesse

E.1. Sous-champs

Afin de répondre aux préoccupations concernant le biais de sélection lié aux différences de résultats entre domaines scientifiques, nous étendons notre spécification d'étude d'événements de référence afin d'y intégrer des effets fixes spécifiques à chaque domaine. Cela nous permet de prendre en compte la possibilité que les chercheurs privilégient certains sous-domaines, connaissent une productivité croissante ou bénéficient de découvertes majeures dans ces domaines grâce à des externalités de connaissances. En contrôlant ces tendances au niveau du domaine, nous pouvons isoler avec plus de confiance les effets causaux que nous cherchons à mesurer dans notre analyse principale. Les résultats de cette spécification alternative, présentés dans la figure A6, démontrent que nos principales conclusions restent robustes malgré l'inclusion de ces effets de domaine. Cette cohérence entre les spécifications fournit des preuves convaincantes que les effets observés représentent de véritables phénomènes plutôt que des artefacts de sélection différentielle ou des trajectoires de résultats spécifiques à un domaine.

E.2. Échantillons alternatifs

Français Pour évaluer la robustesse de nos résultats à la sélection des auteurs, nous reproduisons notre analyse pour (1) seulement les premier et dernier auteurs et (2) tous les auteurs. Pour les premier et dernier auteurs, la part institutionnelle estimée de la variation de la production dans cet échantillon est légèrement plus petite, environ 40 % (figure A8 de l'annexe, ce qui suggère que les effets liés au lieu peuvent jouer un rôle légèrement plus faible parmi les chercheurs occupant des postes d'auteurs principaux. En revanche, lorsque nous élargissons l'échantillon pour inclure tous les auteurs répertoriés, la part institutionnelle augmente à environ 60 % (figure A9 de l'annexe), ce qui implique que les effets institutionnels sont plus prononcés lorsque l'on considère l'ensemble plus large des contributeurs à la production scientifique. Ces résultats suggèrent que si les différences de production induites par l'institution persistent dans les rôles d'auteur, leur ampleur varie selon que l'accent est mis sur les chercheurs principaux ou sur l'ensemble de l'équipe de recherche.

E.3. Spécifications alternatives

Une spécification alternative à notre équation d'estimation principale 2 consiste à ne pas inclure l'ensemble des effets fixes relatifs à l'année (,). L'inclusion de ces effets fixes permet de dégager une tendance commune dans les résultats de recherche.

pour tous les déménageurs, indépendamment de la taille et de la direction du déménagement. Nous montrons dans la figure A7 de l'annexe que l'absence de ces covariables a peu d'effet sur nos résultats.

V. Conclusion

La science fondamentale est la pierre angulaire de l'innovation médicale à long terme, et notre étude souligne que le lieu où la science est menée influence significativement la quantité et l'impact de la recherche produite. Nous notons l'importance des grands pôles de recherche et des institutions particulières dans la production d'une part disproportionnée de la science mondiale, mais constatons une variation nettement plus importante dans la dispersion géographique de la production scientifique moyenne. En suivant la production de recherche des scientifiques qui changent d'institution, nous estimons que 50 à 60 % de la variation de la production des chercheurs est attribuable à des facteurs spécifiques au lieu, ce qui signifie qu'une part significative, au moins 40 %, est la conséquence du fait que les institutions plus productives sélectionnent les chercheurs plus productifs, plutôt que de les rendre plus productifs. Ces effets n'ont pas diminué au fil du temps, même si les collaborations interinstitutionnelles sont devenues plus faciles, et soulignent le rôle important des facteurs institutionnels dans la promotion de la production scientifique dans les sciences de la vie. Par conséquent, réduire la générosité du financement public ou philanthropique dans les institutions qui ont une production par scientifique élevée réduira directement la production de connaissances, en particulier de connaissances commercialement pertinentes. Près des deux tiers des effets institutionnels sont prédits par la présence de chercheurs vedettes, notamment l'infrastructure dont ils ont besoin et les étudiants qu'ils attirent, ce qui accroît la productivité des autres scientifiques. Cela concorde avec les travaux de (Waldinger, 2016), qui ont constaté que le licenciement de scientifiques célèbres sous l'Allemagne nazie a entraîné un déclin persistant de la productivité de la recherche. De même, (Azoulay et al., 2019) soulignent que le décès d'un chercheur vedette entraîne une forte baisse de la productivité de ses collaborateurs.

Français Nos résultats contribuent au programme de recherche plus large sur la fonction de production de l'innovation. Une abondante littérature a examiné d'autres facteurs qui façonnent la production de recherche, notamment les mécanismes de financement, la composition des équipes, la concurrence et les incitations professionnelles (Lach et Schankerman, 2008 ; Azoulay et al., 2013 ; Tabakovic et Wollmann, 2019 ; Azoulay et al., 2019 ; Bloom et al., 2020 ; Myers, 2020 ; Arora et al., 2021 ; Acemoglu, 2023 ; Babina et al., 2023 ; Hill et Stein, 2023, 2024 ; Tham et al., 2024). Nous contribuons à cette entreprise en nous concentrant sur un déterminant clé : le lieu où la recherche se déroule. Français Nos travaux complètent également les recherches sur les sources de brevetage, qui sont un résultat en aval de la recherche en sciences fondamentales (Jaffe, 1989 ; Trajtenberg et al., 1997 ; Jensen et Thursby, 2001 ; Thursby et al., 2007 ; Belenzon et Schankerman, 2013 ; Williams, 2017 ; Bell et al., 2019 ; Koning et al., 2020 ; Akcigit et al., 2021 ; Tartari et Stern, 2021 ; Hausman, 2022 ; Moretti et al., 2023 ; Lerner et al., 2024). Si les forces qui influencent le brevetage façonnent également la recherche fondamentale – car elles sont souvent entreprises par les mêmes personnes et sont susceptibles d'être des activités complémentaires – alors le rôle de la place dans l'innovation médicale au sens large pourrait être encore plus important que nous l'avons estimé.

Dans le contexte de la découverte en sciences de la vie, l'institution d'un chercheur a un impact considérable sur sa production. Par conséquent, si les bailleurs de fonds souhaitent maximiser l'impact de leurs ressources sur la découverte scientifique fondamentale, les ressources institutionnelles (comme un don non associé à un scientifique en particulier) devraient être allouées aux institutions les plus productives. Si les bailleurs de fonds

Les chercheurs sont confrontés au choix suivant : financer deux scientifiques ayant un même rendement ; financer le chercheur de l'institution la plus productive générera davantage de recherche. L'ampleur des effets potentiels est importante : les tableaux A6 et A7 de l'annexe montrent une variation de 3 à 4 x de la production moyenne des chercheurs selon les institutions, même parmi les plus productives, ce qui représente une portion tronquée de la distribution réelle de la production institutionnelle. Nos estimations prédisent que plus de la moitié de cette variation est causale, ce qui a des implications pour les scientifiques souhaitant changer d'institution, notamment d'institutions à haut rendement vers des institutions à faible rendement, en échange d'une offre plus avantageuse. Dans notre analyse, nous avons étudié les migrations de scientifiques aux États-Unis, mais si nos résultats sont généralisés aux migrations dans d'autres pays, il y aura une baisse conséquente de la création de connaissances si les chercheurs basés aux États-Unis partent vers des institutions ex-américaines ayant une production moyenne par chercheur plus faible (l'effet potentiel de ce phénomène peut être observé en comparant la production moyenne des principales institutions américaines dans les tableaux A6 et A7 de l'annexe).

Parallèlement, les établissements pourraient souhaiter accroître leur production. Nos résultats suggèrent que l'agglomération n'est pas une condition préalable à une production de recherche élevée (ce qui fait écho à un point soulevé dans (Ellison et al., 1997) concernant les agglomérations dans le secteur manufacturier). En d'autres termes, les agglomérations ne rendent pas les chercheurs plus productifs et reflètent l'agglomération de scientifiques productifs. Deuxièmement, des facteurs associés à la présence de chercheurs vedettes expliquent cet effet ; le départ de chercheurs vedettes réduit également la production des scientifiques de l'établissement qu'ils ont quitté. Investir dans les ressources qui attirent les chercheurs vedettes pourrait être un moyen d'accroître la production des établissements. Cependant, nous soulignons qu'une telle réaffectation pourrait nécessiter des ressources importantes (notamment à la lumière de Myers (2020), qui souligne que des sommes importantes sont nécessaires pour inciter les scientifiques à changer de domaine, sans parler d'établissements).

Ces résultats suggèrent plusieurs pistes de recherche futures. Une question clé est d'identifier les caractéristiques institutionnelles qui rendent certaines institutions plus productives que d'autres. Notre exploration a révélé que les institutions ayant des dépenses de R&D plus élevées étaient plus productives, mais il ne s'agissait pas d'une affirmation causale ; il s'agissait plutôt d'une simple façon de valider les effets institutionnels. Comprendre le lien entre financement et résultats de la recherche, comme l'indiquent Myers et Lanahan (2022), et explorer la manière dont les différentes institutions structurent leurs environnements de recherche peut éclairer les meilleures pratiques visant à optimiser l'efficacité du financement de la recherche. Enfin, si les environnements institutionnels ont un impact causal sur la production scientifique, comme le montrent nos données, les initiatives ciblant les jeunes chercheurs pourraient être très rentables si elles visaient à encourager les talents au sein des institutions les plus productives (Azoulay et al., 2011, 2019, 2021). Comprendre ces relations serait un moyen fructueux d'accroître la découverte scientifique fondamentale, dont les implications pour le bien-être et la science sont considérables.

Références

Acemoglu, D. (mai 2023). Innovation déformée : le marché oriente-t-il correctement la technologie ? Documents et actes de l'AEA 113, 1–28.

Akcigit, U., D. Hanley et N. Serrano-Velarde (janvier 2021). Retour aux fondamentaux : Retombées de la recherche fondamentale, politique d'innovation et croissance. *The Review of Economic Studies* 88(1), 1–43.

Allison, PD et J.S. Long (1990). Effets des départements sur la productivité scientifique. *American Sociological Review* 55(4), 469–478. Éditeur : [American Sociological Association, Sage Publications, Inc.].

Arora, A., S. Belenzon et L. Sheer (mars 2021). Retombées de connaissances et investissement des entreprises dans Recherche scientifique. *American Economic Review* 111(3), 871–898.

Azoulay, P., C. Fons-Rosen et J.S. Graff Zivin (août 2019). La science fait-elle progresser un enterrement à la fois ? Le temps ? *American Economic Review* 109(8), 2889–2920.

Azoulay, P., E. Fuchs, AP Goldstein et M. Kearney (janvier 2019). Financer la recherche innovante : promesses et défis du modèle ARPA. *Innovation Policy and the Economy*, 19, 69–96. Éditeur : The University of Chicago Press.

Azoulay, P., J.S. Graff Zivin et G. Manso (2011). Incitations et créativité : données issues des sciences de la vie universitaires. *The RAND Journal of Economics* 42(3), 527–554. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-2171.2011.00140.x>.

Azoulay, P., J.S. Graff Zivin et G. Manso (janvier 2013). Évaluation par les pairs des National Institutes of Health : défis et pistes de réforme. *Innovation Policy and the Economy*, 13, 1–22. Éditeur : The University of Chicago Press.

Azoulay, P., WH Greenblatt et ML Heggeness (novembre 2021). Effets à long terme d'une exposition précoce à la recherche : Données probantes des « Bénéficiaires » du NIH. *Politique de recherche* 50(9), 104332.

Azoulay, P. et D. Li (mars 2022). Financement de subventions scientifiques. Dans *Innovation and Public Policy*, rapport de conférence du National Bureau of Economic Research, p. 117–150. Presses de l'Université de Chicago.

Azoulay, P., JSG Zivin et BN Sampat (mars 2012). La diffusion des connaissances scientifiques à travers le temps et l'espace : témoignages des transitions professionnelles des superstars de la médecine. Dans « Le rythme et la direction de l'activité inventive revisités », p. 107–155. Presses de l'Université de Chicago.

Babina, T., AX He, ST Howell, ER Perlman et J. Staudt (mai 2023). Couper le moteur de l'innovation : impact des chocs de financement fédéral sur les brevets, l'entrepreneuriat et les publications universitaires*. *The Quarterly Journal of Economics* 138(2), 895–954.

Belenzon, S. et M. Schankerman (juillet 2013). Diffuser l'information : géographie, politique et savoir Débordements. *Revue d'économie et de statistique* 95(3), 884–903.

Bell, A., R. Chetty, X. Jaravel, N. Petkova et J. Van Reenen (mai 2019). Qui devient inventeur aux États-Unis ? L'importance de l'exposition à l'innovation*. *The Quarterly Journal of Economics* 134(2), 647–713.

Bloom, N., CI Jones, J. Van Reenen et M. Webb (avril 2020). Les idées deviennent-elles plus difficiles à trouver ? Américain *Revue économique* 110(4), 1104–1144.

- Bronnenberg, BJ, J.-P.H. Dubé et M. Gentzkow (mai 2012). L'évolution des préférences de marque : Preuves de la migration des consommateurs. *American Economic Review* 102(6), 2472–2508.
- Bryan, KA, Y. Ozcan et B. Sampat (mai 2020). Citations de brevets dans le texte : Guide d'utilisation. *Recherche Politique* 49(4), 103946.
- Budish, E., BN Roin et H. Williams (juillet 2015). Les entreprises sous-investissent-elles dans la recherche à long terme ? Données probantes Extrait d'essais cliniques sur le cancer. *American Economic Review* 105(7), 2044–2085.
- Carlino, G. et WR Kerr (2015). Agglomération et innovation. *Manuel d'urbanisme régional et urbain. économie* 5, 349–404. ISBN : 1574-0080 Éditeur : Elsevier.
- Chetty, R. et N. Hendren (août 2018). Impacts des quartiers sur la mobilité intergénérationnelle II : Estimations au niveau des comtés*. *The Quarterly Journal of Economics* 133(3), 1163–1228.
- Clarivate (2023). Facteur d'impact des revues 2022, *Journal Citation Reports*. Rapport technique.
- Dietz, J.S. et B. Bozeman (avril 2005). Carrières universitaires, brevets et productivité : expérience en entreprise en tant que capital humain scientifique et technique. *Politique de recherche* 34(3), 349–367.
- Ellison, SF, I. Cockburn, Z. Griliches et J. Hausman (1997). Caractéristiques de la demande de produits pharmaceutiques : étude de quatre céphalosporines. *The RAND Journal of Economics* 28(3), 426–446. Éditeur : [RAND Corporation, Wiley].
- Finkelstein, A., M. Gentzkow et H. Williams (novembre 2016). Sources de variation géographique des soins de santé : données issues de la migration des patients*. *The Quarterly Journal of Economics* 131(4), 1681–1726.
- Furman, JL et S. Stern (août 2011). Grimper sur les épaules des géants : l'impact des institutions sur la recherche cumulative. *American Economic Review* 101(5), 1933–1963.
- Gruber, J., S. Johnson et E. Moretti (septembre 2022). Productivité et coûts locaux en sciences.
- Guzman, J., F. Murray, S. Stern et H. Williams (juillet 2023). Accélérer les écosystèmes d'innovation : promesses et défis des moteurs d'innovation régionaux. Dans *Entrepreneuriat et politique d'innovation et économie*, volume 3. Presses de l'Université de Chicago.
- Hall, BH et J. Lerner (janvier 2010). Chapitre 14 - Le financement de la R&D et de l'innovation. Dans BH Hall et N. Rosenberg (éd.), *Manuel d'économie de l'innovation*, volume 1 du *Manuel d'économie de l'innovation*, vol. 1, pp. 609–639. Hollande-Septentrionale.
- Hausman, N. (juillet 2022). Innovation universitaire et croissance économique locale. *Revue d'économie et de statistique* 104(4), 718–735.
- Hill, R. et C. Stein (2023). Scooped! Estimation des récompenses pour la priorité en science.
- Hill, R. et C. Stein (2024). Course vers le bas : concurrence et qualité en science.
- Howell, ST (avril 2017). Financement de l'innovation : données probantes issues des subventions de R&D. *American Economic Review* 107(4), 1136–1164.
- Hvide, HK et BF Jones (juillet 2018). Innovation universitaire et privilège du professeur. *Américain Revue économique* 108(7), 1860–1898.

- Jaffe, AB (1989). Effets réels de la recherche universitaire. *The American Economic Review* 79(5), 957–970.
Éditeur : American Economic Association.
- Jensen, R. et M. Thursby (mars 2001). Épreuves et prototypes à vendre : l'octroi de licences universitaires
Inventions. *American Economic Review* 91(1), 240–259.
- Jones, BF (janvier 2009). Le fardeau du savoir et la « mort de l'homme de la Renaissance » : l'innovation est-elle
De plus en plus difficile ? *Revue d'études économiques* 76(1), 283–317.
- Jones, BF (février 2010). Âge et grande invention. *Revue d'économie et de statistique*, 92(1), 1–14.
- Kerr, WR et F. Robert-Nicoud (août 2020). Pôles technologiques. *Revue des perspectives économiques* 34(3),
50–76.
- Kesselheim, AS, RM Cook-Deegan, DE Winickoff et MM Mello (août 2013). Brevets génétiques : le
La Cour suprême s'exprime enfin. *The New England journal of medicine* 369(9), 869–875.
- Keys, BJ, N. Mahoney et H. Yang (janvier 2023). Quels sont les facteurs déterminants de la détresse financière des consommateurs ?
et facteurs basés sur la personne. *The Review of Financial Studies* 36(1), 42–69.
- Koning, R., S. Samila et J.-P. Ferguson (mai 2020). Genre de l'inventeur et orientation de l'invention. *AEA*
Documents et actes 110, 250–254.
- Lach, S. et M. Schankerman (2008). Incitations et invention dans les universités. pp. 32.
- Lerner, J., H.J. Manley, C. Stein et H.L. Williams (janvier 2024). Les érudits errants : comprendre
L'hétérogénéité de la commercialisation universitaire.
- Levin, SG et PE Stephan (1991). Productivité de la recherche tout au long du cycle de vie : données probantes pour les
chercheurs universitaires. *The American Economic Review* 81(1), 114–132. Éditeur : American Economic Association.
- Marx, M. et A. Fuegi (2020). Traductions Confiance dans la science : Brevet à la une dans le monde entier ci_eprint :
d'articles scientifiques. [https://](https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.3145) *Revue de gestion stratégique* 41(9), 1572–1594.
onlinelibrary.wiley.com/doi/pdf/10.1002/smj.3145.
- Moretti, E. (octobre 2021). L'effet des pôles de haute technologie sur la productivité des meilleurs inventeurs. Américain.
Revue économique 111(10), 3328–3375.
- Moretti, E., C. Steinwender et J. Van Reenen (février 2023). Le butin intellectuel de la guerre ? R&D de défense, productivité
et retombées internationales. *Revue d'économie et de statistique*, 1–46.
- Myers, K. (octobre 2020). L'élasticité de la science. *Revue américaine d'économie appliquée*, vol. 12(4).
103–134.
- Myers, K. et L. Lanahan (2022). Estimation des retombées de la R&D financée par des fonds publics : données des États-Unis
Département de l'Énergie. *American Economic Review* 112(7), 2392–2423.
- Nelson, RR (1959). L'économie simple de la recherche scientifique fondamentale. *Journal of Political Economy* 67(3), 297–
306. Éditeur : University of Chicago Press.
- Priem, J., H. Piwowar et R. Orr (juin 2022). OpenAlex : un index entièrement ouvert d'ouvrages universitaires, d'auteurs et de sources.
lieux, institutions et concepts. [arXiv:2205.01833 \[cs\]](https://arxiv.org/abs/2205.01833).

Sung, NS, WF Crowley, Jr, M. Genel, P. Salber, L. Sandy, LM Sherwood, SB Johnson, V. Catanese, H. Tilson, K. Getz, EL Larson, D. Scheinberg, EA Reece, H. Slavkin, A. Dobs, J. Grebb, RA Martinez, A. Korn et D. Rimoin (mars 2003). Défis centraux auxquels est confrontée l'entreprise nationale de recherche clinique. *JAMA* 289(10), 1278–1287.

Tabakovic, H. et TG Wollmann (octobre 2019). L'impact de l'argent sur la science : données probantes
Résultats inattendus du football américain. *Journal of Public Economics* 178, 104066.

Tartari, V. et S. Stern (mai 2021). Plus qu'une tour d'ivoire : l'impact des institutions de recherche sur la
Quantité et qualité de l'entrepreneuriat.

Tham, WY, J. Staudt, ER Perlman et SD Cheng (février 2024). Les talents scientifiques s'échappent
Lacunes de financement. *arXiv:2402.07235 [econ]*.

Thursby, M., J. Thursby et S. Gupta-Mukherjee (août 2007). L'octroi de licences a-t-il de réels effets sur la recherche
universitaire ? Une approche du cycle de vie. *Journal of Economic Behavior & Organization* 63(4), 577–598.

Trajtenberg, M., H. , Rebecca, et A. Jaffe (1997, janvier). Brevets universitaires et brevets d'entreprise : une fenêtre sur le
caractère fondamental de l'invention. *Économie de l'innovation et des nouvelles technologies*, 5(1), 19–50. Éditeur : Routledge.
_eprint : <https://doi.org/10.1080/10438599700000006>.

Waldinger, F. (décembre 2016). Bombes, cerveaux et science : le rôle du capital humain et physique dans la création de connaissances
scientifiques. *The Review of Economics and Statistics* 98(5), 811–831.

Williams, HL (août 2017). Comment les brevets affectent-ils les investissements en recherche ? *Revue annuelle de l'éco-
nomics* 9 (Volume 9, 2017), 441–469. Éditeur : Annual Reviews.

Zerhouni, EA (2005, septembre). Recherche biomédicale américaine, sciences fondamentales, translationnelles et cliniques.
JAMA 294(11), 1352–1358.

Zucker, LG, MR Darby et MB Brewer (1998). Le capital humain intellectuel et la naissance des États-Unis
Entreprises de biotechnologie. *The American Economic Review* 88(1), 290–306. Éditeur : American Economic
Association.

Tableau 1 : Villes ayant la plus grande production annuelle par chercheur, 2015-2023

Rang de la ville		Part moyenne de la production des auteurs dans la production mondiale (%)	
	Boston–Cambridge–Newton, MA–NH Bay Area,	1,58	9,79
1	CA Trenton, NJ	1,56	7,67
2	Dallas–Fort	1,42	0,54
3	Worth–Arlington, TX New York–Newark–	1,35	0,86
	Jersey City, NY–NJ–PA San Diego–Carlsbad, CA Seattle–	1,19	6,67
4	Tacoma–Bellevue, WA New	1,05	2,25
5	Haven–Milford, CT St. Louis, MO–IL	1,02	1,68
6	Worcester, MA–CT Ithaca, NY	1,02	1,20
7	Houston–The	0,97	1,39
8 9	Woodlands–Sugar Land,	0,97	0,28
10	TX Austin–	0,85	0,78
11	Round Rock, TX Los Angeles–Long Beach–Anaheim,	0,82	1,43
12	CA Memphis, TN–MS–AR Santa	0,81	0,30
13	Cruz–Watsonville, CA Philadelphie–Camden–	0,76	1,85
14	Wilmington, PA–NJ–DE–MD	0,74	0,34
15	Baltimore–Columbia–Towson, MD	0,73	0,15
16	Chicago–Naperville–Elgin, IL–IN–WI Rochester, MN	0,73	1,34
17		0,67	1,29
18		0,58	0,98
19 20		0,58	0,09

Notes : Le tableau présente les 20 premières villes en termes de production moyenne par auteur-année en recherche scientifique fondamentale dans les revues des sciences de la vie. La production de recherche est pondérée par le nombre total de chercheurs dans chaque lieu. Les revues incluses sont : Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neurone, Cellule souche, PLOS One, Oncogene, Journal of Chimie biologique et le journal FASEB entre 2015 et 2023. La part de la production mondiale est calculée à partir du total des publications sur la période 2015-2023.

Tableau 2 : Principales institutions dans les dix villes ayant la plus grande production scientifique fondamentale, 2015-2023

Rang	Institution	Production moyenne des chercheurs
I. Boston-Cambridge-Newton, MA-NH		
1.	Institut du cancer Dana-Farber	2,81
2.	Institut de technologie du Massachusetts	2,33
3.	Université de Harvard	1,64
4.	Général de Masse Brigham	1,52
5.	Hôpital pour enfants de Boston	1,15
II. Région de la baie de San Francisco, Californie		
1.	Université de Stanford	1,86
2.	Instituts Gladstone	1,77
3.	Université de Californie à Berkeley	1,33
4.	Université de Californie à San Francisco	1,31
5.	Laboratoire national Lawrence Berkeley	0,67
III. Trenton, New Jersey		
1.	Université de Princeton	1,42
IV. Dallas-Fort Worth-Arlington, Texas		
1.	Centre médical du sud-ouest de l'Université du Texas	1,35
V. New York-Newark-Jersey City, NY-NJ-PA		
	Centre de cancérologie Memorial Sloan Kettering 1.	2,24
	Novartis (États-Unis) 2. 3. 4. 5.	1,72
	École de médecine Icahn du Mont Sinai	1,69
	Laboratoire de Cold Spring Harbor	1,42
	Université Rockefeller	1,34
VI. San Diego-Carlsbad, Californie		
1.	Institut Salk d'études biologiques	1,73
2.	Université de Californie à San Diego	1,00
3.	Institut de recherche Scripps	0,91
4.	Institut de découverte médicale Sanford Burnham Prebys	0,51
VII. Seattle-Tacoma-Bellevue, Washington		
	Institut Allen 1.	2,40
2.	Centre de cancérologie Fred Hutch	1,42
3.	Université de Washington	0,85
VIII. New Haven-Milford, Connecticut		
	Université Yale 1.	1,01
IX. Saint-Louis, MO-IL		
1.	Université de Washington à Saint-Louis	1,06
2.	Université de Saint-Louis	0,40
X. Worcester, MA-CT		
	École de médecine Chan de l'Université du Massachusetts 1.	0,97

Remarques : Le tableau présente les principales institutions dans les dix villes les plus productives en fonction de la production moyenne des chercheurs en recherche scientifique fondamentale dans la vie. Revues scientifiques. La mesure est pondérée par le nombre de chercheurs dans chaque zone géographique. Les revues incluses sont : Cell, Nature, Science, Nature Biologie cellulaire, génétique naturelle, médecine naturelle, biotechnologie naturelle, biologie chimique naturelle, neurosciences naturelles, neurone, cellules souches cellulaires, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal entre 2015 et 2023.

Tableau 3 : Conception des moteurs : statistiques récapitulatives

	Signifier	Écart type
A. Niveau individuel		
Tout le monde (N = 296 757)		
Années de publication active	1,79	2,03
Taille moyenne de l'équipe	3,62	0,66
Taille annuelle de la ville	7386,67	7404,94
Productivité annuelle	0,39	2,08
Productivité à vie	0,79	4,59
Nombre de mouvements	0,13	0,33
Personnes n'ayant pas déménagé (N = 258 948)		
Années de publication active	1,48	1,51
Taille moyenne de l'équipe	3,63	0,67
Taille annuelle de la ville	7369,74	7570,54
Productivité annuelle	0,38	2,16
Productivité à vie	0,60	3,57
Déménageurs (N = 37 809)		
Années de publication active	3,90	3,42
Taille moyenne de l'équipe	3,56	0,56
Taille annuelle de la ville	7502,61	6150,92
Productivité annuelle	0,49	1,42
Productivité à vie	2,13	8,70
Déménageurs ponctuels (N = 25 845)		
Années de publication active	4,12	3,58
Taille moyenne de l'équipe	3,56	0,56
Taille annuelle de la ville	7504,57	6127,13
Productivité annuelle	0,49	1,32
Productivité à vie	2,24	9,21
B. Au niveau de la ville		
Nombre d'établissements	169,46	173,15
Taille annuelle de la ville	7796,07	8134,21
Productivité annuelle	0,38	2,04
Productivité à vie	47,25	480,80

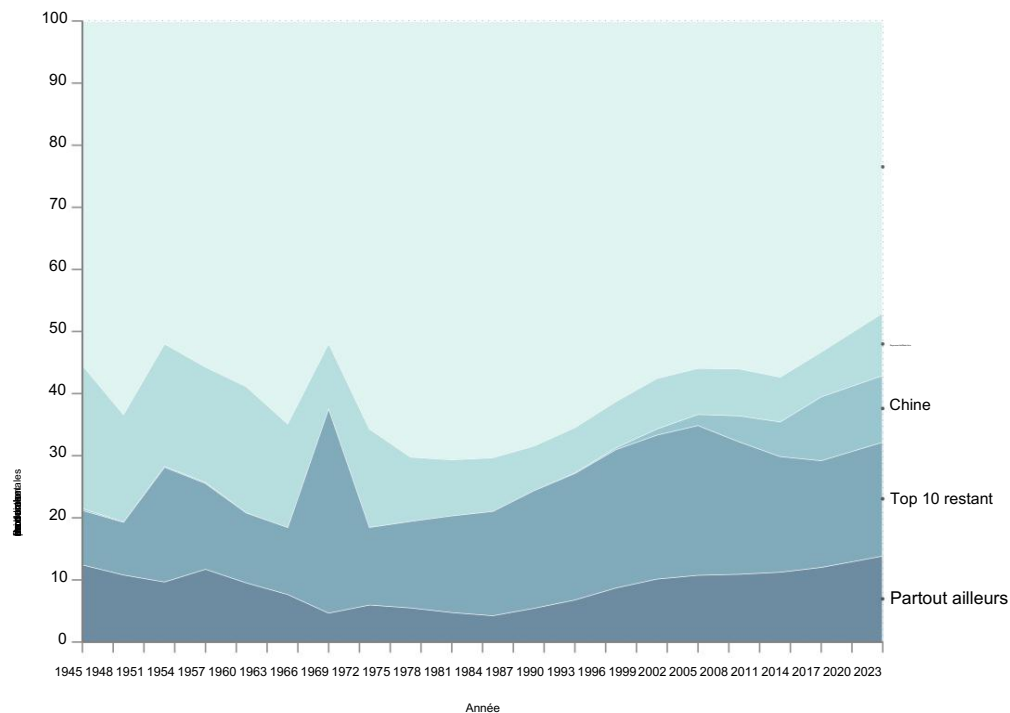
Notes : Le tableau présente les statistiques récapitulatives de notre échantillon d'auteurs pour la période 1945-2023. Le panneau A présente les statistiques au niveau individuel et le panneau B présente les statistiques au niveau de la ville.

Tableau 4 : Décompositions de la production des chercheurs

Panneau A : Décomposition de la variance de la production moyenne des chercheurs				
Variance de la production totale des chercheurs	0,16			
Attribuable aux effets institutionnels	0,08			
Attribuable aux effets du chercheur	0,04			
Attribuable aux effets de l'institution et du chercheur	0,28			
Part de la variance qui serait réduite si :				
Les effets des chercheurs ont été rendus égaux	0,49			
Les effets institutionnels ont été rendus égaux	0,73			
Panneau B : Décomposition additive de la production moyenne des chercheurs				
Haut et bas	5%	10%	25%	50%
Différence dans la production des chercheurs				
Dans l'ensemble	3,52	2,79	1,87	1,14
chercheurs	1,66	1,33	0,84	0,53
Institutions	1,85	1,46	1,04	0,60
Part de la différence attribuable à				
chercheurs	0,47	0,48	0,45	0,47
Institutions	0,53	0,52	0,55	0,53

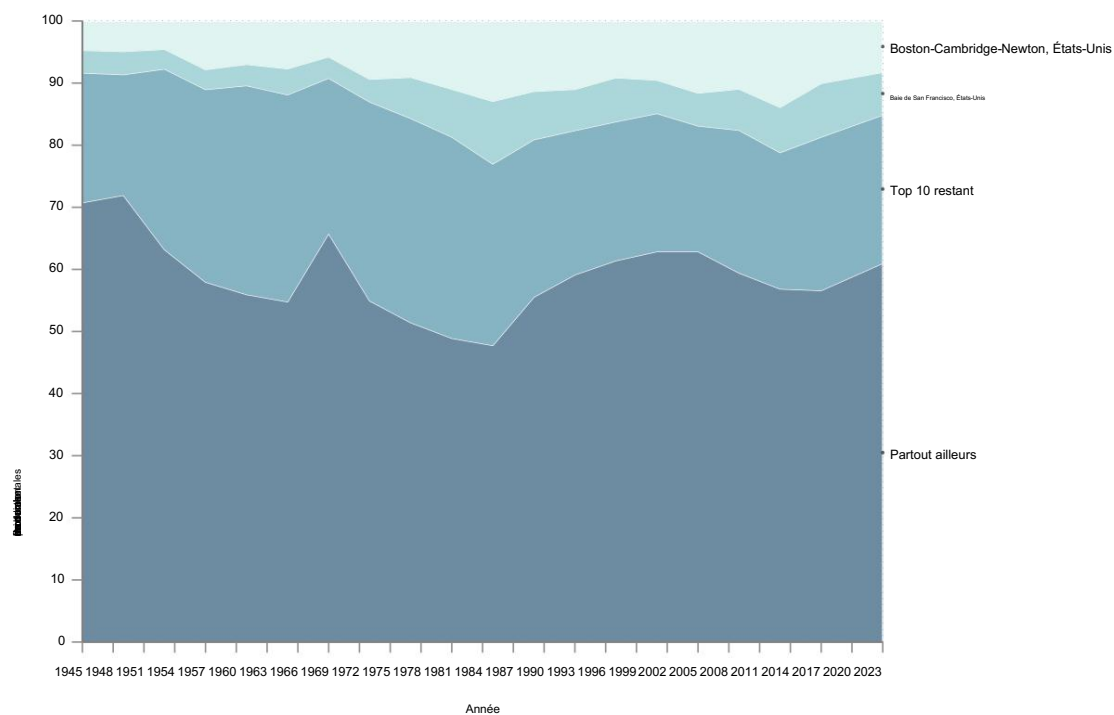
Notes : Le panneau A présente la décomposition de la variance des résultats de recherche entre les établissements. La première partie présente la variance globale et ses composantes, tandis que la seconde illustre la part de variance qui serait réduite si les effets du chercheur ou de l'institution étaient égaux. Le panneau B décompose la différence de productivité des chercheurs entre les groupes de percentiles supérieurs et inférieurs. La première partie présente la différence absolue de productivité, tandis que la seconde partie montre la part de cette différence attribuable aux chercheurs individuels par rapport aux institutions. L'échantillon comprend tous les auteurs (ayant déménagé et n'ayant pas déménagé) de 1945 à 2023. Nous réduisons le bruit dans nos estimations en nous limitant aux institutions comptant au moins 100 auteurs par an (Panneau A) et au moins 100 auteurs et 10 les plus grands déménageurs (panneau B).

Figure 1 : Tendances nationales en matière de production scientifique fondamentale, 1945-2023



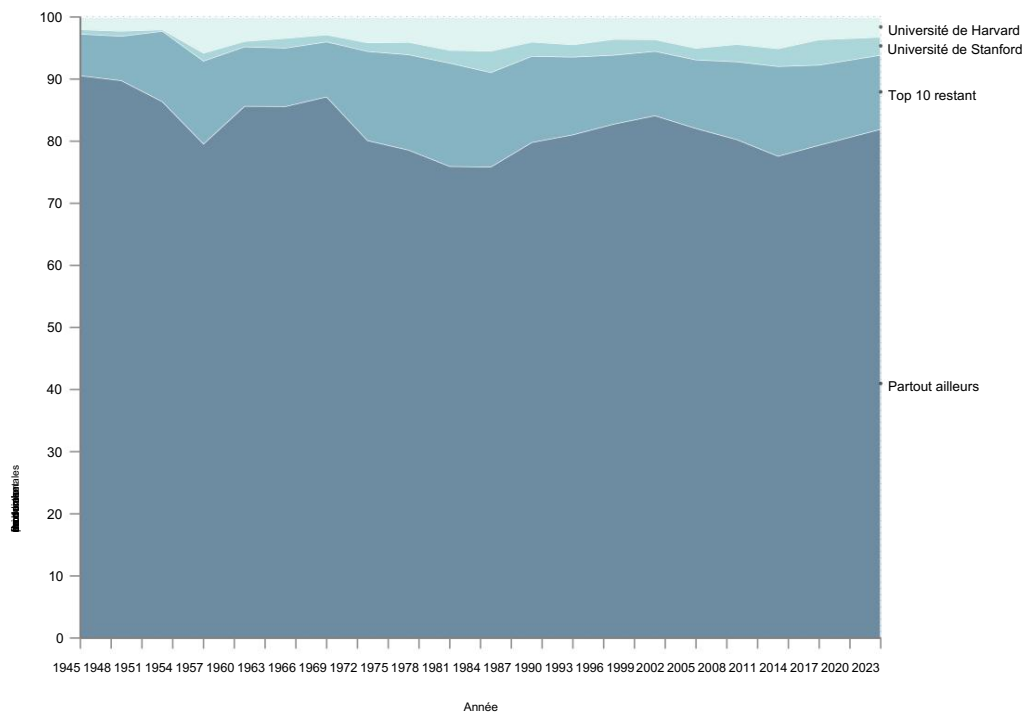
Français : Remarques : La figure montre les tendances au niveau des pays dans la part de la production totale de recherche en sciences fondamentales de 1945 à 2023 en utilisant des tranches de trois ans et des publications dans un ensemble de 15 principales revues de sciences de la vie (Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal). Notre mesure de la productivité a été ajustée en fonction du nombre d'auteurs, du nombre de citations et du dernier facteur d'impact sur 5 ans de la revue. Nous ne comptons un article dans la productivité d'un scientifique que s'il est le premier ou le dernier auteur mentionné sur une publication. Les parts sont présentées séparément pour les deux premiers producteurs (ainsi que pour la Chine), les 10 premiers producteurs restants et partout ailleurs.

Figure 2 : Tendances de la production scientifique fondamentale à l'échelle des villes, 1945-2023



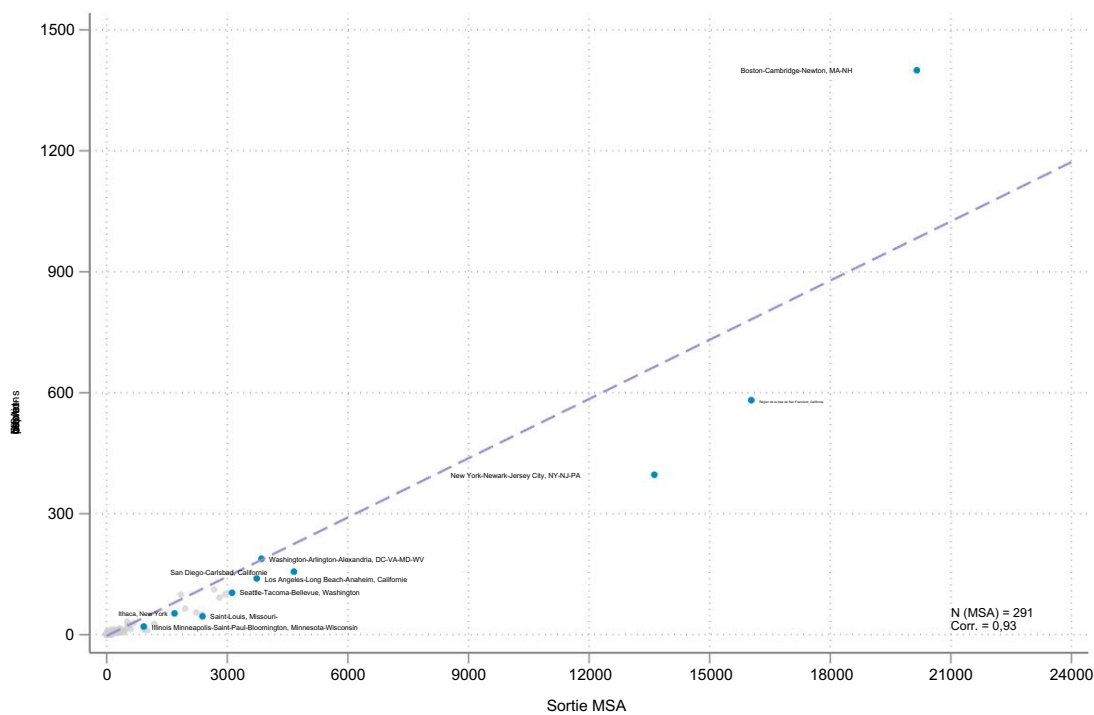
Remarques : La figure montre les tendances à l'échelle des villes en termes de part de la production totale de recherche en sciences fondamentales de 1945 à 2023, en utilisant des tranches de trois ans et des publications dans un ensemble de 15 revues de sciences de la vie de premier plan (Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal). Notre mesure de la productivité a été ajustée en fonction du nombre d'auteurs, du nombre de citations et du dernier facteur d'impact sur 5 ans de la revue. Nous ne comptons un article dans la productivité d'un scientifique que s'il est le premier ou le dernier auteur mentionné dans une publication. Les parts sont présentées séparément pour les deux premiers producteurs (ainsi que pour la Chine), les 10 premiers producteurs restants et partout ailleurs.

Figure 3 : Tendances de la production scientifique fondamentale au niveau des institutions, 1945-2023



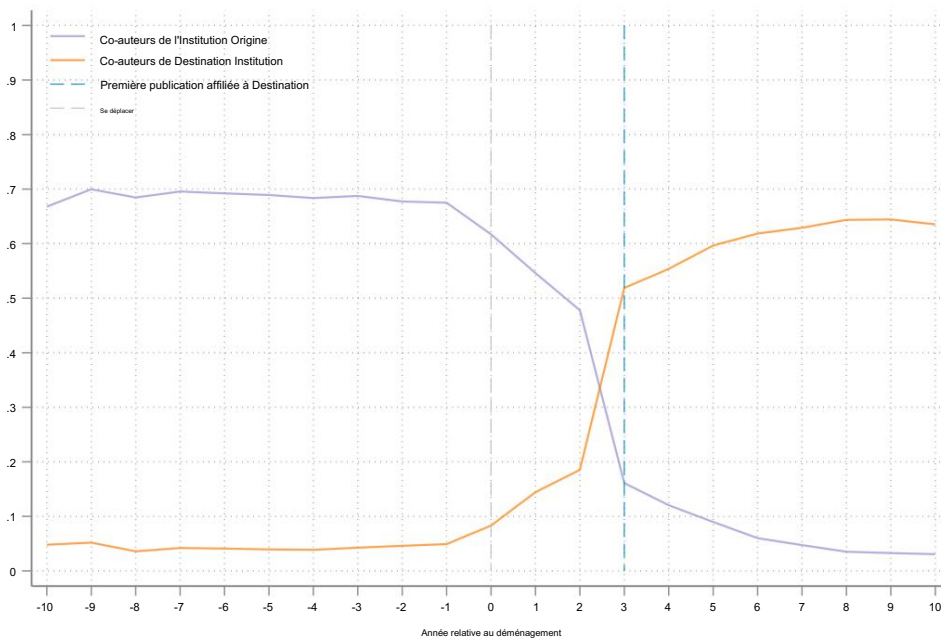
Remarques : La figure montre les tendances au niveau des institutions en termes de part de la production totale de recherche en sciences fondamentales de 1945 à 2023, en utilisant des tranches de trois ans et des publications dans un ensemble de 15 revues de sciences de la vie de premier plan (Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal). Notre mesure de la productivité a été ajustée en fonction du nombre d'auteurs, du nombre de citations et du dernier facteur d'impact sur 5 ans de la revue. Nous ne comptons un article dans la productivité d'un scientifique que s'il est le premier ou le dernier auteur mentionné dans une publication. Les parts sont présentées séparément pour les deux premiers producteurs (ainsi que pour la Chine), les 10 premiers producteurs restants et partout ailleurs.

Figure 4 : Recherches scientifiques fondamentales citées dans les brevets, 1945-2023



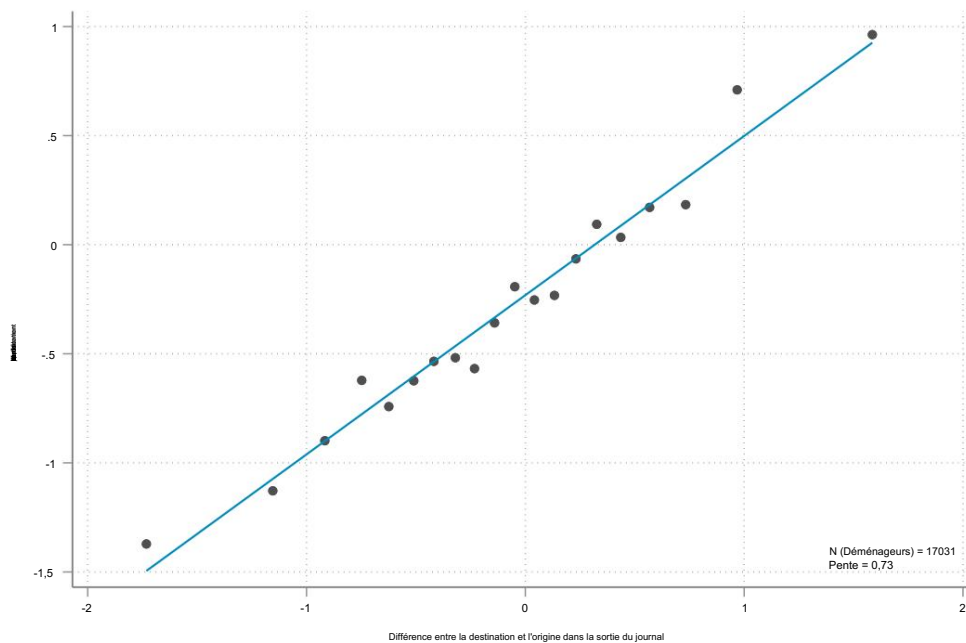
Remarques : La figure présente un diagramme de dispersion du nombre d'articles cités dans les brevets par rapport à la mesure ajustée des citations de la recherche scientifique fondamentale produite au niveau de la MSA. Environ 15,5 % des Les articles de notre échantillon ont déjà été cités dans un brevet. Notre échantillon comprend N = 291 MSA.

Figure 5 : Part des coauteurs en sciences fondamentales de la vie provenant de l'institution d'origine par rapport à l'institution de destination



Remarque : La figure représente la part des coauteurs affiliés à l'établissement d'origine et de destination du candidat au fil du temps. L'année de transfert correspond mécaniquement aux trois années précédant le premier transfert. publication au lieu de destination du déménageur.

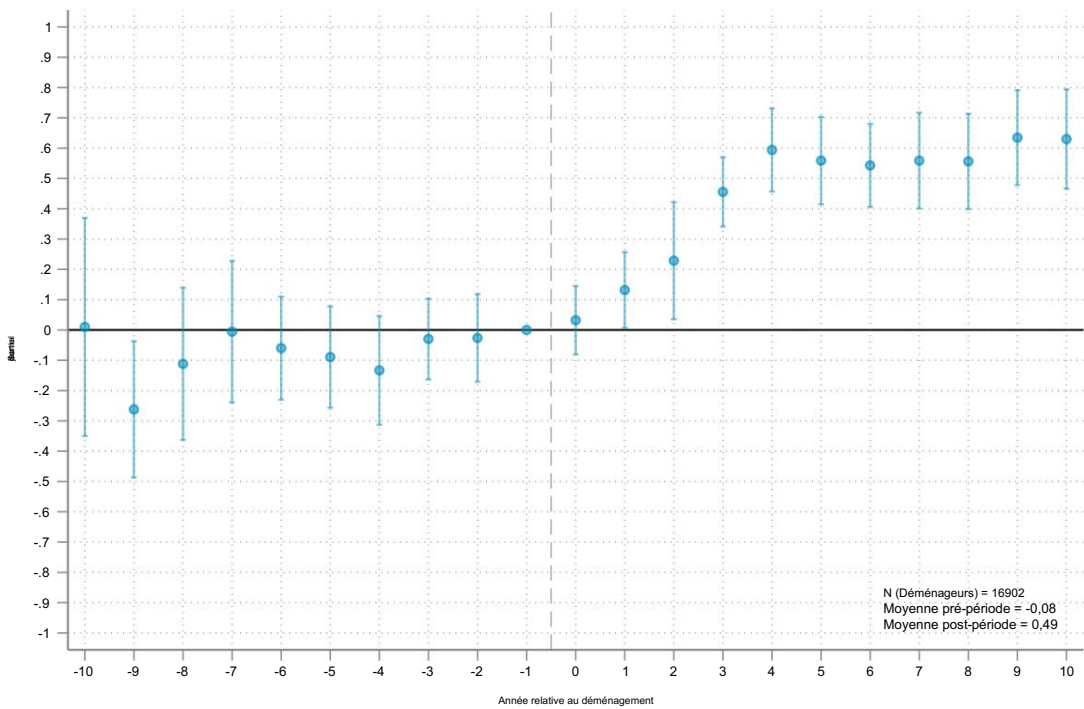
Figure 6 : Évolution de la sortie du journal en fonction de la taille du déplacement



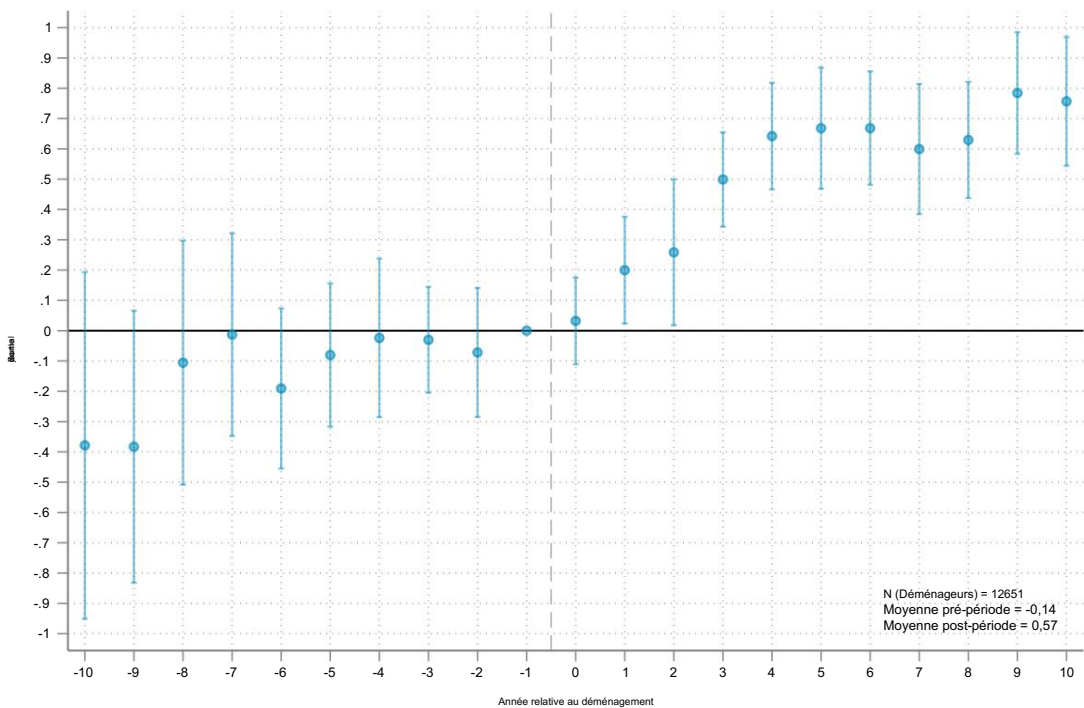
Remarques : La figure illustre l'évolution de la production logarithmique en fonction de la taille du mouvement d'un acteur. La taille du mouvement est définie comme la différence de production moyenne entre l'établissement d'origine et l'établissement de destination d'un acteur.

Figure 7 : Changement de sortie en réponse au déplacement

(a) 1945-2023



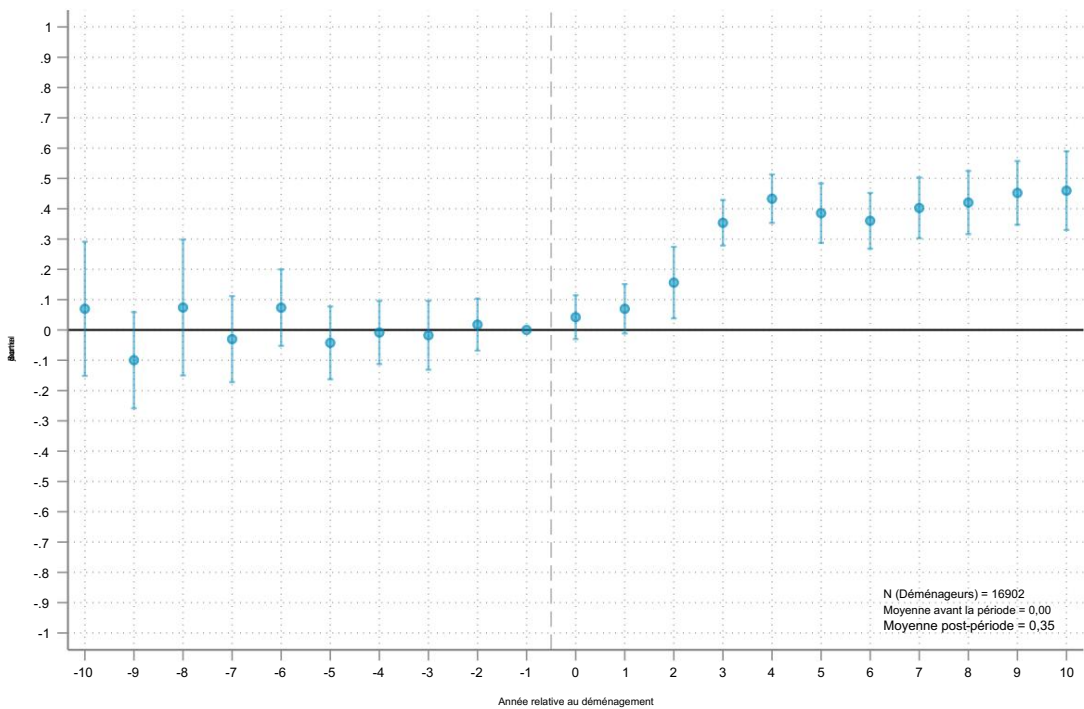
(b) 1995-2023



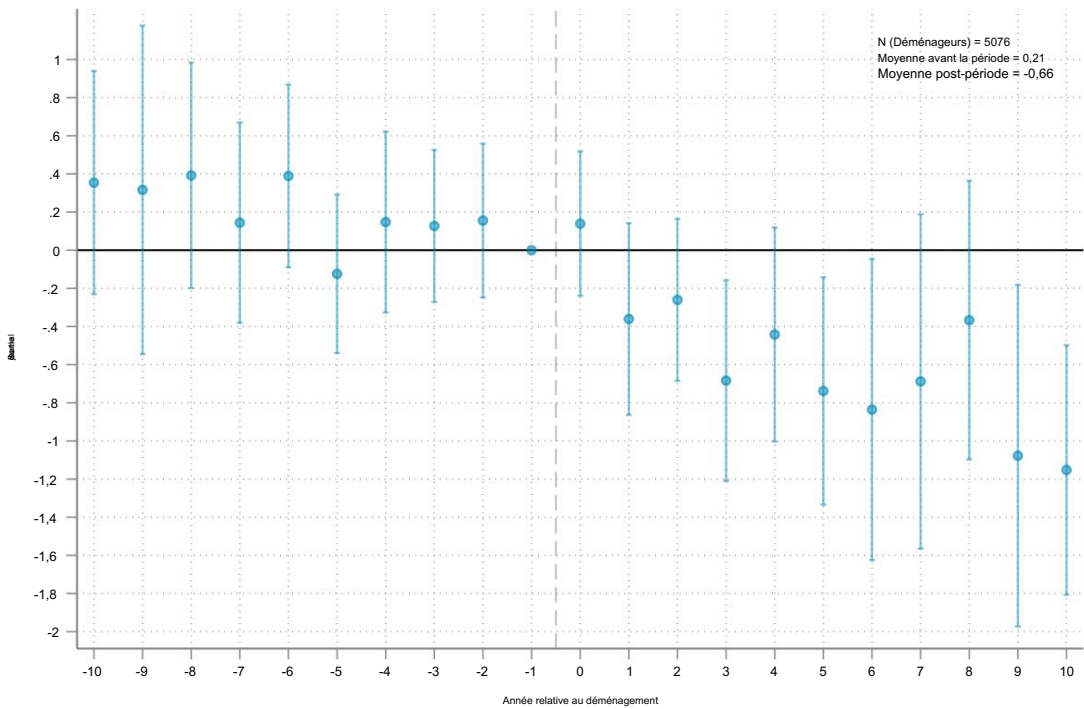
Remarques : La figure montre les coefficients estimés à partir de l'équation 2. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est la sortie logarithmique et nous contrôlons un ensemble d'années, d'auteurs, et des effets fixes relatifs à l'année dans notre spécification. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur les scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023. Le panneau (a) présente les résultats pour les déménageurs publiant à tout moment entre 1945 et 2023 et le panneau (b) montre les résultats pour les déménageurs publiant entre 1995 et 2023.

Figure 8 : Impact des chercheurs vedettes

(a) Taille de déplacement définie par la sortie de Star Researcher



(b) Rôle du départ de l'étoile sur la production des coauteurs



Notes : Le panneau (a) présente les coefficients estimés à partir de l'équation 2 en utilisant la définition de la taille du mouvement telle que définie par la productivité des étoiles à l'emplacement. Le panneau (b) présente les coefficients d'un Événement canonique – étude qui suit l'évolution de la productivité logarithmique des coauteurs de l'étoile au cours des années entourant le mouvement. Pour les deux, le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est le logarithme de la productivité, et notre spécification prend en compte un ensemble d'effets fixes liés à l'année, à l'auteur et à l'année relative. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur des scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023.

Appendice

A. Pièces supplémentaires

Tableau A1 de l'annexe : Tailles des échantillons de revues

Échantillon	Journal	% Fondamental	# Fondamental
SNC	Science	37,10	39 454
	Nature	64,36	64 034
	Cellule	93,89	16 163
Revue de premier plan en sciences de la vie	Neurone	93,14	8 464
	Génétique de la nature	90,59	5 775
	Médecine naturelle	57,11	4 625
	Biotechnologie de la nature	43,14	3 445
	Neurosciences de la nature	87,08	4 252
	Biologie cellulaire naturelle	84,97	3 182
	Nature Chimie Biologie	78,45	2 235
	Cellule souche	89,24	1 642
Autres revues des sciences de la vie	PLOS One	81,44	210 334
	Journal de biochimie	95,07	171 693
	Oncogène	97,92	15 604
	Le Journal de la FASEB	69,06	12 718

Remarques : Le tableau présente la part de la recherche en sciences fondamentales publiée dans chaque revue. Notre échantillon complet pour chaque revue comprend tous les articles et essais cliniques publiés dans OpenAlex entre 1945 et 2023.

Annexe Tableau A2 : Robustesse de la pondération des mesures

	Niveau national	Niveau municipal	Niveau institutionnel
Résultats de la recherche			
Citations d'articles dans les brevets non pondérées et pondérées par le facteur d'impact	0,9934	0,9495	0,9300
	0,9953	0,9042	0,8667
Toutes les citations vs. Citations de la première page	0,9996	0,9967	0,9907
Toutes les citations vs. Citations du corps	0,9993	0,9933	0,9810

Remarques : Le tableau montre la corrélation entre différentes mesures de la part de la recherche scientifique fondamentale produite au niveau du pays, de la ville et de l'institution pour les articles publiés dans un ensemble de 15 premiers Revues des sciences de la vie (Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS) One, Oncogene, Journal of Biological Chemistry et le FASEB Journal).

Tableau A3 de l'annexe : Corrélation entre la production et les citations d'articles dans les brevets

	Niveau de l'État	Niveau de la ville	Niveau de l'institution
Productivité de la recherche et citations d'articles dans les brevets			
Facteur d'impact et pondération des citations par rapport à toutes les citations	0,9950	0,9916	0,9598
Facteur d'impact et pondération des citations par rapport aux citations de la première page	0,9948	0,9929	0,9682
Facteur d'impact et pondération des citations par rapport aux citations du corps du texte	0,9939	0,9857	0,9337

Remarques : Le tableau montre la corrélation entre notre principale mesure de la productivité des sciences fondamentales et les différentes versions du nombre de citations d'articles dans les brevets au niveau de l'État, de la ville et de l'institution. pour les articles publiés dans un ensemble de 15 revues de sciences de la vie parmi les plus importantes (Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neurosciences, Neuron, Cellule souche, PLOS One, Oncogène, Journal of Biological Chemistry et le FASEB Journal).

Annexe Tableau A4 : Les plus grandes institutions américaines par production annuelle moyenne et part de la production mondiale, 2015–2023

Rang Institution		Part de la production annuelle dans la production mondiale (%)	
1	Université de Harvard	678,78	3,63
2	Université de Stanford	632,12	3,38
3	Institut de technologie du Massachusetts	389,78	2,09
4	Université de Californie à San Francisco	349,68	1,87
5	Général de Masse Brigham	291,98	1,56
6	Instituts nationaux de la santé Centre de	272,94	1,46
7	cancérologie Memorial Sloan Kettering Université Johns	223,44	1,20
8	Hopkins Université Yale Université de	222,75	1,19
9	Californie à San Diego	219,00	1,17
10	Université Columbia Université de Pennsylvanie	218,64	1,17
11	Institut du cancer Dana-Farber	217,52	1,16
12	Université de Californie à Berkeley	193,20	1,03
13	Université Washington à Saint-Louis	189,01	1,01
14	Université de Washington École de médecine	186,16	1,00
15	Icahn du Mont Sinai Université du Texas Centre	178,82	0,96
16	médical Southwestern NYU Langone	177,00	0,95
17	Health Université Cornell Centre de génomique de New	168,34	0,90
18	York Université du Texas Centre de cancérologie MD Anderson Université de	154,93	0,83
19	Californie à Los Angeles	146,74	0,79
20	Mammoth Biosciences	145,96	0,78
21	Université Rockefeller Université du	140,14	0,75
22	Michigan–Ann Arbor Université de Caroline du Nord à Chapel Hill Université	130,37	0,70
23	Duke Université de Princeton Université de Chicago	116,44	0,62
24	Institut de technologie de	114,05	0,08
25	Californie VA St. Louis Health	112,19	0,60
26	Care System Baylor College of Medicine Hôpital	105,21	0,56
27	pour enfants de Boston Université du Minnesota Université	100,49	0,54
28	Northwestern Université	99,16	0,53
29	de Pittsburgh Institut de	95,87	0,51
30	recherche Scripps Hôpital de	88,14	0,47
31	recherche pour enfants St. Jude Institut Salk	86,62	0,46
32	d'études biologiques Université de Californie du	83,23	0,33
33	Sud Université Vanderbilt Université du	75,96	0,41
34	Texas à Austin Université du	72,28	0,39
35	Massachusetts École de médecine	71,24	0,38
36	Chan Campus de recherche Janelia	64,11	0,34
37	Institut d'immunologie de La Jolla	61,65	0,33
38	Centre de cancérologie Fred Hutch	61,23	0,33
39	Université Google Emory Institut Allen	60,18	0,32
40		59,65	0,32
41		55,60	0,30
42		53,44	0,29
43		52,72	0,28
44		51,83	0,28
45		48,73	0,26
46		46,75	0,25
47		45,69	0,24
48		45,12	0,24
49		41,78	0,22
50		41,62	0,22

Notes : Le tableau présente les 50 principaux établissements selon la production annuelle moyenne de recherche en sciences fondamentales dans les revues scientifiques. La production de recherche est mesurée par le nombre total de publications. Entre 2015 et 2023. La part de la production mondiale est calculée à partir de la production totale sur la période 2015-2023. Les revues incluses sont : Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal.

Annexe Tableau A5 : Les plus grandes institutions internationales par production annuelle moyenne et part de Production mondiale, 2015–2023

Rang	Institution	Pays	Part de la production annuelle dans la production mondiale (%)	
1	Académie chinoise des sciences	Chine	405,31	2.17
2	Université d'Oxford	Royaume-Uni	230,76	1.23
3	Université de Cambridge	Royaume-Uni	223,67	1.20
4	Société Max Planck	Allemagne	222,95	1.19
5	Institut Karolinska	Suède	132,44	0,71
6	Institut Weizmann des sciences	Israël	127,17	0,68
7	Association Helmholtz des centres de recherche allemands	Allemagne	125,49	0,67
8	Université de Pékin	Chine	113,63	0,61
9	Institut Wellcome Sanger	Royaume-Uni	107,49	0,58
10	DeepMind	Royaume-Uni	101,32	0,54
11	University College de Londres	Royaume-Uni	95,11	0,51
12	Association Leibniz	Allemagne	93,77	0,50
13	Université de Copenhague	Danemark	85,97	0,46
14	Université du Queensland	Australie	85,44	0,46
15	Institut de génomique intégrative et expérimentale d'Allemagne		84,56	0,11
16	Université de Heidelberg	Allemagne	78,70	0,42
17	Imperial College de Londres	Royaume-Uni	76,80	0,41
18	Université Tsinghua	Chine	75,28	0,40
19	Conseil de la recherche médicale	Royaume-Uni	73,35	0,39
20	Université de Toronto	Canada	66,05	0,35
21	Université de Hong Kong	Hong Kong	63,07	0,34
22	L'Université de Tokyo	Japon	61,08	0,33
23	ETH Zurich	Suisse	60,17	0,32
24	École Polytechnique Fédérale de Lausanne	Suisse	59,37	0,32
25	Centre National de la Recherche Scientifique	France	59,08	0,32
26	Université d'Édimbourg	Royaume-Uni	56,74	0,30
27	King's College de Londres	Royaume-Uni	54,24	0,29
28	Institut Pasteur	France	54,02	0,29
29	L'Institut néerlandais du cancer	Pays-Bas	48,67	0,26
30	Université Sun Yat-sen	Chine	47,87	0,26
31	UNSW Sydney	Australie	47,61	0,25
32	Université technique du Danemark	Danemark	45,98	0,25
33	Inserm	France	44,42	0,24
34	Charité - Universitätsmedizin Berlin	Allemagne	44,07	0,24
35	L'Institut Francis Crick	Royaume-Uni	43,94	0,24
36	Université du Zhejiang	Chine	43,43	0,23
37	Académie royale des arts et des sciences des Pays-Bas	Pays-Bas	43,05	0,23
38	Institut des sciences et technologies de Barcelone	Espagne	43,05	0,23
39	Université McGill	Canada	42,66	0,23
40	KU Leuven	Belgique	42,20	0,23
41	RIKEN	Japon	41,51	0,22
42	Laboratoire européen de biologie moléculaire	Allemagne	41,30	0,23
43	Université de Lund	Suède	40,94	0,22
44	Collège médical de l'Union de Pékin	Chine	40,84	0,22
45	LanzaTech	Nouvelle-Zélande	39,88	0,03
46	Université d'Helsinki	Finlande	38,02	0,20
47	Université de Zurich	Suisse	37,78	0,20
48	Université de Kyoto	Japon	37,19	0,20
49	Université Louis-et-Maximilien de Munich	Allemagne	36,12	0,19
50	Université de Westlake	Chine	35,96	0,12

Notes : Le tableau présente les 50 principales institutions internationales selon la production annuelle moyenne de recherche en sciences fondamentales dans les revues de sciences de la vie. La production de recherche est mesurée par le nombre total de Publications entre 2015 et 2023. La part de la production mondiale est calculée à partir de la production totale sur la période 2015-2023. Les revues concernées sont : Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Médecine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal.

Tableau A6 de l'annexe : Institutions américaines classées selon la production moyenne des chercheurs, 2015-2023

Rang	Institution	Production du chercheur
1	Institut du cancer Dana-Farber	2,81
2	Institut Allen	2,40
3	Institut de technologie du Massachusetts	2,33
4	Centre de cancérologie Memorial Sloan Kettering	2,24
5	Université de Stanford	1,86
6	Instituts Gladstone	1,77
7	Institut Salk d'études biologiques	1,73
8	Novartis (États-Unis)	1,72
9	École de médecine Icahn du Mont Sinai	1,69
10	Université de Harvard	1,64
11	Général de Masse Brigham	1,52
12	Centre de cancérologie MD Anderson de l'Université du Texas	1,49
13	Centre de cancérologie Fred Hutch	1,42
14	Université de Princeton Institut	1,42
15	de technologie de Californie Laboratoire de Cold	1,42
16	Spring Harbor Centre médical Southwestern	1,42
17	de l'Université du Texas Université Rockefeller Université de Californie,	1,35
18	Berkeley Université de Californie,	1,34
19	San Francisco Campus de recherche Janelia	1,33
20	Laboratoire Jackson Hôpital pour enfants de Boston	1,31
21	Université Columbia Centre médical	1,26
22	national City Of Hope NYU	1,20
23	Langone Health Hôpital de recherche	1,15
24	pour enfants St. Jude	1,15
25	Université de Pennsylvanie Université Washington à	1,13
26	Saint-Louis Université Yale	1,11
27	Université de Californie, San Diego Université du	1,08
28	Massachusetts École de médecine	1,06
29	Chan	1,06
30		1,02
31		1,00
32		0,97
33	Centre médical Cedars-Sinai	0,95
34	Université de Chicago	0,94
35	Institut de recherche Scripps	0,91
36	Université de Washington	0,85
37	Université Cornell	0,85
38	MSD (États-Unis)	0,85
39	Centre médical Beth Israel Deaconess	0,84
40	L'Université du Texas à Austin	0,81
41	Université Johns Hopkins	0,79
42	Université Carnegie Mellon	0,76
43	La branche médicale de l'Université du Texas à Galveston	0,76
44	Université de Californie à Santa Cruz	0,73
45	Université de Californie du Sud	0,73
46	Collège de médecine Baylor	0,73
47	Instituts nationaux de la santé	0,69
48	Laboratoire national Lawrence Berkeley	0,67
49	Collège de médecine Albert Einstein	0,66
50	Université de Californie à Los Angeles	0,65

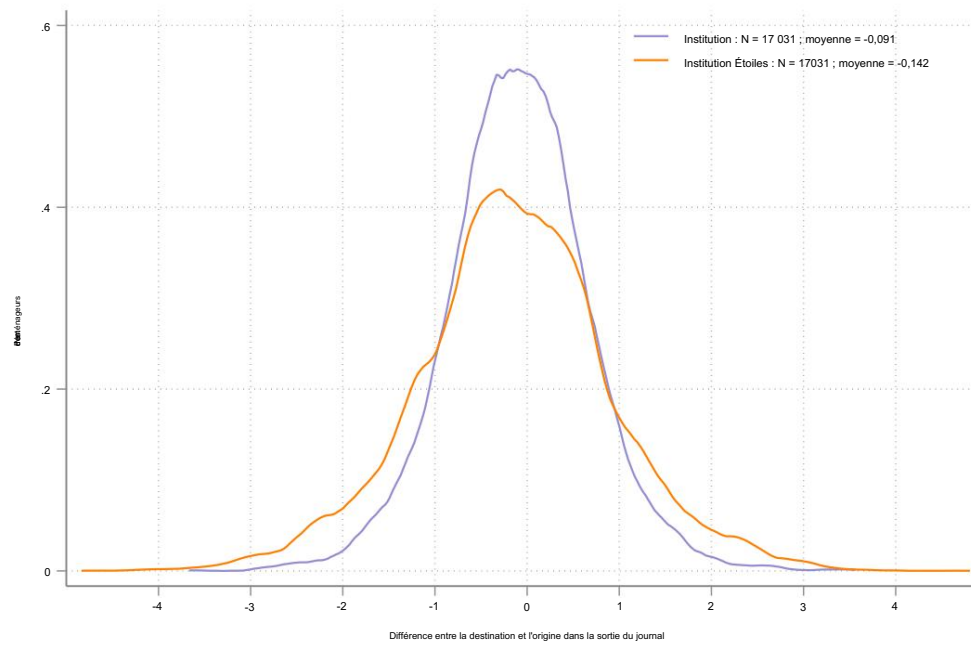
Notes : Le tableau présente les 50 principales institutions selon la productivité moyenne par auteur-année en recherche fondamentale dans les revues de sciences de la vie. La mesure de la productivité est pondérée par le nombre de chercheurs de chaque institution. Les revues concernées sont : Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal entre 2015 et 2023.

Annexe Tableau A7 : Institutions internationales classées selon la production moyenne des chercheurs, 2015-2023

Rang	Institution	Pays	Production du chercheur
1	Institut Wellcome Sanger	Royaume-Uni	2,66
2	Université de Trente	Italie	2,24
3	Académie royale des arts et des sciences des Pays-Bas	Pays-Bas	2,24
4	Centre médical universitaire	Pays-Bas	2,21
5	L'Institut néerlandais du cancer	Pays-Bas	1,96
6	Académie autrichienne des sciences	Autriche	1,63
7	Institut Weizmann des sciences	Israël	1,63
8	Centre de cancérologie Princess Margaret	Canada	1,60
9	Université de médecine de Chongqing	Chine	1,57
10	Institut des sciences et technologies de Barcelone	Espagne	1,56
11	Université technique du Danemark	Danemark	1,50
12	Recherche sur le cancer au Royaume-Uni	Royaume-Uni	1,49
13	École Polytechnique Fédérale de Lausanne	Suisse	1,36
14	Institut royal de technologie KTH	Suède	1,32
15	Novartis	Suisse	1,29
16	Institut Friedrich Miescher	Suisse	1,29
17	Université de Hong Kong	Hong Kong	1,26
18	Institut Pasteur	France	1,19
19	Conseil de recherche biomédicale	Singapour	1,17
20	Université d'Islande	Islande	1,16
21	Université de Cambridge	Royaume-Uni	1,13
22	Conseil de la recherche médicale	Royaume-Uni	1,09
23	Université d'Oxford	Royaume-Uni	1,05
24	Roche (Suisse)	Suisse	1,05
25	Centre médical universitaire de Fribourg	Allemagne	1,02
26	L'Institut Francis Crick	Royaume-Uni	1,00
27	Université du Queensland	Australie	0,97
28	Université Tsinghua	Chine	0,95
29	Institut de recherche sur le cancer	Royaume-Uni	0,94
30	Association Leibniz	Allemagne	0,88
31	UNSW Sydney	Australie	0,88
32	Société Max Planck	Allemagne	0,87
33	Université de Pékin	Chine	0,84
34	Association Helmholtz des centres de recherche allemands	Allemagne	0,84
35	Hôpital pour enfants malades	Canada	0,82
36	Université de Stockholm	Suède	0,81
37	Institut Karolinska	Suède	0,78
38	King's College de Londres	Royaume-Uni	0,77
39	Réseau universitaire de santé	Canada	0,76
40	Institut Curie	France	0,76
41	Technion – Institut israélien de technologie	Israël	0,76
42	ETH Zurich	Suisse	0,75
43	Institut de recherche médicale QIMR Berghofer	Australie	0,71
44	Institut de recherche médicale Walter et Eliza Hall	Australie	0,69
45	Université Keio	Japon	0,68
46	Institut de recherche médicale Garvan	Australie	0,67
47	Académie chinoise des sciences	Chine	0,67
48	Institut coréen avancé des sciences et technologies, Corée du Sud		0,67
49	RIKEN	Japon	0,66
50	Université de Dundee	Royaume-Uni	0,66

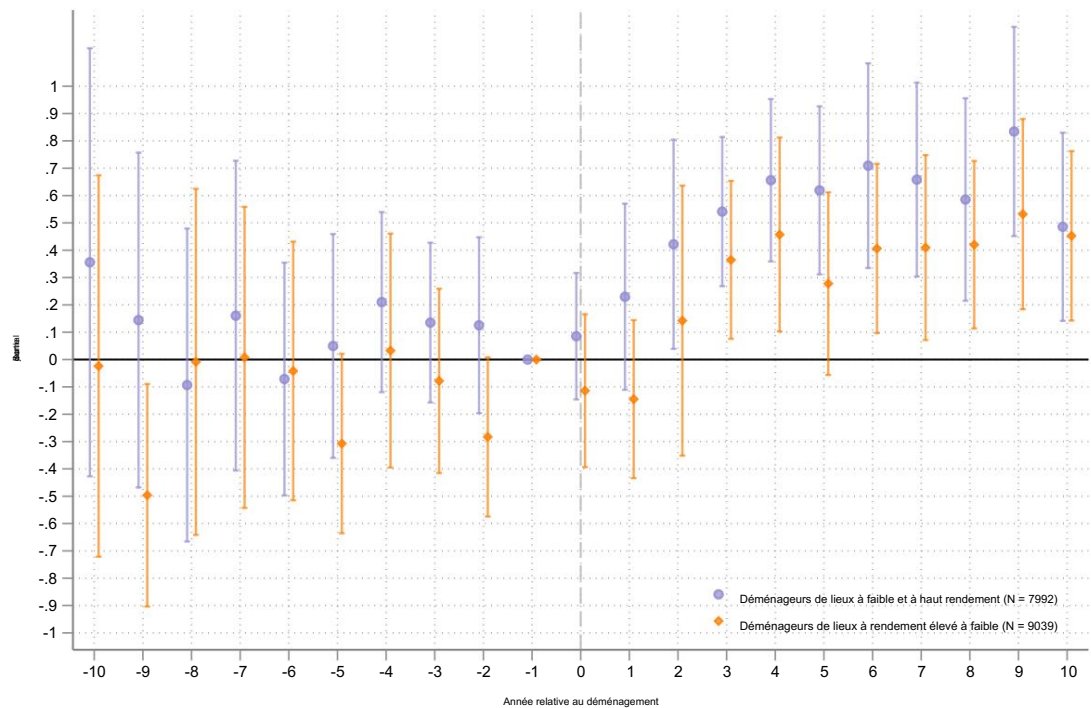
Notes : Le tableau présente les 50 principales institutions internationales selon la productivité moyenne des chercheurs en recherche fondamentale dans les revues spécialisées en sciences de la vie. La mesure de la productivité est pondérée par le nombre de chercheurs dans chaque institution. Les revues concernées sont : Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry et le FASEB Journal entre 2015 et 2023.

Annexe Figure A1 : Répartition des



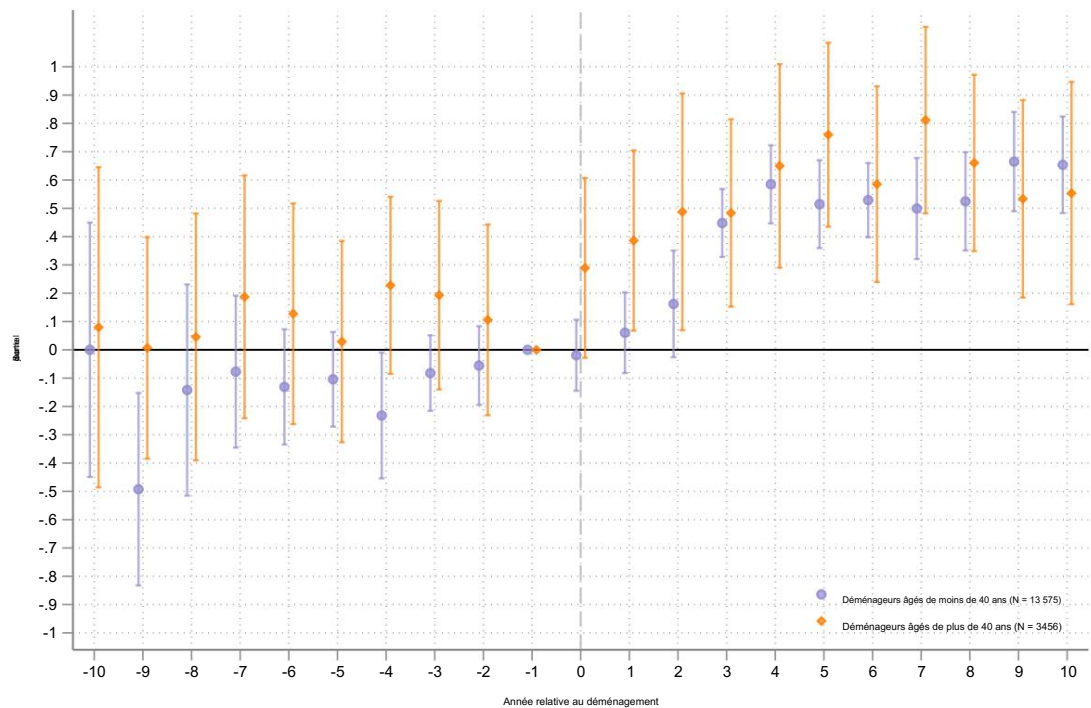
Remarques : La figure montre la distribution des données calculées à différents niveaux pour les scientifiques en mouvement. Elle mesure ici l'ampleur du mouvement et est construite comme la différence moyenne de productivité. Entre le lieu de destination et le lieu d'origine. Les distributions au niveau de l'établissement sont indiquées en violet et celles des étoiles au niveau de l'établissement sont indiquées en orange.

Annexe Figure A2 : Modification de la sortie en fonction de la taille du déplacement



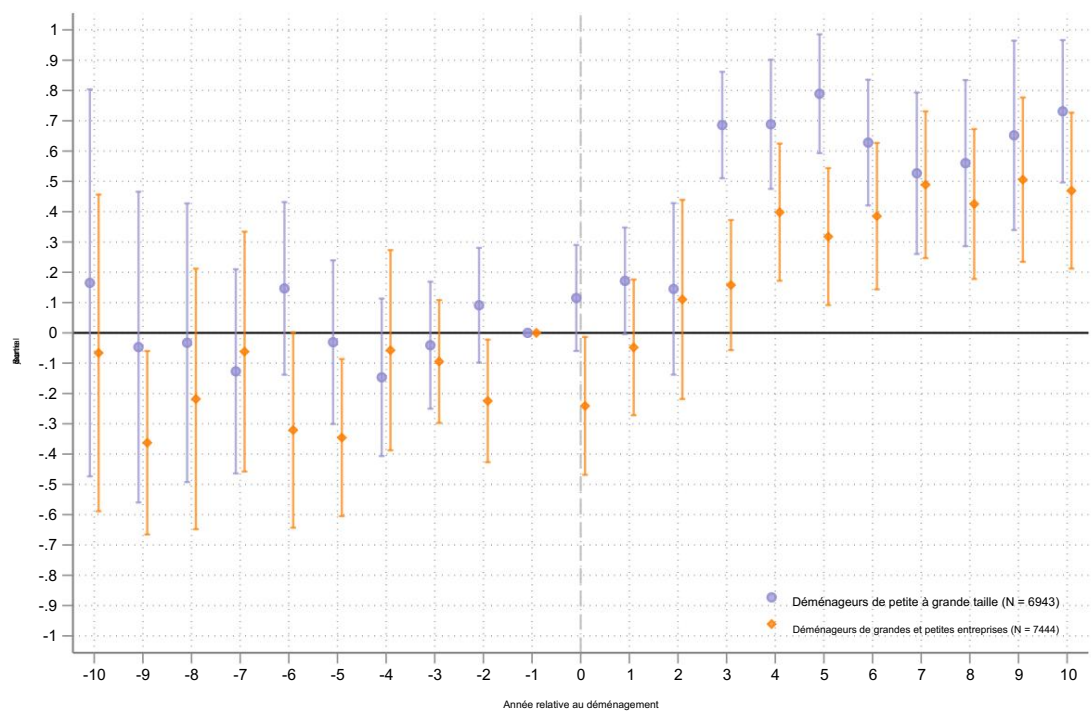
Remarques : La figure présente les coefficients estimés à partir de l'équation 2 pour les personnes ayant déménagé, en fonction de la taille du déménagement, pour la période comprise entre 1945 et 2023. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est le logarithme de la productivité, et notre spécification prend en compte un ensemble d'effets fixes liés à l'année, à l'auteur et à l'année relative. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur des scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023.

Annexe Figure A3 : Variation de la production en fonction de l'âge



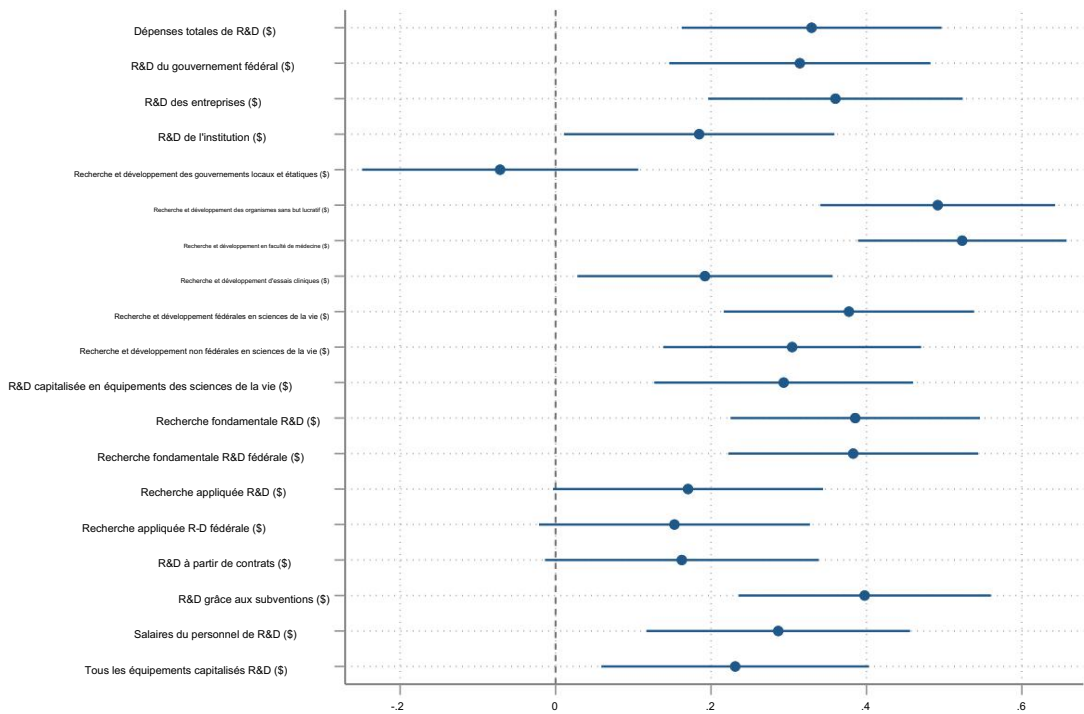
Notes : La figure montre les coefficients estimés à partir de l'équation 2 pour les personnes ayant déménagé, répartis par âge, pour la période 1945-2023. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est le logarithme de la productivité, et notre spécification prend en compte un ensemble d'effets fixes liés à l'année, à l'auteur et à l'année relative. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur des scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023.

Annexe Figure A4 : Variation de la production en fonction de la taille de la ville



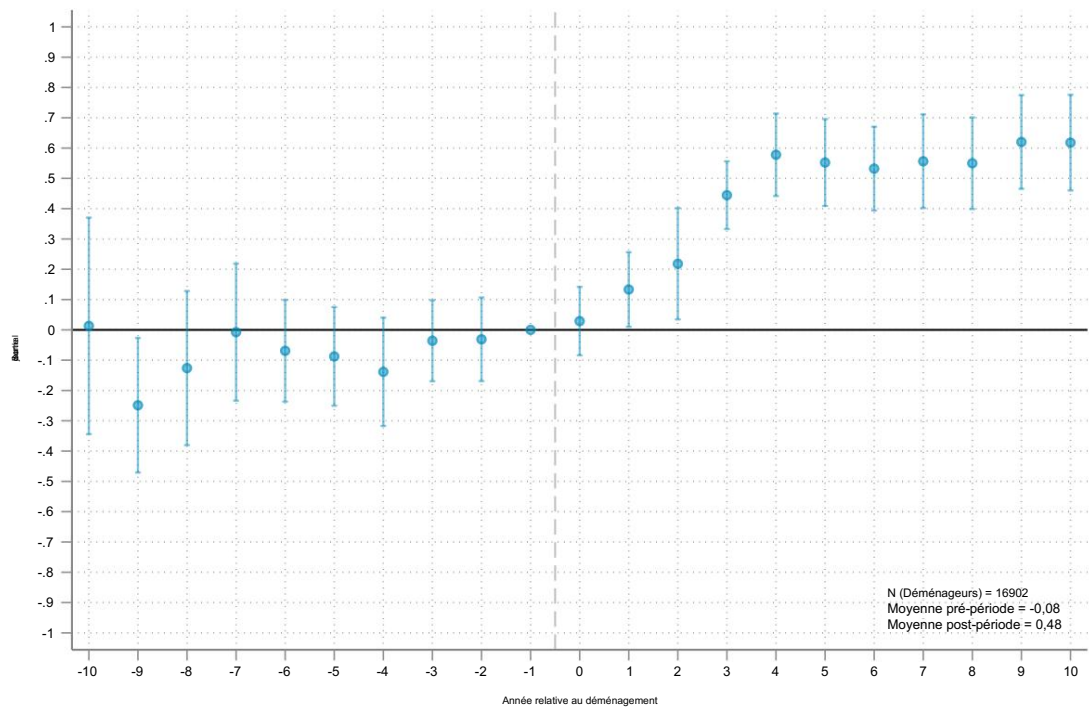
Remarques : La figure présente les coefficients estimés à partir de l'équation 2 pour les personnes ayant déménagé, en fonction de l'évolution de la taille des villes pour la période 1945-2023. Le coefficient pour l'année relative -1 est normalisé à 0. Lors du calcul de la taille de la ville, nous ignorons les chercheurs situés aux lieux d'origine et de destination du migrant. La variable dépendante est le logarithme de la productivité, et nous contrôlons un ensemble d'années, d'auteurs et Effets fixes relatifs à l'année dans notre spécification. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur les scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023.

Annexe Figure A5 : Corrélat des effets institutionnels



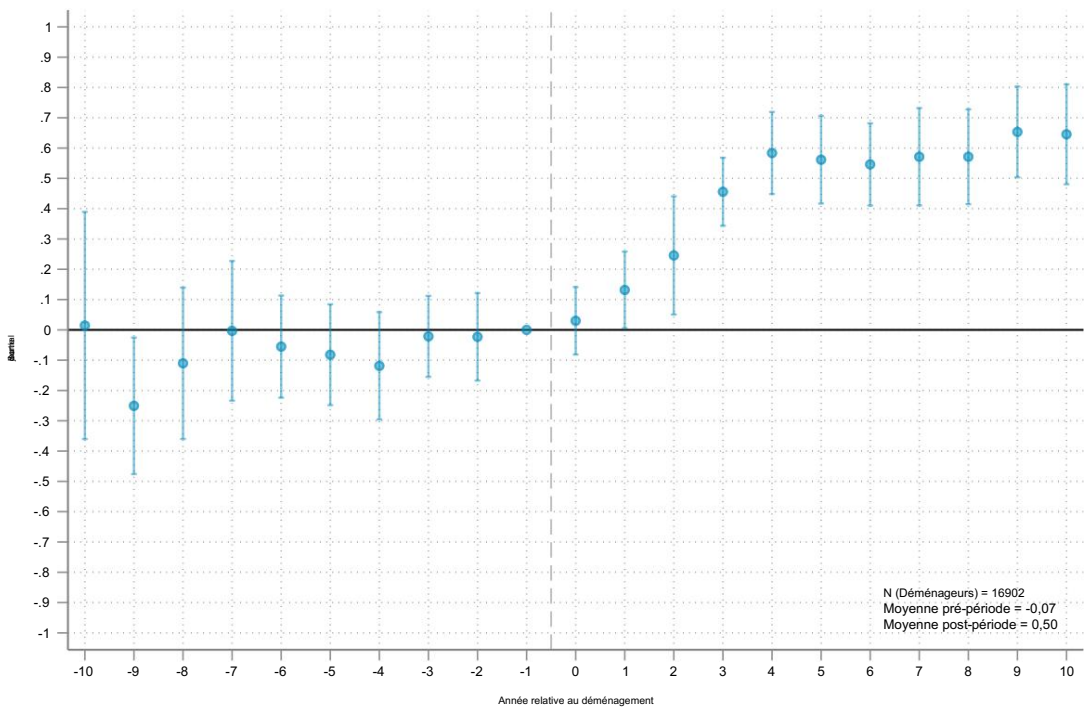
Remarques : La figure montre les résultats de la régression des effets estimés des institutions sur divers facteurs locaux issus de l'enquête NSF Higher Education Research and Development (HERD).

Annexe Figure A6 : Spécification incluant les effets fixes sur le terrain



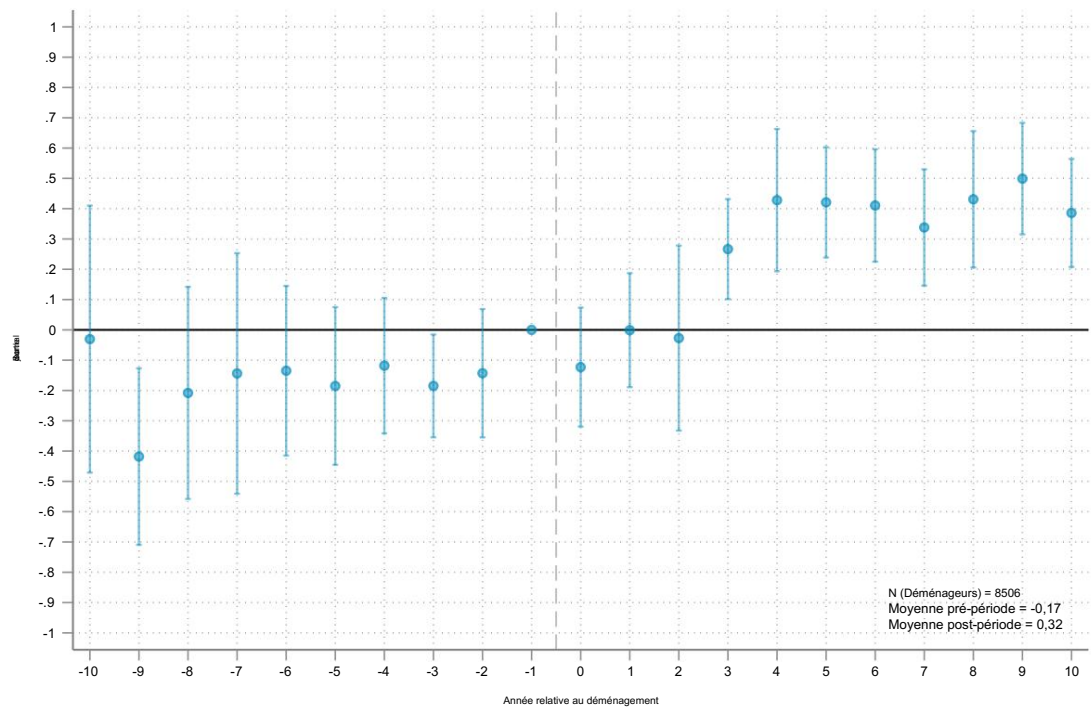
Notes : La figure présente les coefficients estimés à partir de l'équation 2, en incluant les effets fixes des sous-domaines, pour la période 1945-2023. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est le logarithme de la productivité et nous contrôlons un ensemble d'effets fixes liés à l'année, à l'auteur et à l'année relative dans notre spécification. Les erreurs types sont regroupées au niveau de l'établissement. Nous se concentrer sur les scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023.

Annexe Figure A7 : Spécification sans effets fixes relatifs à l'année



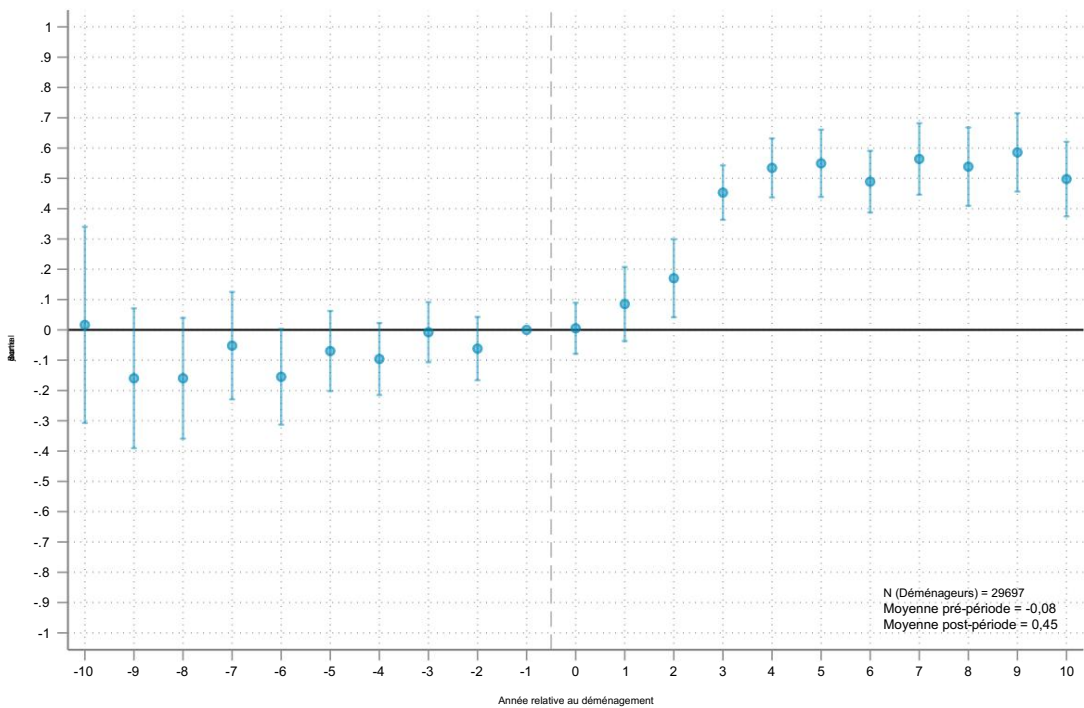
Notes : La figure montre les coefficients estimés à partir de l'équation 2 sans inclure les effets fixes de l'année relative pour la période 1945-2023. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est la productivité logarithmique et nous contrôlons un ensemble d'effets fixes année et auteur dans notre spécification. Les erreurs types sont regroupées au niveau de l'établissement. se concentrer sur les scientifiques qui n'ont déménagé qu'une seule fois entre 1945 et 2023.

Annexe Figure A8 : Premier et dernier auteurs



Notes : La figure présente les coefficients estimés à partir de l'équation 2 pour l'échantillon des premiers et derniers auteurs. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante est logarithmique, productivité et nous contrôlons un ensemble d'effets fixes d'année, d'auteur et d'année relative dans notre spécification. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur les scientifiques ayant uniquement déménagé une fois entre 1945 et 2023.

Annexe Figure A9 : Tous les auteurs



Notes : La figure montre les coefficients estimés à partir de l'équation 2 pour l'échantillon de tous les auteurs publiant dans les 15 premières revues. Le coefficient pour l'année relative -1 est normalisé à 0. La variable dépendante La variable est le logarithme de la productivité, et notre spécification prend en compte un ensemble d'effets fixes liés à l'année, à l'auteur et à l'année relative. Les erreurs types sont regroupées au niveau de l'établissement. Nous nous concentrons sur les scientifiques. qui n'a déménagé qu'une seule fois entre 1945 et 2023.

B. Pondération des publications en fonction de leur impact

Chaque observation dans nos données représente un auteur publiant un article dans une revue. Nous mettons en œuvre une méthode à deux niveaux. approche de pondération qui tient compte à la fois du prestige global de la revue et de l'influence de l'article spécifique lors de l'évaluation des résultats de recherche d'un auteur.

En laissant désigner les revues et désigner les articles, définir

- être le nombre brut d'articles dans notre échantillon
- être le nombre brut d'articles dans la revue
- être le dernier facteur d'impact sur 5 ans de la revue
- être le nombre annuel moyen de citations pour un article dans une revue
- être le nombre d'auteurs pour l'article

Nous ajustons d'abord la distribution de la représentation des revues dans notre échantillon en fonction de leurs facteurs d'impact. Cette redistribution donne lieu à un échantillon qui déplace le poids vers les revues ayant des facteurs d'impact plus élevés, quel que soit leur nombre initial d'articles. Cela permet de tenir compte du fait que les revues à fort impact peuvent publier moins souvent que les revues à faible impact. Par exemple, PLOS One et le Journal of Biological Chemistry ont des tailles d'échantillons de revues brutes qui représentent chacune 31 % à 37 % de notre échantillon complet, mais sont relativement petites facteurs d'impact de 3,8 et 4,8 respectivement. En revanche, les deux revues ayant les facteurs d'impact les plus élevés de notre échantillon, Nature Medicine et Nature ont des facteurs d'impact supérieurs à 60 mais ne sont représentés que par 0,8 % et 10 % dans nos données brutes. échantillon. Ainsi, pour la revue, nous repondérons son nombre d'articles comme

$$= \frac{\sum}{\sum} \sum$$

où les sommes sont sur les revues de notre échantillon. Remarque : $\sum = \sum =$

Pour intégrer le nombre de citations d'un article, nous calculons d'abord un poids de citation relatif qui est le nombre moyen de citations annuelles d'un article par rapport au nombre moyen total de citations annuelles dans notre échantillon.

$$= \frac{\sum}{\sum \sum}$$

Afin de tenir compte du fait qu'il peut y avoir des facteurs spécifiques à la revue qui influencent les taux de citation, nous créons un poids de citation relatif spécifique à la revue pour chaque article qui quantifie la citation de chaque article impact dans le contexte de sa revue spécifique. Un poids relatif de citation plus élevé, propre à la revue, indique qu'un article est cité plus fréquemment que d'autres articles de la même revue. Nous calculons cela comme suit :

$$= \frac{\sum}{\sum}$$

où le dénominateur est la somme de tous les articles de la revue. Ainsi, pour chaque revue, la somme de tous les articles de la revue est de 1. En créant ce poids de citation intra-revue, nous sommes effectivement accorder plus de poids aux articles qui n'ont peut-être pas reçu un nombre élevé de citations parce qu'ils ont été publiés au sein d'une revue à faible impact mais qui était relativement influent au sein de cette revue à faible impact.

Enfin, nous divisons ces pondérations par article par le nombre d'auteurs. Notre impact final :

la mesure de sortie pondérée par le facteur et la citation est égale à

$$\frac{1}{\dots}$$

C. Construction de sous-domaines des sciences de la vie

Nous implémentons K-means++, qui améliore l'algorithme original en choisissant un centroïde initial plus intelligent. Cela tend à produire des clusters mieux séparés et équilibrés que l'algorithme original. Pour chaque auteur-année, nous combinons l'ensemble du texte (termes du maillage, titre, résumé) de leur source. papiers.

Année Bin: 1945–1964

1. Neurosciences : cerveau, physiologie, métabolisme, pharmacologie, foie, inhibition, tissu, animal, acide, médicament, chimie, sang, modèle, action, technique
2. Physiologie musculaire et biomécanique : physiologie, muscle, métabolisme, fonction, pharmacologie, foie, cellule, sang, porc, mécanisme, action, inhibition, influence, température, in vitro
3. Physiologie comparée : foie, tissu, animal, composé, métabolisme, muscle, solution, concentration, action, substance, fait, souche, urine, espèce, eau
4. Biochimie : acide, métabolisme, amino, synthèse, foie, formation, composé, pharmacologie, tissu, cent, extrait, isolé, solution, protéine, dérivé
5. Microbiologie : métabolisme, synthèse, formation, foie, bactéries, tissu, nucléotide, plante, glucose, mécanisme, composé, oxydation, in vitro, pharmacologie, vitamine
6. Biologie cellulaire : cellule, métabolisme, culture, tissu, acide, pharmacologie, croissance, foie, synthèse, milieu, sang, in vitro, animal, structure, démontré
7. Hématologie : sang, sérum, plasma, pharmacologie, cellule, métabolisme, chimie, acide, concentration, animal, lapin, mécanisme, foie, action, trois
8. Biologie moléculaire : protéines, métabolisme, synthèse, chimie, acide, foie, sérum, cellule, plasma, fraction, pharmacologie, amino, solution, composant, formation
9. Endocrinologie : hormone, pharmacologie, croissance, métabolisme, action, physiologie, injection, tissu, animal, urine, administration, chimie, sang, influence, substance
10. Biologie structurale : chimie, acide, structure, métabolisme, chromatographie, pharmacologie, tissu, synthèse, composant, contenu, nucléotide, plante, propriété, sérum, isolement
11. Pharmacologie : pharmacologie, action, médicament, métabolisme, acide, inhibition, agent, croissance, influence, dérivé, composé, induit, vitamine, expérimental, foie
12. Virologie : virus, souche, cellule, culture, acide, plante, isolé, pharmacologie, tissu, agent, induit, chimie, animal, protéines, métabolisme
13. Biologie des rayonnements : rayonnement, pharmacologie, cellule, induit, lumière, biologique, produit, solution, acide, métabolisme, action, tissu, corps, donné, plante
14. Enzymologie : enzyme, foie, métabolisme, synthèse, acide, chimie, phosphate, formation, tissu, pharmacologie, inhibition, action, propriété, protéine, extrait
15. Physiologie : facteur, sang, sérum, pharmacologie, physiologie, métabolisme, foie, chimie, vitamine, croissance, plasma, certain, animal, acide, protéine

Année Bin: 1965–1984

1. Biophysique et biologie membranaire : concentration, sang, taux, membrane, liaison, foie, cerveau, complexe, chaîne, muscle, structure, cytochrome, facteur, site, élevé
2. Métabolisme et biochimie : métabolisme, biosynthèse, pharmacologie, enzymologie, cerveau, foie, calcium, cellule, sang, synthèse, muscle, protéine, glucose, érythrocyte, sodium
3. Pharmacologie et action des médicaments : récepteur, liaison, métabolisme, cellule, membrane, affinité, cerveau, site, physiologie, hormone, médicament, pharmacologie, protéine, complexe, spécifique
4. Génétique moléculaire et microbienne : ADN, séquence, génétique, polymérase, cellule, métabolisme, synthèse, fragment, gène, protéine, coli, biosynthèse, site, génome, virus
5. Virologie et génétique virale : virus, cellule, immunologie, génétique, anticorps, antigène, maladie, ADN, ARN, souche, particule, protéine, culture, lignée, isolé
6. Enzymologie et fonction des protéines : enzyme, substrat, purifié, moléculaire, poids, acide, foie, sous-unité, gel, complexe, protéine, concentration, site, déshydrogénase, taux
7. Neurophysiologie et électrophysiologie : physiologie, muscle, métabolisme, cellule, cerveau, pharmacologie, potentiel, calcium, membrane, stimulation, cytologie, mécanisme, sodium, sang, fonction
8. Génétique : chromosome, génétique, gène, cellule, ADN, souris, région, physiologie, séquence, cytologie, long, métabolisme, lymphocyte, génétique, lignée
9. Biologie structurale : acide, amino, résidu, peptide, séquence, chaîne, protéine, structure, enzyme, isolé, synthèse, cellule, chromatographie, produit, moléculaire
10. Génomique et expression génétique : gène, génétique, séquence, ADN, région, expression, transcription, nucléotide, ARNm, ARN, cellule, cloné, mutation, protéine, souris
11. Immunologie : immunologie, antigène, anticorps, lymphocyte, cellule, sérum, souris, surface, spécifique, in vitro, génétique, virus, pharmacologie, sang, maladie
12. Signalisation cellulaire : protéine, kinase, membrane, cellule, liaison, moléculaire, sous-unité, gel, synthèse, poids, purifié, phosphorylation, acide, peptide, fraction
13. Neuroparmacologie : médicament, pharmacologie, métabolisme, action, induit, cellule, inhibition, cerveau, physiologie, comportement, récepteur, sang, in vitro, calcium, muscle
14. Biologie cellulaire : cellule, culture, lignée, cytologie, croissance, membrane, souris, synthèse, milieu, protéine, cultivé, surface, in vitro, tumeur, fibroblaste
15. Biologie moléculaire : ARN, ARNm, séquence, polymérase, synthèse, métabolisme, ADN, biosynthèse, cellule, transcription, nucléotide, gène, génétique, protéine, in vitro

Année Bin: 1985–2004

1. Biologie structurale et biophysique : liaison, domaine, peptide, structure, résidu, protéine, site, interaction, acide, complexe, séquence, région, mutant, affinité, cellule
2. Biologie du cancer et oncologie : cellule, expression, protéine, apoptose, croissance, activation, récepteur, lignée, tumeur, facteur, gène, anticorps, kinase, décès, cancer
3. Signalisation cellulaire et transduction du signal : kinase, phosphorylation, cellule, protéine, activation, tyrosine, récepteur, signalisation, phosphorylé, voie, domaine, insuline, facteur, croissance, site

4. Génétique moléculaire et régulatrice : promoteur, transcription, gène, cellule, expression, élément, facteur, site, protéine, liaison, région, transcriptionnel, séquence, ADN, nucléaire
5. Dynamique membranaire et cellulaire : protéine, cellule, séquence, acide, domaine, ADNc, membrane, amino, anticorps, complet, liaison, gène, complexe, région, mutant
6. Métabolisme cellulaire et bioénergétique : cellule, protéine, métabolisme, membrane, acide, immunologie, facteur, complexe, expression, chimie, médicament, muscle, transport, ARN, maladie
7. Biologie et pharmacologie des récepteurs : récepteur, cellule, liaison, protéine, ligand, métabolisme, domaine, affinité, activation, agoniste, expression, signalisation, mutant, physiologie, facteur
8. Neurosciences et neurobiologie : physiologie, neurone, métabolisme, cellule, cytologie, neuronal, cerveau, génétique, protéine, récepteur, neuronal, synaptique, apoptose, pharmacologie, facteurs de transcription
9. Génétique et troubles génomiques : génétique, gène, mutation, chromosome, syndrome, métabolisme, cancer, maladie, carcinome, protéine, facteurs de transcription, locus, souris, protéines de liaison à l'ADN, expression
10. Biologie des canaux ioniques et électrophysiologie : canal, Ca^{2+} , courant, cellule, calcium, sous-unité, membrane, protéine, récepteur, potassium, ion, intracellulaire, ovocyte, activation, neurone
11. Immunologie et immunogénétique moléculaire : alpha, bêta, sous-unité, cellule, récepteur, protéine, liaison, chaîne, gène, séquence, acide, ARNm, expression, complexe, peptide
12. Génomique et génomique fonctionnelle : gène, séquence, chromosome, génome, expression, cellule, génétique, protéine, région, exon, ARNm, ADN, ARN, mutation, souris
13. Biologie des suppresseurs de tumeurs et génétique du cancer : p53, cellule, apoptose, tumeur, gène, protéine, ADN, expression, suppresseur, cancer, génétique, mutant, arrêt, mutation, métabolisme
14. Dommages à l'ADN et stabilité du génome : ADN, réplication, réparation, protéine, polymérase, liaison, complexe, cellule, brin, site, séquence, génétique, gène, base, structure
15. Enzymologie et mécanismes catalytiques : enzyme, substrat, acide, résidu, protéine, site, mutant, actif, catalytique, purifié, structure, amino, séquence, sous-unité, réductase

Année Bin: 2005–2024

1. Biologie structurale : domaine, liaison, structure, protéine, résidu, site, complexe, interaction, peptide, enzyme, structure, substrat, récepteur, cristal, chimie
2. Génomique : génome, génétique, variant, gène, locus, population, SNP, génétique, à l'échelle du génome, polymorphisme, séquence, maladie, variation, chromosome, trait
3. Écologie et biologie environnementale : espèce, plante, différent, zone, eau, communauté, physiologie, population, élevé, condition, taux, modèle, performance, concentration, température
4. Génétique et médecine génomique : mutation, gène, génétique, séquençage, mutant, cancer, cellule, syndrome, maladie, protéine, identifiée, variante, génétique, tumeur, famille
5. Biologie moléculaire : gène, expression, cellule, transcription, promoteur, ARN, génétique, protéine, transcriptionnel, facteur, voie, identifié, plante, régulation, exprimé
6. Virologie : virus, infection, vaccin, anticorps, viral, cellulaire, immunitaire, infecté, souche, maladie, respiratoire, immunologie, hôte, protéine, antigène
7. Biologie du cancer : cancer, tumeur, cellule, sein, expression, métastase, poumon, croissance, carcinome, gène, thérapie, survie, progression, protéine, tissu

8. Biologie cellulaire : protéine, cellule, membrane, complexe, interaction, domaine, fonction, expression, mutant, métabolisme, kinase, liaison, phosphorylation, complet, acide
9. Épidémiologie : femme, enfant, santé, mortalité, maladie, résultat, VIH, facteur, prévalence, soins, cohorte, taux, supérieur, participant, score
10. Neurosciences : neurone, cerveau, neuronal, cortex, cellule, neural, physiologie, récepteur, mémoire, réseau, comportement, souris, expression, maladie, fonction
11. Métabolisme et endocrinologie : foie, expression, cellule, graisse, maladie, blessure, régime alimentaire, gène, lipide, métabolisme, acide, tissu, insuline, sérum, métabolique
12. Cellules souches et biologie régénérative : cellule, expression, souche, protéine, prolifération, apoptose, tumeur, ligne, récepteur, activation, cancer, signalisation, gène, croissance, voie
13. Immunologie : cellule, récepteur, expression, métabolisme, signalisation, activation, voie, médicament, macrophage, protéine, inflammation, maladie, fonction, mécanisme, inhibiteur
14. Physiologie et biologie des systèmes : muscle, cellule, expression, protéine, gène, mitochondrial, fonction, métabolisme, tissu, insuline, signalisation, activation, masse, physiologie, souris
15. Stabilité de l'ADN et du génome : ADN, réparation, cellule, dommage, réplication, protéine, gène, liaison, complexe, génome, séquence, site, chromatine, polymérase, génétique