

# **Analyse des publications et des brevets sur le cancer dans le monde et en France, 2000-2020**

## **VOLUME 1 : Analyse des publications**

**Juillet 2024**

## Préambule

Ce rapport présente les résultats issus d'une étude commanditée par l'Institut National du Cancer et réalisée par l'OST de septembre 2022 à juin 2023.

Les études de l'OST sont le fruit du travail d'équipe. Cette étude a été plus directement réalisée par Tessa ENOCK LEVI, Agéonor LAHATTE, Esther LARDREAU, Céline PELTIER, et Frédérique SACHWALD. Isabelle MEZIERES a contribué à la préparation du rapport en tant qu'assistante de projet.

Cette étude a été réalisée avec l'appui du Pôle Recherche de l'INCa (Karima BOUROUGAA, Carla ESTAQUIO et Bruno QUESNEL) et en sollicitant l'avis d'experts de domaine du cancer (en particulier, Pr Isabelle BOUTRON, professeur d'épidémiologie à l'Université de Paris, Responsable de l'équipe Méthode, INSERM- Centre d'épidémiologie et de statistique – CRESS et Directrice de Cochrane France ; Sophie GOURGOU, Responsable de l'unité de Biométrie, ICM ; Pr Marie PREAU Professeure de psychologie sociale de la santé et directrice adjointe de l'Unité Inserm U1296 rattachée à l'Université Lyon 2 ; Claude SARDET, Institut de recherche en cancérologie de Montpellier - IRCM U1194 et Virginie WESTEEL, UMR 1098, coordonnatrice de Cochrane Poumon et Présidente du CA de l'intergroupe francophone de cancérologie thoracique - IFCT).

## Sommaire

<b>Synthèse</b> .....	<b>4</b>
<b>Introduction</b> .....	<b>6</b>
<b>1. Les publications mondiales sur le cancer depuis 2000</b> .....	<b>6</b>
Constitution du corpus de publications sur le cancer .....	6
Publications scientifiques dans le domaine du cancer, 2000-20 .....	6
<b>2. Les principaux pays publiant dans le domaine du cancer</b> .....	<b>8</b>
Evolution des publications des premiers pays contributeurs .....	8
Les co-publications internationales dans le monde et en France .....	10
Evolution des publications en accès ouvert dans le domaine du cancer .....	12
<b>3. Spécialisation des pays dans la recherche sur le cancer</b> .....	<b>13</b>
<b>4. Mesures de l'impact scientifique des publications sur le cancer</b> .....	<b>14</b>
Indice d'impact moyen des dix premiers pays publiants .....	14
Publications les plus citées au monde .....	15
<b>5. Analyse des publications par type de recherche</b> .....	<b>16</b>
Contexte et problématique .....	16
Méthode d'élaboration des sous-corpus .....	16
Principaux résultats .....	17
<b>Annexe 1 : Données et Méthodologie</b> .....	<b>24</b>
<b>Annexe 2 : Méthode détaillée de constitution du corpus Cancer</b> .....	<b>28</b>
<b>Annexe 3 : Requêtes utilisées pour construire les sous corpus par type de Recherche</b> .....	<b>31</b>
<b>Références Bibliographiques</b> .....	<b>32</b>

## Synthèse

L'étude des publications scientifiques relatives au cancer s'appuie sur la constitution d'un corpus de publications spécifique construit en collaboration avec un groupe d'experts.

### **Caractérisation des publications mondiales dans le domaine du cancer**

Entre 2000 et 2020, les publications sur le cancer représentent 8 % du total mondial et 20 % des publications en recherche biomédicale. Le nombre de publications scientifiques relatives au cancer dans le monde, a été multiplié par 3 entre 2000 et 2020. Ces publications appartiennent à plus de 80 % aux disciplines de la Recherche Médicale ou de la Biologie Fondamentale. Toutefois, on peut noter l'émergence de publications en Chimie. Les publications scientifiques de la France sur le cancer augmentent légèrement moins vite que le total mondial ; elles ont été multipliées par 2, comme celles des pays européens. La part des publications de recherche médicale y est encore plus élevée.

### **Dynamiques nationales**

L'augmentation du nombre de publications est tirée notamment par la Chine qui passe du 11<sup>ème</sup> rang mondial en 2000 au 1<sup>er</sup> rang en 2017. La forte croissance du nombre de publications de la Chine se traduit par une spécialisation accrue du pays dans le domaine du cancer. La Chine a désormais une part de ses publications scientifiques supérieure de 20 % à la part de ce domaine dans les publications mondiales, soit un indice de spécialisation de 1,2. Les publications de la Chine enregistrent aussi une amélioration de leur indice moyen d'impact scientifique ; elle se positionne désormais autour de la moyenne mondiale

Les Etats-Unis sont désormais le deuxième pays publiant le plus après la Chine. Ils restent spécialisés dans le domaine du cancer, avec un indice de spécialisation supérieur à 1,1. Les indicateurs d'impact des publications américaines sont élevés ; leur indice d'impact moyen est, comme celui du Royaume-Uni, supérieur de plus de 30 % à la moyenne mondiale. Le Japon est le troisième pays publiant le plus sur le cancer. C'est un des pays les plus spécialisés dans le domaine avec l'Italie. L'impact des publications japonaises sur le cancer est néanmoins relativement faible avec des indices inférieurs de 20 % à la moyenne mondiale sur l'ensemble de la période.

Parmi les pays européens, la France, le Royaume-Uni et l'Allemagne voient leur rang reculer en nombre de publications, au contraire de l'Italie qui progresse nettement et passe devant ces trois pays. L'Italie est très spécialisée dans le domaine du cancer durant toute la période étudiée (indice 1,4). La France maintient son indice de spécialisation légèrement au-dessus de la moyenne mondiale (1), contrairement à l'Allemagne et au Royaume-Uni qui perdent leur spécialisation au cours de la décennie 2010. Le Royaume-Uni a des indices d'impact en forte croissance qui positionnent le pays juste derrière les Etats-Unis. La France, l'Italie et l'Allemagne qui avaient des indices d'impact inférieur à la moyenne mondiale ont progressé et atteignent en 2019 des indices d'impact supérieurs à la moyenne mondiale (1).

Des pays de profils divers accroissent sensiblement leur nombre de publications dans le domaine du cancer, comme la Corée du Sud, l'Inde, l'Iran ou le Brésil. Le phénomène concerne la Corée du Sud à partir de 2005 et le pays a désormais un profil similaire à celui du Japon, avec une forte spécialisation et un indice d'impact des publications assez faible. L'Inde, malgré une forte croissance du nombre de publications, reste peu spécialisée dans le domaine.

### **Accroissement de la part des co-publications internationales et des publications en accès ouvert**

La part des co-publications internationales dans le domaine du cancer augmente dans tous les pays, sauf en Chine où elle a diminué depuis les années 2000 pour ne représenter que 17 % du total. Les Etats-Unis, malgré leur très grand nombre de publications sur le cancer, ont une part de co-publications internationales en croissance et nettement plus élevée que la Chine (42 % en 2020). Les pays ayant les taux de co-publications les plus élevés (> 60 %) sont les mêmes que dans l'ensemble des domaines de recherche : Suisse, Suède, Pays-Bas,

Royaume-Uni, Australie, Canada. La part des co-publications internationales de la France est similaire à celle de l'Allemagne ou de l'Espagne, entre 58 et 60 %.

Les pratiques de publications sur le cancer en accès ouvert se développent dans tous les pays étudiés. La part des publications sur le cancer en accès ouvert augmente dès 2003 aux Etats-Unis pour se stabiliser à partir de 2013. Le Japon, la Corée du Sud et l'Italie ont aussi vu leur part de publications en accès ouvert se développer à partir de 2006 et ont désormais rejoint le niveau des Etats-Unis. La part des publications françaises en accès ouvert augmente progressivement et accélère depuis 2018, ce qui lui permet d'approcher les taux des pays plus précoces en fin de période.

### **Recherche fondamentale et recherche clinique dans le domaine du cancer**

Les publications en recherche fondamentale et en recherche clinique étudiées représentent 40 % de l'ensemble du corpus cancer. Le nombre de publications augmente fortement, en particulier en recherche clinique, dont la part dans le corpus cancer passe de 13 % en 2000 à 23 % en 2020, la part de la recherche fondamentale diminuant de 24 % à 21 %.

Les principaux pays publiant dans le domaine du cancer ont une part de publications en recherche fondamentale comprise entre 20 et 25 % en 2020, sauf le Japon où cette part n'est que de 15 %. La part des publications en recherche clinique augmente pour les principaux pays publiant, pour atteindre de 19 % à 28 %.

Les pays les plus spécialisés en recherche clinique dans le domaine du cancer sont la Chine, le Japon, le Royaume-Uni, l'Italie et la France. En termes d'impact, les Etats-Unis, suivis du Royaume-Uni et du Canada, ont des indices d'impact nettement supérieurs à la moyenne mondiale depuis 20 ans. Si l'indice d'impact de la Chine s'améliore nettement en recherche fondamentale pour être supérieur à la moyenne mondiale à partir de 2018, il reste bien en deçà de la moyenne en recherche clinique. La France progresse régulièrement en matière d'impact, en particulier en recherche clinique où elle rejoint le trio de tête Etats-Unis, Royaume-Uni, Canada.

## Introduction

Dans le cadre de l'analyse des publications et des brevets relatifs au cancer depuis 2000, ce volume présente les indicateurs concernant les publications. La première partie présente la méthode de construction du corpus mondial de publications sur le cancer et ses caractéristiques. La deuxième partie positionne les principaux pays publiant dans le domaine du cancer y compris leurs co-publications et leur engagement dans les publications en accès ouvert. La troisième partie évalue le degré d'engagement relatif des pays dans le domaine du cancer avec l'indice de spécialisation de leurs publications. La quatrième partie présente des mesures de l'impact scientifique des vingt pays publiant le plus dans le domaine du cancer. La cinquième partie présente les spécificités en termes de publications en recherche fondamentale et en recherche clinique. Le cas de la France est plus particulièrement analysé dans chacune des parties du rapport.

Les annexes fournissent une description de la méthode de constitution des corpus (annexe 1) et des précisions méthodologiques sur le calcul des indicateurs (annexe 2).

## 1. Les publications mondiales sur le cancer depuis 2000

### **Constitution du corpus de publications sur le cancer**

Pour délimiter un corpus de publications scientifiques relatives à un sujet de recherche, l'analyse est menée au grain de la publication individuelle. Il s'agit de déterminer si le vocabulaire utilisé dans la publication relève de la thématique de recherche étudiée. Cette méthode consiste à définir un ensemble de publications à partir de leur contenu sémantique. La sélection est faite à l'aide d'une liste de « mots-requêtes » qui sont des termes caractéristiques de la thématique et dont on cherche la présence dans le titre, les mots-clés auteurs ou le résumé des publications.

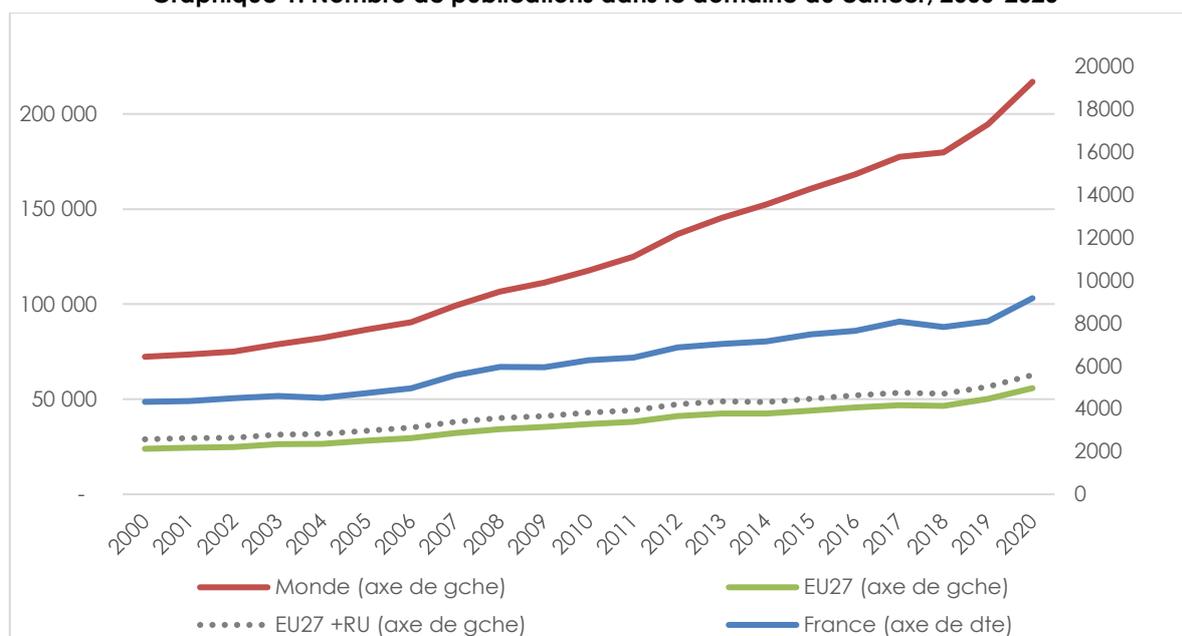
Du fait de la polysémie de certains mots-requêtes, le corpus ainsi construit comporte généralement des documents hors du domaine d'intérêt. Il est donc nécessaire de réduire le bruit représenté par ces publications scientifiques non pertinentes qui peut aussi masquer certains aspects d'une thématique. Il faut donc aussi analyser le « silence » représenté par les publications du domaine non présentes dans le corpus et voir dans quelle mesure il est possible de le compléter en identifiant de nouveaux mots requêtes.

Une approche itérative a été adoptée en sollicitant le groupe d'experts du domaine associés à cette étude à différentes étapes de la constitution du corpus cancer. Les différentes étapes de construction du corpus sont décrites en annexe 2.

### **Publications scientifiques dans le domaine du cancer de 2000 à 2020**

Le corpus cancer compte 2,6 M de publications scientifiques sur le cancer de 2000 à 2020. Il représente 8 % de l'ensemble des publications scientifiques sur cette période et 20 % des publications dans le domaine de la recherche biomédicale (Biologie Fondamentale et Recherche Médicale).

**Graphique 1. Nombre de publications dans le domaine du cancer, 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte entier - 2020 : année complète à 95 %

Le nombre de publications concernant le cancer dans le monde est en croissance, particulièrement sur la dernière décennie (graphique 1). La croissance est moins importante dans l'Union européenne et en France : le nombre de publications a été multiplié par 3 dans le monde, 2,3 pour l'UE 27 et par 2,1 pour la France.

Les publications concernant le cancer sont parues à 65 % dans des revues ou actes de conférences en Recherche Médicale et à 16 % en Biologie Fondamentale (tableau 1). La part des publications de ces deux disciplines scientifiques est en légère baisse depuis 20 ans dans les publications concernant le cancer, passant de 71 % à 65 % pour la Recherche Médicale et de 17 % à 16 % pour la Biologie Fondamentale. Cela s'explique par la hausse de publications relatives à d'autres disciplines, notamment en Chimie, dont la part passe de 2,4 % du corpus cancer en 2000-2005 à 7,4 % en 2016-2020. On observe aussi une augmentation (plus réduite) de la part des publications en Informatique, Mathématiques, Physique et Science pour l'Ingénieur. La part des publications en Sciences Humaines et Sociales, en légère croissance, reste inférieure à 2 %.

La part des publications relevant de la spécialité du WoS ONCOLOGIE représente 27 à 31 % du corpus en croissance de 2000 à 2020. Les autres spécialités représentent moins de 7 % du corpus. On peut noter la forte croissance au niveau mondiale des publications relevant des spécialités CHIMIE GENERALE, CHIMIE, PHARMACEUTIQUE, MEDECINE EXPERIMENTALE ET TRANSLATIONNELLE et BIOLOGIE CELLULAIRE.

**Tableau 1 : Distribution des publications mondiales sur le cancer par discipline, 2000-2020**

	2000-05	2006-10	2011-15	2016-20	Croissance 2001-05 à 2016-20
RECHERCHE MÉDICALE	71,24 %	69,98 %	67,59 %	64,84 %	- 9 %
BIOLOGIE FONDAMENTALE	17,21 %	15,67 %	15,97 %	16,15 %	- 6 %
CHIMIE	2,40 %	3,72 %	5,63 %	7,43 %	209 %
BIOLOGIE APPLIQUÉE-ÉCOLOGIE	3,70 %	3,90 %	3,95 %	4,16 %	12 %
SCIENCES POUR L'INGÉNIEUR	1,66 %	2,24 %	2,20 %	2,25 %	35 %
SCIENCES SOCIALES	1,46 %	1,61 %	1,66 %	1,61 %	11 %
PHYSIQUE	0,81 %	1,06 %	1,23 %	1,24 %	54 %
INFORMATIQUE	0,63 %	0,78 %	0,77 %	1,05 %	67 %
SCIENCES DE L'UNIVERS	0,50 %	0,50 %	0,51 %	0,80 %	61 %
MATHÉMATIQUES	0,20 %	0,25 %	0,24 %	0,25 %	25 %
SCIENCES HUMAINES	0,20 %	0,28 %	0,26 %	0,22 %	10 %

Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

La part des publications scientifiques en Recherche Médicale est encore plus élevée parmi les publications scientifiques sur le cancer de la France (tableau 2) : elle baisse très légèrement depuis 2000 et représente 74 % des publications sur le cancer de 2016-2020 (tableau 2).

**Tableau 2 : Distribution des publications de la France sur le cancer par discipline, 2000-2020**

	2000-05	2006-10	2011-15	2016-20	Croissance 2001-2005 à 2016-2020
RECHERCHE MÉDICALE	74,17 %	75,37 %	73,55 %	73,61 %	- 1 %
BIOLOGIE FONDAMENTALE	16,95 %	14,33 %	14,83 %	14,04 %	- 17 %
CHIMIE	2,26 %	2,88 %	3,64 %	4,37 %	93 %
BIOLOGIE APPLIQUÉE-ÉCOLOGIE	3,06 %	2,50 %	2,38 %	2,40 %	- 21 %
SCIENCES POUR L'INGÉNIEUR	1,21 %	1,87 %	2,06 %	1,75 %	45 %
SCIENCES SOCIALES	0,74 %	0,92 %	1,18 %	1,36 %	84 %
PHYSIQUE	0,60 %	0,84 %	1,01 %	0,96 %	58 %
INFORMATIQUE	0,51 %	0,58 %	0,61 %	0,67 %	31 %
SCIENCES DE L'UNIVERS	0,32 %	0,33 %	0,27 %	0,46 %	41 %
MATHÉMATIQUES	0,13 %	0,25 %	0,31 %	0,26 %	103 %
SCIENCES HUMAINES	0,05 %	0,12 %	0,16 %	0,13 %	151 %

Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95%

La part de la recherche en Biologie Fondamentale, équivalente à celle du monde en début de 2000 à 2005 (17 %), diminue plus rapidement en France et ne concerne plus que 14 % de 2016 à 2020 (contre 16 % dans le monde). La part des publications en Biologie Appliquée-Ecologie, équivalente à celle du monde en 2000-2005, est aussi en baisse, passant à 2,4 %, alors qu'elle est en légère hausse pour les publications mondiales sur le cancer, atteignant 4 %.

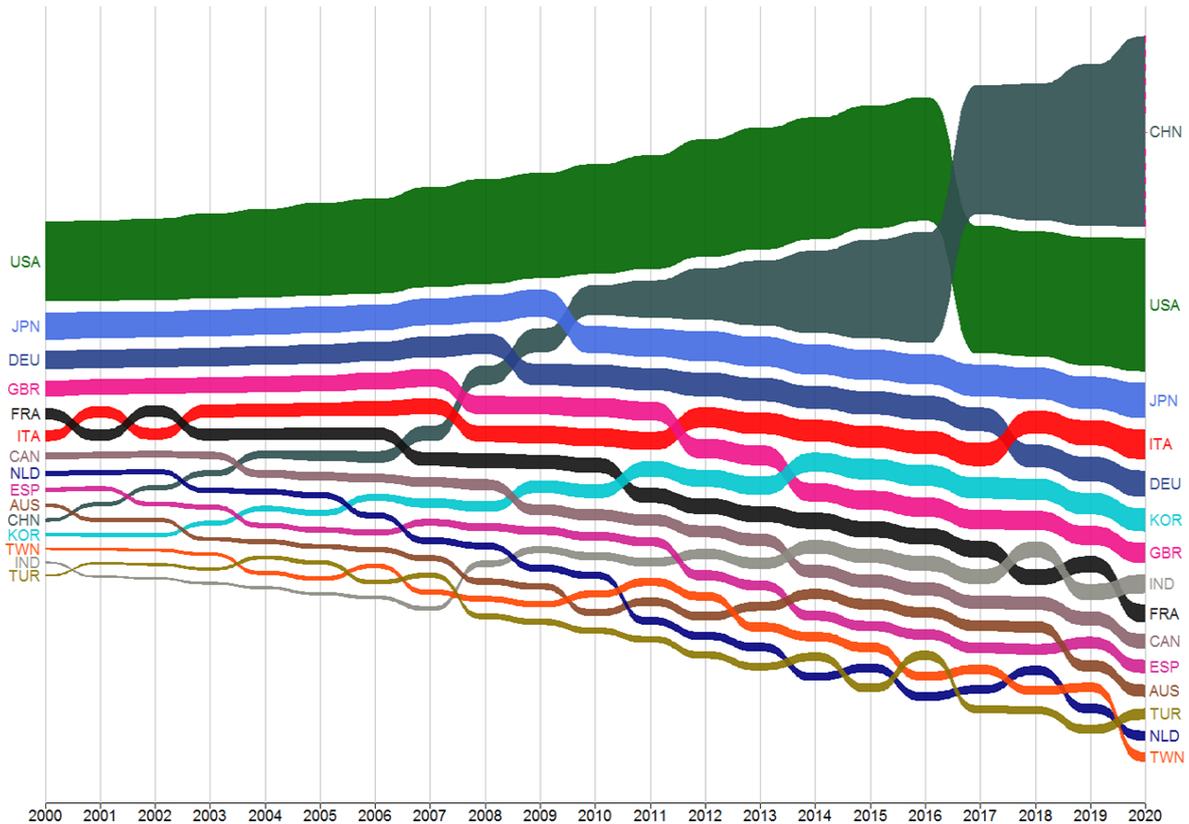
La part des publications en Chimie est passé de 2,3 % à 4,4 %, mais cette croissance a été moins importante que parmi les publications sur le cancer dans le monde où cette part atteint 7,4 % des publications sur le cancer. Comme pour le corpus Monde, la part des publications relevant de la spécialité du WoS ONCOLOGIE représente 25 à 31 % du corpus en croissance de 2000 à 2020. Les autres spécialités représentent à nouveau moins de 7 % du corpus. Les spécialités du corpus cancer en forte croissance pour la France sont différentes. Il s'agit des spécialités OBSTETRIQUE ET GYNECOLOGIE, SANTE PUBLIQUE, SANTE AU TRAVAIL ET RISQUES ENVIRONNEMENTAUX et RADIOLOGIE, MEDECINE NUCLEAIRE ET IMAGERIE MEDICALE.

## 2. Les principaux pays publiant dans le domaine du cancer

### Evolution des publications des premiers contributeurs

La Chine est passée du 11<sup>ème</sup> rang des pays publiant sur le cancer en 2000 au 1<sup>er</sup> rang en 2018, devant les Etats-Unis (Graphique 2). La Corée du Sud et l'Inde enregistrent aussi des progressions remarquables. Pour ces trois pays, l'évolution dans le domaine du cancer correspond à l'évolution des publications scientifiques tous domaines confondus [1]. Parmi les acteurs historiques, le Japon passe du second rang au troisième, alors que l'Allemagne, le Royaume-Uni et le Canada reculent de 2 à 3 rangs. La France passe elle du rang de 4<sup>ème</sup> producteur au 9<sup>ème</sup>. Au contraire, l'Italie voit son nombre de contributions augmenter et passe devant la France, le Royaume-Uni et l'Allemagne.

**Graphique 2. Nombre de publications des 15 premiers pays publiant dans le domaine du cancer, 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

Le tableau 3 confirme la très forte croissance de la Chine, dont la part de contributions dans le domaine du cancer est multipliée par 9. Les contributions de l'Inde augmentent de 155 % et celles de la Corée du Sud de 88 %. Le tableau souligne aussi l'émergence rapide de pays ayant encore une part de contributions faibles, comme l'Iran (part de publications multipliée par 12), le Brésil (+ 85 %) et la Pologne (+ 38 %).

Les pays à hauts revenus enregistrent des baisses de leurs parts de contributions importantes, entre 37 % et 48 % pour les Etats-Unis, le Japon, l'Allemagne, le Royaume-Uni et la France. Les parts de l'Italie et de l'Espagne enregistrent des baisses plus faibles, avec respectivement 17 et 5 %.

**Tableau 3 : Part mondiale des 20 premiers pays publiant dans le domaine du cancer, 2000-2020**

Pays	2000-05	2006-10	2011-15	2016-20	Croissance 2001-05 à 2016-20
Chine	2,6	5,9	14,5	24,3	819 %
Etats-Unis	33,7	30,6	25,9	21,3	- 37 %
Japon	10,4	8,0	6,4	5,4	- 48 %
Italie	5,0	5,0	4,5	4,2	- 17 %
Allemagne	7,4	6,3	5,0	4,1	- 45 %
Corée du Sud	1,9	3,2	4,0	3,6	88 %
Royaume-Uni	6,3	5,4	4,2	3,3	- 48 %
Inde	1,1	1,8	2,5	2,8	155 %
France	4,6	4,2	3,5	2,8	- 39 %
Canada	3,0	3,1	2,7	2,3	- 22 %
Espagne	2,0	2,3	2,2	1,9	- 5 %
Australie	1,8	1,9	2,1	1,9	6 %
Iran	0,1	0,5	0,9	1,4	1 119 %
Turquie	1,4	1,7	1,7	1,5	12 %
Brésil	0,8	1,4	1,6	1,6	85 %
Pays Bas	2,2	2,1	1,8	1,6	-29 %
Taiwan	1,5	1,8	2,0	1,6	7%
Pologne	0,8	1,0	1,1	1,1	38 %
Suisse	1,2	1,0	0,9	0,8	- 29 %
Suède	1,5	1,2	1,0	0,8	- 44 %

Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

### **Les co-publications internationales dans le monde et en France**

La part des co-publications internationales est le pourcentage des publications du domaine cancer d'un pays qui sont co-publiées avec une ou plusieurs institutions à l'étranger. Dans la plupart des pays, les co-publications internationales se sont développées de 2000 à 2020 (tableau 4). La Chine fait exception ; la forte croissance du nombre de publications chinoises s'accompagne d'une baisse de la part des co-publications internationales de la Chine qui passent de 25,9% à 16,6 %. Les autres pays asiatiques, Japon, Corée du Sud et Taiwan, ont une part de co-publications en croissance mais qui reste plus relativement basse (entre 25 et 34 %).

Les Etats-Unis ont une part de co-publications internationales de 42,2 %, en augmentation sensible, mais qui reste inférieure à celle des pays européens – ce qui n'est pas spécifique au domaine du cancer.

Les petits pays européens ont des parts de co-publications internationales supérieures à 65 % (Suisse, Suède, Pays-Bas). Le Royaume-Uni, le Canada et l'Australie ont aussi, comme dans toutes les disciplines, une proportion de co-publications internationales élevée (entre 60 et 70 %). L'Allemagne, la France et l'Espagne ont des parts de co-publications internationales de l'ordre de 60 %. L'Italie, dont la croissance des publications a été plus importante, a une part de co-publications internationales plus faible (48,5 %).

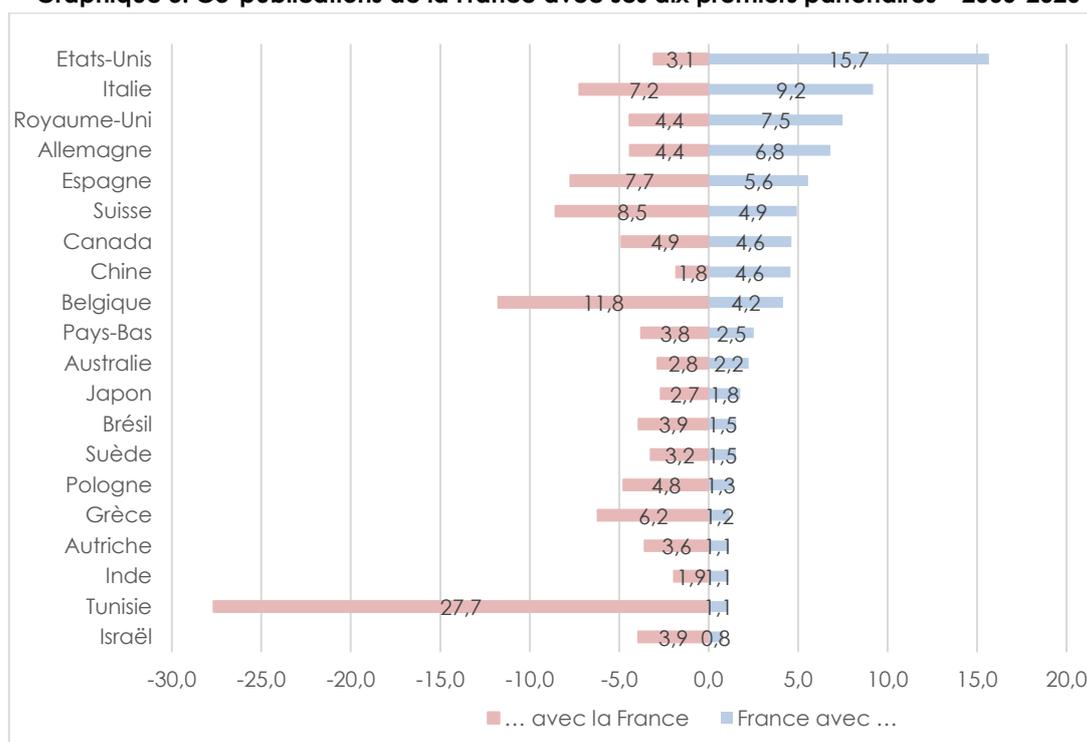
**Tableau 4 : Part des co-publications internationales des vingt premiers pays publiant sur le cancer, 2000-2020**

	2000	2005	2010	2015	2020
Chine	25,9	28,4	25,2	18,6	16,6
Etats-Unis	19,3	23,2	28,9	36,9	42,2
Japon	13,5	17,9	19,5	21,7	25,1
Italie	28,4	32,9	37,7	45,1	48,5
Allemagne	26,2	36,0	43,6	51,9	59,7
Corée du Sud	19,7	20,5	18,4	22,6	26,0
Royaume-Uni	29,7	36,5	46,7	60,2	70,5
Inde	13,0	14,7	15,6	22,9	32,5
<b>France</b>	<b>28,4</b>	<b>35,8</b>	<b>40,8</b>	<b>50,1</b>	<b>59,0</b>
Canada	37,4	42,7	48,8	56,0	63,3
Espagne	23,9	34,0	40,1	51,4	58,6
Australie	31,5	37,1	45,9	51,6	63,0
Iran	20,8	26,1	19,0	24,3	34,6
Turquie	11,4	10,2	11,9	17,1	24,9
Brésil	30,6	28,5	24,8	35,7	42,8
Pays-Bas	37,6	41,5	50,6	58,8	65,4
Taiwan	16,4	15,6	21,0	25,4	34,3
Pologne	31,4	33,9	31,1	34,0	42,1
Suisse	50,1	58,5	69,2	74,1	79,2
Suède	41,2	47,8	63,2	68,2	73,7

Source : Base OST, Web of Science, calculs OST. Compte entier - 2020 : année complète à 95 %

Le premier pays partenaire de la France est les Etats-Unis, avec 16 % des co-publications internationales françaises (graphique 3).

**Graphique 3. Co-publications de la France avec ses dix premiers partenaires – 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte entier - 2020 : année complète à 95 %

Les autres principaux pays partenaires sont les trois grands voisins de la France qui représentent chacun entre 7 et 9 % des co-publications (9,2 % pour l'Italie, 7,5 % pour le Royaume-Uni et 6,8 % pour l'Allemagne)

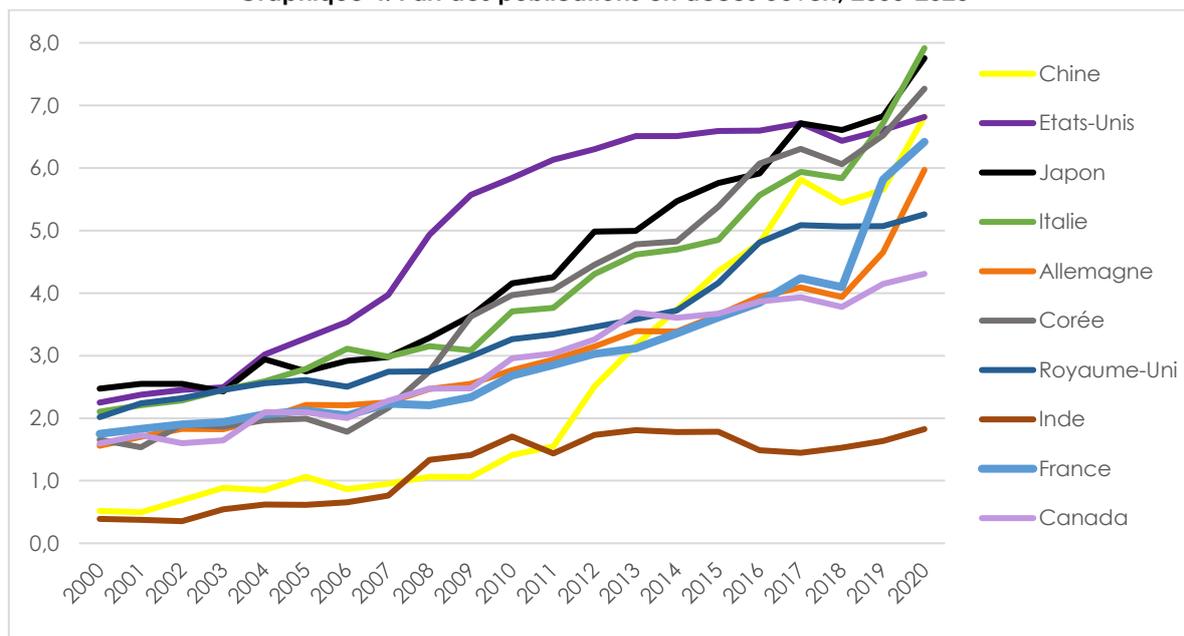
La part des co-publications de la France avec la Chine, le Japon, la Corée du Sud et l'Inde est relativement réduite eu égard au volume de publications scientifiques dans le domaine du cancer de ces pays.

On peut aussi identifier des pays ayant une position singulière vis-à-vis de la France comme la Suisse, la Belgique, la Grèce, la Tunisie ou encore Israël. Il s'agit de petits pays par le nombre de publications, mais dont la part des publications en collaboration avec la France est particulièrement élevée probablement pour des raisons de proximité géographique, linguistique, culturelles ou historique.

### Evolution des publications en accès ouvert dans le domaine du cancer

La part des publications scientifiques sur le cancer en accès ouvert est en croissance pour la plupart des pays (graphique 4). Cette croissance a été plus précoce aux Etats-Unis dès 2003. Le Japon, la Corée du Sud et l'Italie ont aussi vu leur part de publications en accès ouvert augmenter fortement à partir de 2006 pour rejoindre et dépasser le niveau des Etats-Unis à partir de 2017 (pour le Japon) et 2018 pour l'Italie et la Corée du Sud. La Chine dont le taux de publications sur le cancer en accès ouvert est resté inférieur à 1,5 % jusqu'en 2011 a rejoint les Etats-Unis en 2020. Le Royaume-Uni, la France et l'Allemagne ont vu leur taux de publications en accès ouvert augmenter progressivement avec une très nette augmentation depuis 2018 pour la France et l'Allemagne qui approchent les taux des pays plus précoces en fin de période.

**Graphique 4. Part des publications en accès ouvert, 2000-2020**

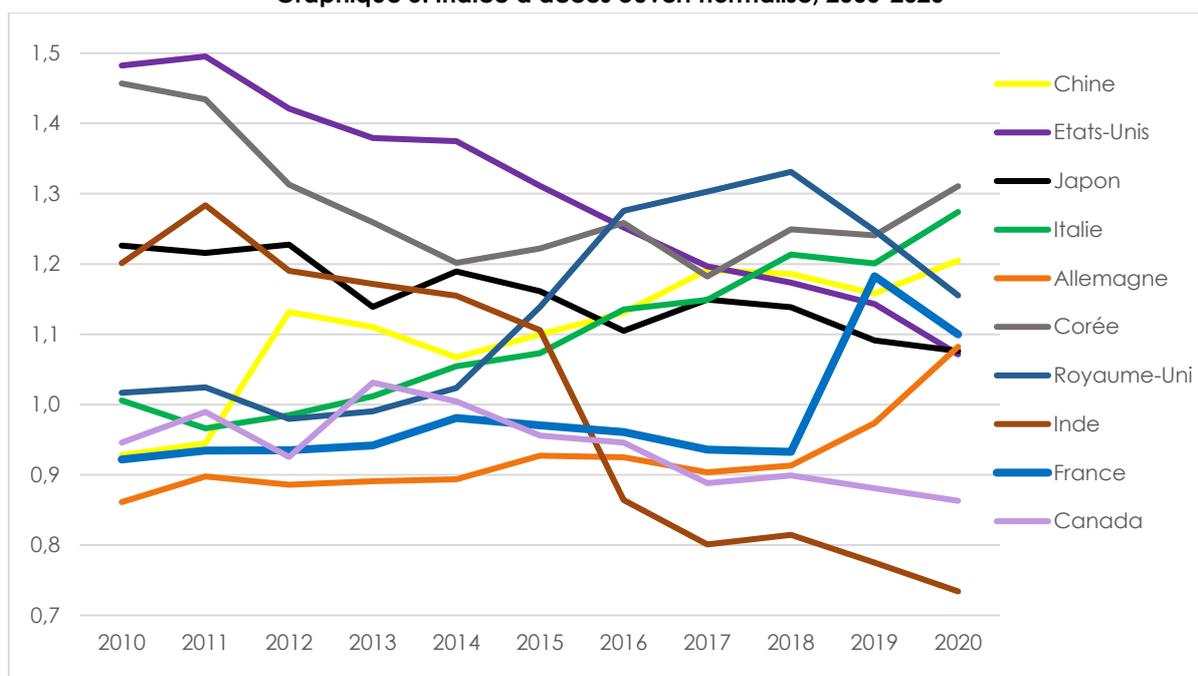


Source : Base OST, Web of Science, calculs OST. Compte entier - 2020 : année complète à 95 %

La publication en accès ouvert est diversement pratiquée selon les disciplines. L'indice normalisé permet de comparer les pays entre eux malgré des profils disciplinaires différents. Il rapporte la part de publications en accès ouvert d'un pays dans un domaine rapporté à la même part pour le total des publications mondiales. La valeur neutre de l'indice d'accès ouvert est 1.

Les indices normalisés d'accès ouvert des 10 premiers pays publiant dans le domaine du cancer montrent une certaine convergence : compris entre 0,95 et 1,5 en 2000, ils sont compris entre 1,1 et 1,3 en 2020, hors Canada et Inde qui sont les deux seuls pays à avoir un indice inférieur 1 (graphique 5). Notons que la valeur de la dernière année peut être influencée par la proportion de publications de statut bronze qui évolue dans le temps.

**Graphique 5. Indice d'accès ouvert normalisé, 2000-2020**



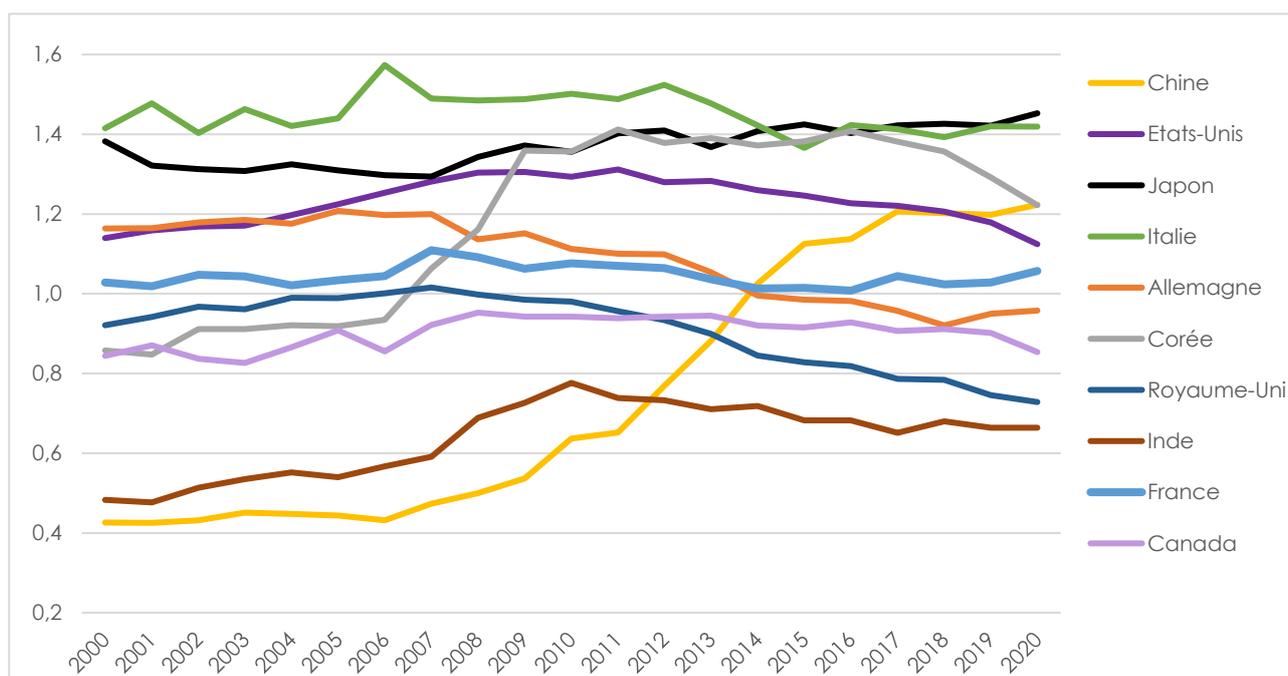
Source : Base OST, Web of Science, calculs OST. Compte entier - 2020 : année complète à 95 %

### 3. Spécialisation des pays dans la recherche sur le cancer

Les pays sont plus ou moins engagés dans la recherche sur le cancer par rapport à l'ensemble des domaines de recherche. L'indice de spécialisation mesure le poids relatif de la recherche sur le cancer d'un pays par rapport au poids moyen de la recherche sur le cancer dans le monde.

Le graphique 6 montre que l'Italie est l'un des pays les plus constamment spécialisés dans le domaine du cancer, suivi de près par le Japon. L'Italie a une part de ses publications dans le domaine 40 % supérieure à la moyenne mondiale.

**Graphique 6. Indice de spécialisation dans le domaine du cancer des 10 premiers pays publiant, 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

L'indice de spécialisation des Etats-Unis a varié, tout en étant parmi les plus élevés, compris entre 1,1 et 1,3.

La France maintient son indice de spécialisation entre 1 et 1,1, alors que l'Allemagne perd sa spécialisation, à la fin des années 2000. Le Royaume-Uni voit son indice de spécialisation encore plus baisser pour descendre en dessous de 0,8 en fin de période.

L'indice de spécialisation de la Chine a augmenté fortement de 2011 à 2017 pour atteindre 1,2 : le nombre de publications scientifiques de la Chine a considérablement augmenté tous domaines scientifiques confondus, mais proportionnellement encore plus dans le domaine du cancer.

L'évolution des indices de spécialisation peut notamment s'expliquer par les politiques publiques nationales, et notamment les politiques de recherche dans le domaine du cancer. Au sein même de la recherche biomédicale, le cancer peut bénéficier d'engagements plus ou moins forts, en partie liés à la prévalence de différents cancers dans le pays. La France est plus spécialisée dans le domaine du cancer, que certains pays à hauts revenus qui sont pourtant plus spécialisés que la France dans la recherche médicale ou le domaine plus large de la recherche biomédicale [2]. Ce profil pourrait être lié aux efforts consentis dans le cadre des plans cancer. La France a notamment particulièrement développé les essais cliniques dans le domaine du cancer sur la période 2010-2019 [3].

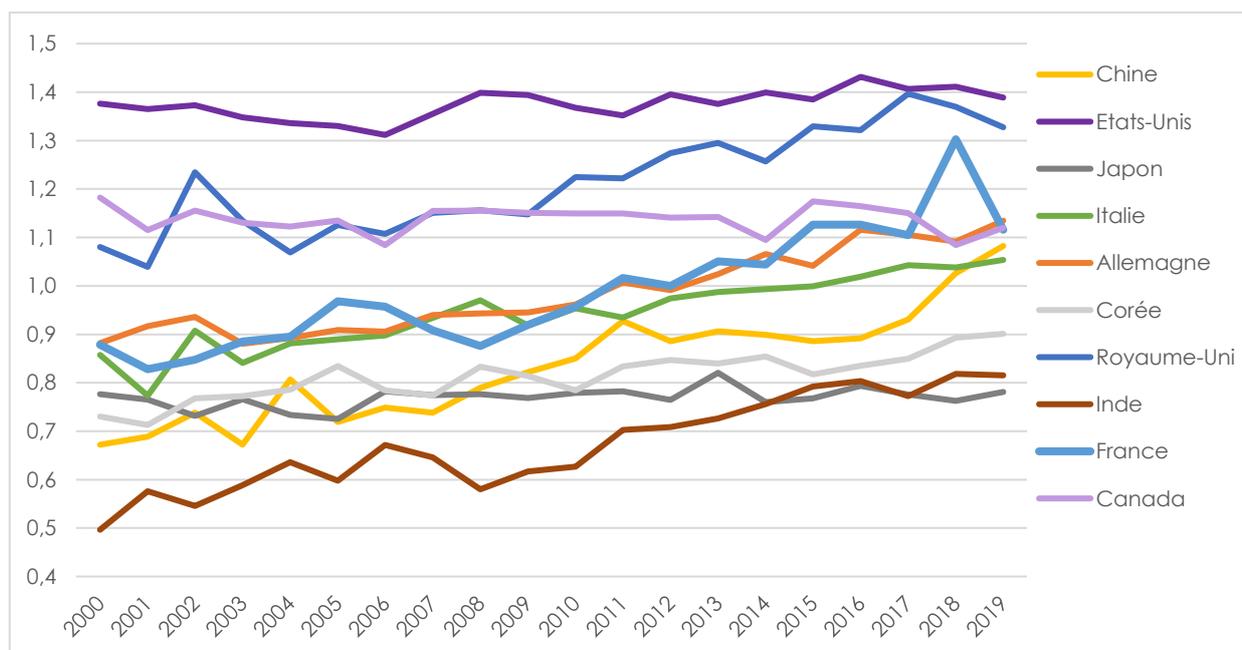
#### 4. Mesures de l'impact scientifique des publications sur le cancer

##### Indice d'impact moyen des dix premiers pays publiants

L'indice d'impact moyen est le rapport de la moyenne des citations normalisées des publications d'un pays et de la moyenne des citations normalisées des publications mondiales. La normalisation des citations est réalisée par domaine scientifique. Un indice supérieur à 1 signifie que les publications d'un pays ont en moyenne un impact plus élevé que celui de l'ensemble des publications mondiales. Un indice inférieur à 1 implique le contraire.

Les Etats-Unis sont depuis 20 ans le pays qui a l'indice d'impact dans le domaine du cancer le plus élevé à 1,4 (graphique 7). Le Royaume-Uni a connu une forte croissance de son indice d'impact qui est désormais très proche de celui des Etats-Unis. Le troisième pays ayant historiquement un indice d'impact supérieur à 1 est le Canada dont l'indice est resté entre 1,1 et 1,2 depuis 2000.

Graphique 7. Indice d'impact des publications sur le cancer des 10 premiers pays publiant, 2000-2019



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

La France, l'Allemagne et l'Italie qui avaient en 2000 un indice d'impact proche de 1 sont plutôt autour de 1,1 en fin de période.

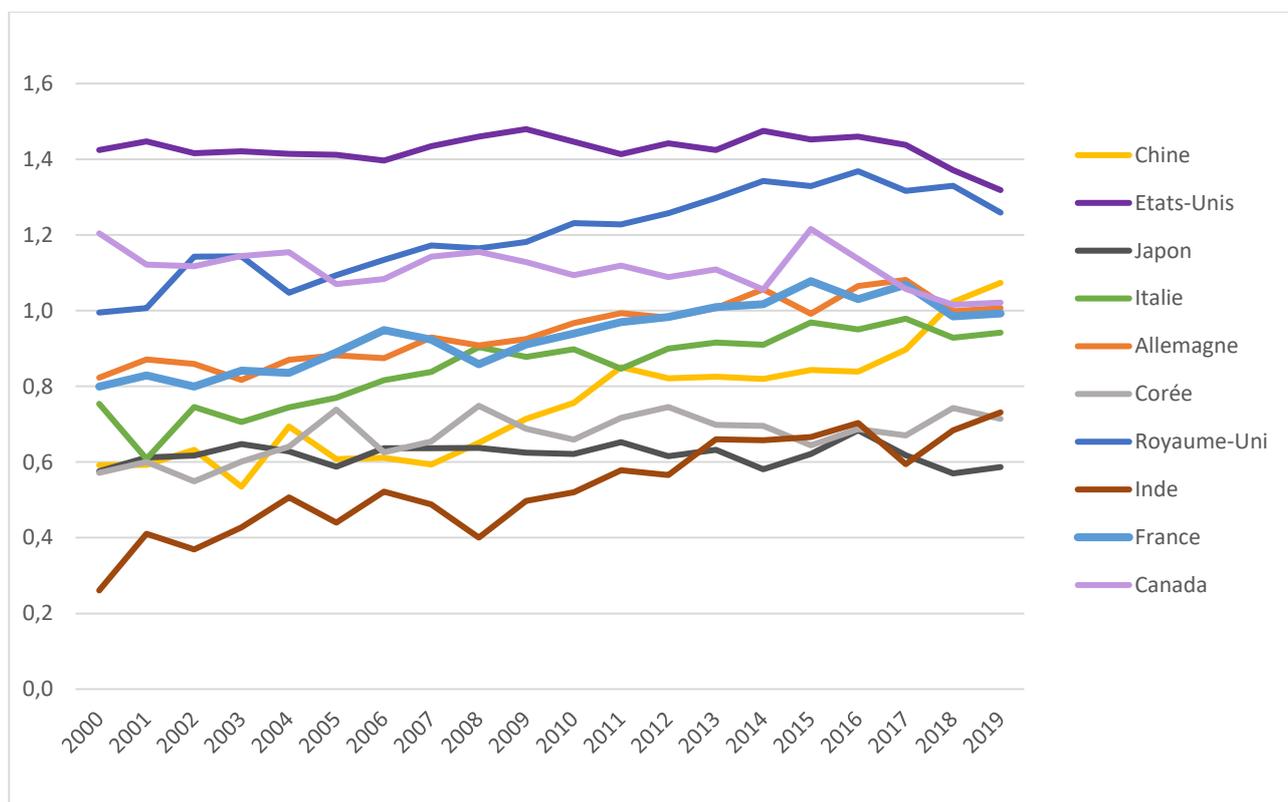
La Chine qui avait un indice d'impact faible en 2000 (0,7) a fortement progressé pour atteindre 1,1. Le Japon et la Corée du Sud, très spécialisés dans le domaine du cancer, ont des indices d'impact entre 0,8 et 0,9. L'Inde qui avait l'indice le plus faible les a rejoints en fin de période.

### Publications les plus citées au monde

L'indice d'activité d'un pays dans le décile des publications les plus citées au monde est complémentaire à l'impact moyen. Il est défini comme le ratio entre la part des publications d'un pays dans le décile et la part des publications mondiales dans cette même classe. Un indice d'activité supérieur à 1 signifie que le pays a une proportion plus importante de publications parmi les 10 % des publications les plus citées que celle du monde.

Le graphique 8 confirme les observations relatives à l'indice d'impact moyen. La dispersion entre les 10 pays publiant le plus dans le domaine du cancer est similaire. Les Etats-Unis et le Royaume-Uni présentent les indices les plus élevés, à 1,3 en fin de période. L'indice du Canada se tasse pour rejoindre la valeur neutre de 1. Il est rejoint par la France, l'Italie et l'Allemagne qui avaient un indice d'activité inférieur à la moyenne en 2000. Le Japon et la Corée du Sud ont un indice d'activité faible qui oscille entre 0,6 et 0,8 durant la période. La Chine améliore considérablement son indice d'activité, de 0,6 à 1.

**Graphique 8. Indices d'activité des dix premiers pays publiant dans le domaine du cancer, 2000-2019**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

## 5. Analyse des publications par type de recherche

### Contexte et problématique

L'objectif initial était d'identifier et de caractériser les publications scientifiques dans le domaine du cancer en distinguant les travaux de recherche fondamentale, les travaux de recherche clinique et les travaux de recherche translationnelle. L'OST a animé, avec la Direction du Pôle Recherche de l'INCa, plusieurs réunions avec le groupe d'experts du cancer associés à cette étude afin de définir les contours de ces trois types de recherche.

Les experts ont souligné les difficultés liées à la définition de ces sous-domaines. La recherche clinique semblait possible à appréhender avec des mots-clés du type « essais cliniques », « cohorte de patients ». La recherche translationnelle peut en revanche répondre à plusieurs définitions selon qu'elle se positionne en complément de travaux de recherche fondamentale ou en complément de travaux de recherche clinique. Par ailleurs, les experts ont souligné que cette distinction est de moins en moins pertinente, puisque les travaux tendent aujourd'hui à s'hybrider : près de la moitié des articles de recherche fondamentale contiennent une partie translationnelle ou préclinique. De même, les articles de recherche clinique sont souvent associés à des travaux de recherche translationnelle.

La direction du Pôle Recherche de l'INCa adhère aux remarques des experts sur la typologie des recherches sur le cancer et est consciente de leur potentiel recouvrement. Néanmoins, elle a souhaité distinguer les publications scientifiques selon ces trois types car ils correspondent à des programmes de financement différents de l'INCa.

### Méthode d'élaboration des sous-corpus

#### Recherche Fondamentale

L'OST a défini le sous-corpus recherche fondamentale comme étant les publications du corpus cancer qui sont rattachées à la grande discipline « BIOLOGIE FONDAMENTALE ». Ce sous-corpus correspond à 574 970 publications – 24 % du corpus cancer en 2000 et 21 % en 2020. Entre 30 % et 50 % de ces publications selon les années sont aussi rattachées à la grande discipline « RECHERCHE MEDICALE ».

Le groupe d'experts a examiné un échantillon de 100 publications de ce sous-corpus. Il cible bien les publications de recherche fondamentale à 80 %. Environ 15 % de publications pourraient aussi être classées en recherche translationnelle et 5 % en recherche clinique. Le groupe d'experts a ensuite examiné un ensemble de publications hors de ce sous-corpus. Ces publications ne relevaient pas de la recherche fondamentale ou alors il s'agissait de publications à la frontière avec la recherche translationnelle.

Compte tenu du recoupement entre recherche translationnelle et recherche fondamentale et du degré de précision attendu, ce sous-corpus est apparu satisfaisant.

Le détail des spécialités scientifiques prises en compte pour définir la grande discipline « BIOLOGIE FONDAMENTALE » est précisé en annexe 2.

#### Recherche Clinique

Le processus de construction du corpus de publications relevant de la recherche clinique a été itératif.

Deux corpus ont été construits :

- Un premier corpus défini à partir d'une liste de mots-clés caractéristiques issue de l'étude réalisée par l'OST en 2017 pour le compte de l'INCa
- Un deuxième corpus défini à partir d'une requête MeSH rassemblant les publications du corpus cancer typées « Clinical Trial » [Publication Type] et les publications identifiées par le terme MeSH «Clinical Trials as Topic».

Le premier corpus rassemble près de 50 % du corpus cancer et a été considéré par les experts comme beaucoup trop large. Certains mots-clés comme PATIENT ou CLINICAL ramènent beaucoup de bruit (publications hors recherche clinique). Le deuxième corpus rassemble 6 % des publications et le groupe d'experts a estimé que la requête MeSH est trop restrictive car elle est focalisée sur les publications relatives à un essai clinique. Or, la recherche clinique est beaucoup plus large : les essais cliniques correspondent à un cadre très spécifique et règlementé. Une partie de la recherche clinique qui n'implique pas un essai clinique n'est donc pas prise en compte dans cette requête. En particulier, les experts ont suggéré d'intégrer à la recherche clinique les travaux réalisés sur de nouvelles applications techniques (imagerie, chirurgie) et a proposé pour cela d'ajouter des mots-clés comme CANCER SURGERY et TUMOR SURGERY.

A l'issue de cette analyse, un troisième corpus a été défini en combinant les deux méthodes : une requête textuelle s'appuyant sur une version actualisée de la liste de mots-clés issue de la requête de 2017 (suppression ou précision de certains termes trop génériques et ajout de termes complémentaires) et la requête utilisant les champs MeSH. Un échantillon complémentaire de publications hors de ce troisième corpus de publications a été analysé par les experts, ce qui a permis d'identifier des publications de recherche clinique non prises en compte et de compléter la liste de mots-clés.

Le corpus définitif a été construit en combinant la requête MeSH et la liste de mots-clés actualisée<sup>1</sup> et rassemble 476 403 publications, soit environ 18 % du corpus cancer.

### Recherche Translationnelle

Il est difficile d'obtenir une définition de la recherche translationnelle qui fasse consensus. La recherche translationnelle est à la frontière entre différents types de recherche et le point de vue est différent selon le domaine de recherche. [4,5]

Le groupe de travail a proposé de constituer un corpus spécifique sur l'interface entre recherche fondamentale et recherche clinique qui correspondrait à la recherche pré-clinique reposant sur un test/ échantillon humain (et non animal). Une requête textuelle a pour cela été définie avec une dizaine de mots-clés et a permis de construire un corpus correspond à 119 671 publications (environ 4 % du corpus cancer). La grande majorité des publications de ce corpus correspondent bien au domaine recherché, toutefois, la définition retenue est très restrictive et ne saurait représenter l'ensemble de la recherche translationnelle. Il a été décidé de ne pas présenter d'indicateurs sur ce sous-corpus.

## **Principaux résultats**

Les principaux indicateurs présentés ci-dessous concernent les corpus de publications en recherche fondamentale et en recherche clinique. Ces deux corpus correspondent à environ 40 % du corpus cancer global. Les publications restantes représentant environ 60 % du corpus correspondent notamment à la recherche translationnelle, à la recherche en épidémiologie ou encore en santé publique. Toutefois ces différents domaines n'étant pas totalement disjoints, il est probable qu'il reste des publications de recherche clinique et de recherche fondamentale hors de ces deux corpus et les indicateurs doivent être interprétés avec prudence.

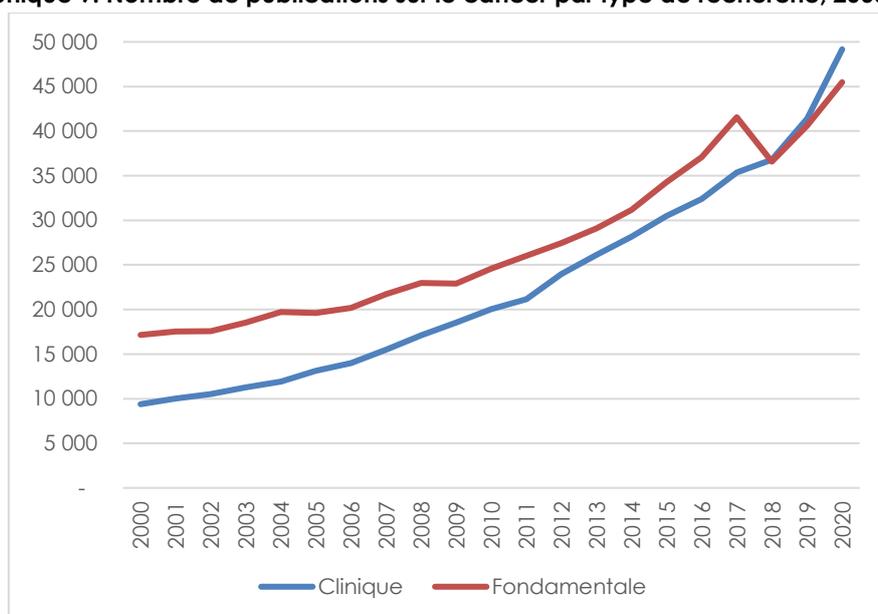
### **1. Nombre de publications**

Le volume de publications en recherche clinique et en recherche fondamentale augmente fortement de 2000 à 2020 (graphique 9), la croissance étant plus importante pour les publications de recherche clinique qui passent de 13 % à 23 % du corpus Cancer alors que la part des publications de recherche fondamentale se stabilise autour de 21 % du corpus Cancer (graphique 10).

---

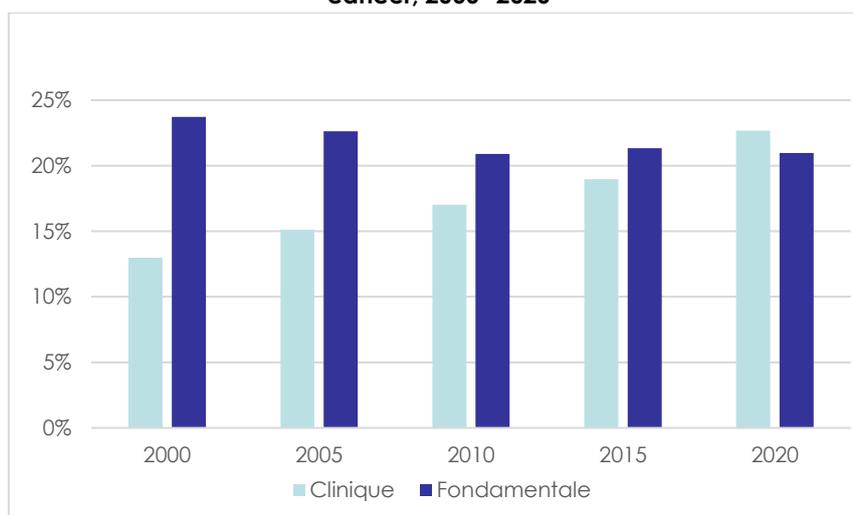
<sup>1</sup> La liste des mots-clés actualisée utilisée pour cette requête est précisée en annexe.

**Graphique 9. Nombre de publications sur le cancer par type de recherche, 2000-2020**



Source: Base OST, Web of Science, calculs OST. Compte entier<sup>2</sup> - 2020 : année complète à 95 %

**Graphique 10. Part de la recherche clinique et de la recherche fondamentale dans les publications sur le cancer, 2000- 2020**



Source: Base OST, Web of Science, calculs OST. Compte entier - 2020 : année complète à 95 %

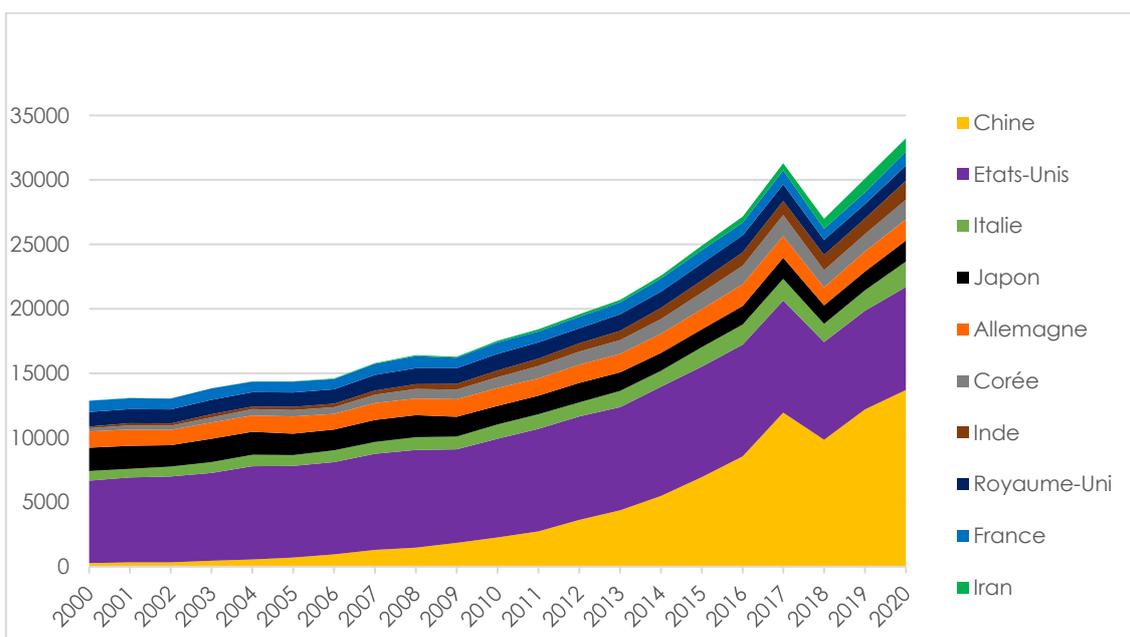
Le positionnement des premiers pays publiant en recherche fondamentale / clinique dans le domaine du cancer est similaire au positionnement des pays publiant dans le domaine du cancer, tous travaux confondus (voir section 2 du rapport).

Les graphiques 11 et 12 permettent de visualiser notamment la croissance très rapide de la Chine sur les deux types de recherche.

L'ordre des pays est similaire à l'ordre sur l'ensemble des publications sur le cancer. On peut toutefois noter quelques différences : dans le domaine de la recherche fondamentale, l'Italie publie plus que le Japon et l'Inde et l'Iran sont aussi mieux positionnés. Dans le domaine de la recherche clinique, le Royaume-Uni publie plus que la Corée du Sud et l'Espagne apparaît dans les 10 premiers pays publiant devant le Canada.

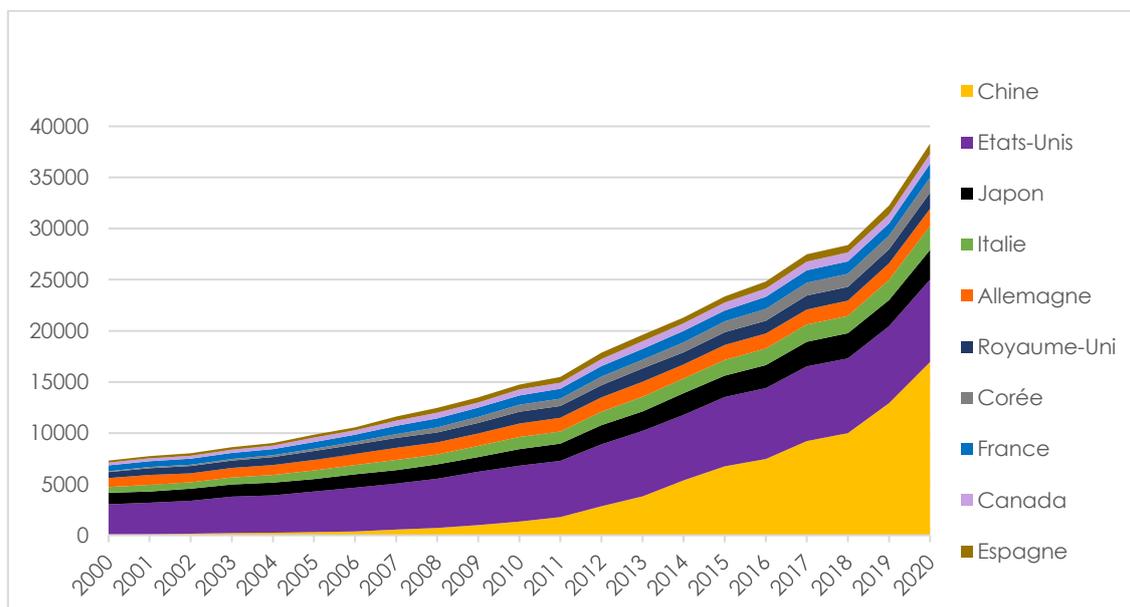
<sup>2</sup> Les comptes entiers et comptes fractionnaires sont définis dans l'annexe 1

**Graphique 11. Nombre de publications sur le cancer en recherche fondamentale des 10 premiers pays publiant sur le cancer, 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

**Graphique 12. Nombre de publications sur le cancer en recherche clinique des 10 premiers pays publiant sur le cancer, 2000-2020**

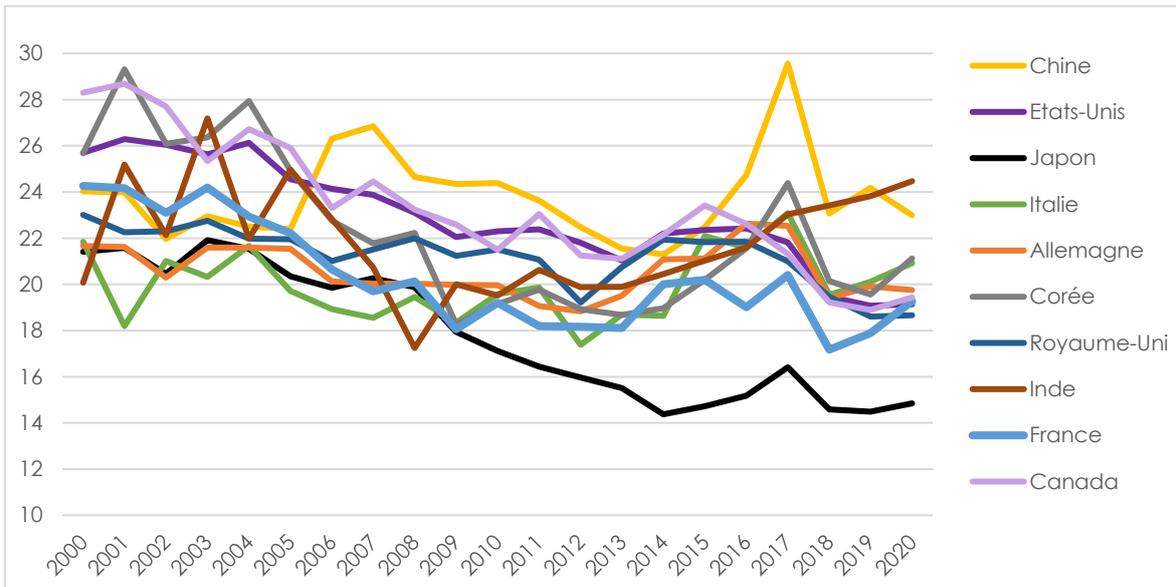


Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

## 2. Profil des pays par type de recherche sur le cancer

Les graphiques 13 et 14 permettent de visualiser le poids de chaque type de recherche au sein des publications des dix premiers pays publiant dans le domaine du cancer. La part de la recherche fondamentale dans la recherche sur le cancer après une légère décroissance de 2000 à 2010 est stabilisée dans la plupart des pays, sauf au Japon où elle continue à décroître.

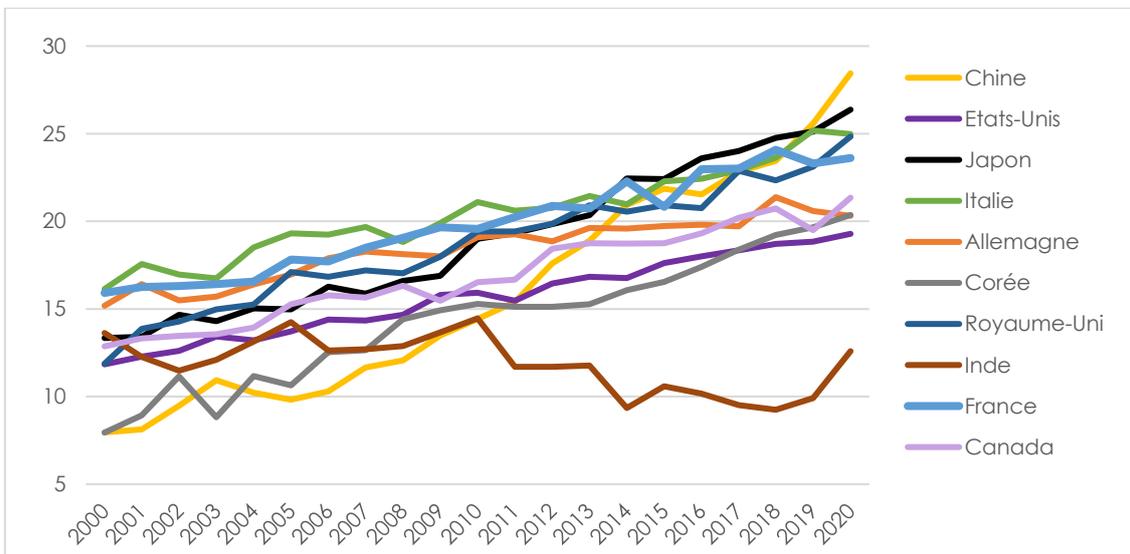
**Graphique 13. Part en % des publications en recherche fondamentale sur le cancer, 10 premiers pays publiants, 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

La part de la recherche clinique progresse dans tous les pays sauf en Inde. Cette part progresse particulièrement en Chine où elle passe de 8 à 28 %, soit la part la plus élevée de l'échantillon de pays. Cette part augmente au Japon et au Royaume-Uni passant de 13 % (Resp. 12 %) en 2000 à 26 % (Resp. 25 %) en 2020. L'Italie et la France avec une part de recherche clinique déjà très importante en début de période conservent une part importante de recherche clinique avec 25 % et 24 % de publications en recherche clinique en 2020. Au contraire l'Allemagne, les Etats-Unis et le Canada ont une part de recherche clinique qui augmente moins vite que les autres pays.

**Graphique 14. Poids de la recherche clinique dans la production scientifique sur le cancer des 10 premiers pays publiants, 2000-2020**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

Les indices de spécialisation permettent de préciser les résultats précédents. Les pays fortement spécialisés en recherche fondamentale dans le domaine du cancer sont la Chine, l'Inde, l'Iran, la Pologne, Taiwan et dans une moindre mesure l'Italie. La spécialisation en recherche fondamentale de la plupart des pays européens, ainsi que des Etats-Unis et du Canada, est en baisse (tableau 5).

**Tableau 5 : Indice de spécialisation en recherche fondamentale des 20 premiers pays publiant dans le domaine du cancer, 2000-2020**

	2001-2005	2006-2010	2011-2005	2016-2020
Chine	0,99	1,17	1,05	1,1
Etats-Unis	1,09	1,07	1,05	0,91
Japon	0,9	0,82	0,69	0,71
Italie	0,87	0,94	1,04	1
Allemagne	0,95	0,96	0,99	0,94
Corée du Sud	1,1	0,92	0,95	1,01
Royaume-Uni	0,97	1,03	1,02	0,89
Inde	1,1	0,93	0,99	1,17
France	0,98	0,92	0,95	0,92
Canada	1,14	1,03	1,1	0,93
Espagne	1,04	1	1,12	0,99
Australie	1,1	1,1	1,04	0,98
Iran	0,77	0,72	1,12	1,33
Turquie	0,49	0,53	0,51	0,65
Brésil	1,09	0,96	1,05	1,03
Pays-Bas	0,85	0,91	0,82	0,75
Taiwan	0,77	0,96	1,11	1,13
Pologne	1,13	1,06	1,02	1,41
Suisse	1,04	1,01	1,09	0,98
Suède	1,03	1,29	1,05	0,92

Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

La Chine et le Japon voient leur spécialisation en recherche clinique augmenter très fortement pour devenir les pays les plus spécialisés en 2020 (tableau 6). Les Pays-Bas maintiennent une spécialisation nette, juste inférieure à celle du Japon. Les autres pays historiquement spécialisés en recherche clinique comme l'Italie, le Royaume-Uni, la France et l'Espagne conservent une spécialisation positive, mais moindre. Au contraire, la spécialisation en recherche clinique régresse en-dessous de la référence mondiale (1) pour l'Allemagne, le Canada, la Turquie, le Brésil et Taiwan.

Les Etats-Unis ne sont spécialisés ni en recherche fondamentale, ni en recherche clinique. En effet, en 2020, plus de 60 % de leurs publications sur le cancer ne concernent ni la recherche clinique ni la recherche fondamentale alors que cette part de publications hors domaine est moindre pour les autres pays, surtout la Chine dont moins de 50 % des publications sont rattachées hors des domaines clinique ou fondamental.

**Tableau 6 : Indice de spécialisation en recherche clinique des 20 premiers pays publiants dans le domaine du cancer, 2000-2020**

	2001-2005	2006-2010	2011-2005	2016-2020
Chine	0,65	0,85	1,15	1,25
Etats-Unis	0,91	0,94	0,93	0,85
Japon	0,99	1,12	1,18	1,16
Italie	1,28	1,24	1,17	1,1
Allemagne	1,12	1,12	1,04	0,9
Corée du Sud	0,7	0,9	0,87	0,9
Royaume-Uni	1,13	1,14	1,1	1,1
Inde	0,94	0,85	0,56	0,56
France	1,18	1,15	1,1	1,04

Canada	1,01	0,97	0,99	0,94
Espagne	1,04	0,98	0,99	1,03
Australie	0,98	0,9	0,96	0,99
Iran	0,95	0,83	0,67	0,75
Turquie	1,56	1,26	0,92	0,81
Brésil	1,03	1,03	0,92	0,89
Pays-Bas	1,13	1,12	1,29	1,13
Taiwan	1,05	0,87	0,83	0,9
Pologne	0,84	0,98	0,85	0,74
Suisse	0,9	1,08	1,03	1,01
Suède	0,83	0,91	1,02	0,95

Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

Ces indicateurs de spécialisation sont à interpréter avec précaution car les corpus de recherche fondamentale et de recherche clinique ne correspondent pas à la totalité de la production dans le domaine du cancer mais 40 à 50 % des publications sur le cancer selon les pays.

La spécialisation observée de la Chine à la fois en recherche fondamentale et en recherche clinique, signifie que la Chine publie une proportion moindre d'articles d'autres domaines (recherche translationnelle, épidémiologie, santé publique). Symétriquement, la baisse de spécialisation des Etats-Unis en recherche fondamentale et en recherche clinique traduit probablement un développement fort de publications dans ces domaines alternatifs.

### 3. Impact des publications

Les indices d'impact des différents pays en recherche fondamentale et en recherche clinique (graphiques 15 et 16) montrent des résultats déjà observés sur l'ensemble de la recherche sur le cancer avec quelques nuances en particulier sur la Recherche Clinique.

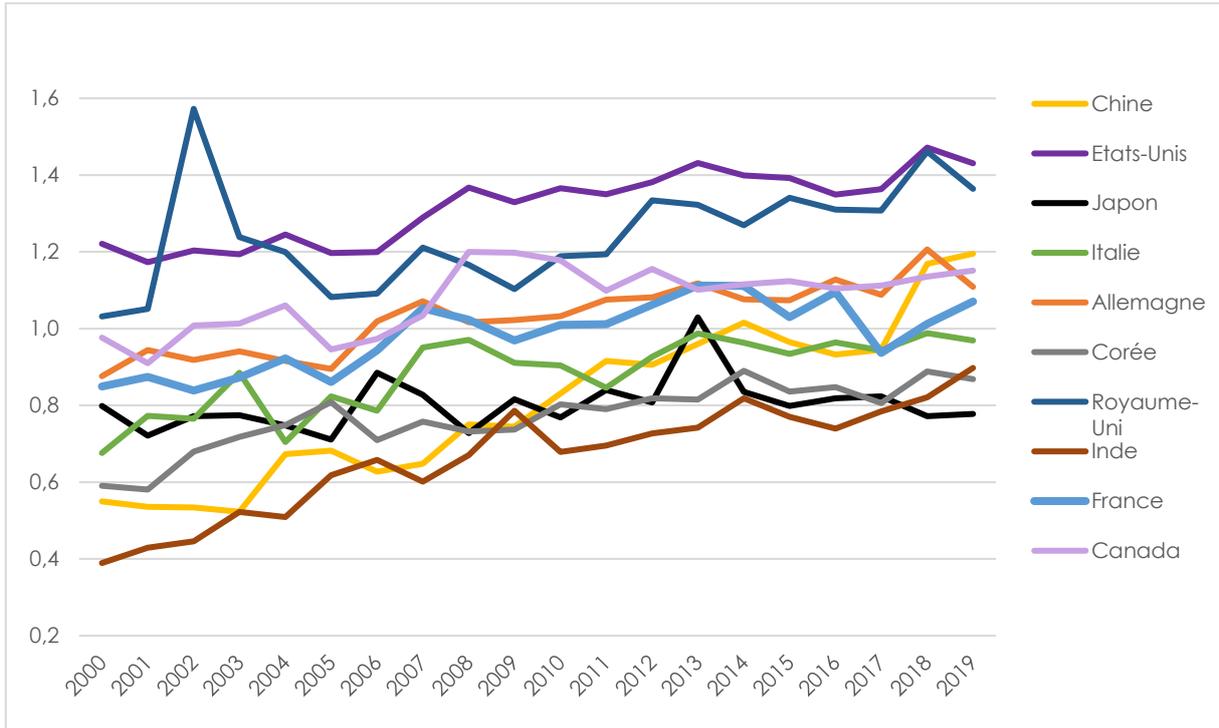
On retrouve en tête les Etats-Unis suivis du Royaume-Uni avec des indices d'impact nettement supérieurs à 1 depuis 20 ans, en recherche fondamentale comme en recherche clinique. Le Canada qui est en troisième position pour l'ensemble de la recherche sur le cancer est mieux positionné sur la recherche clinique, domaine dans lequel son indice d'impact, compris entre 1,2 et 1,4, est très proche de celui des Etats-Unis et du Royaume-Uni.

La France, l'Allemagne et l'Italie progressent régulièrement en termes d'impact, mais cette progression est plus importante en recherche clinique, en particulier pour la France qui rejoint le trio de tête Etats-Unis, Royaume-Uni, Canada avec un indice d'impact supérieur à 1,3.

La Chine, qui avait un indice d'impact inférieur à 0,6 en recherche fondamentale en 2000 progresse fortement, comme pour l'ensemble du domaine cancer, pour dépasser la moyenne mondiale à partir de 2018. En revanche dans le domaine de la recherche clinique, l'indice d'impact de la Chine se stabilise autour de 0,8 à partir de 2005 et ne progresse pas.

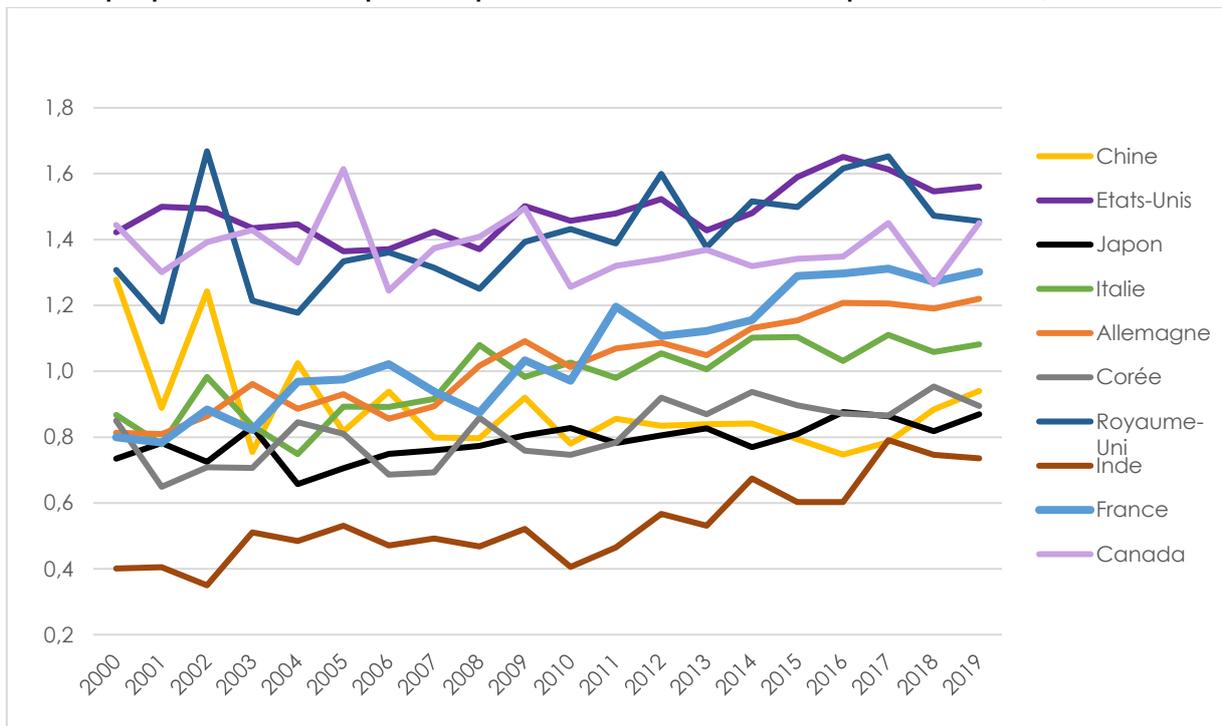
Comme pour l'ensemble du domaine cancer, le Japon et la Corée du Sud ont des indices d'impact faibles qui restent inférieurs à 1, en recherche fondamentale comme en recherche clinique.

**Graphique 15. Indice d'impact des publications en recherche fondamentale sur le cancer, 2000-2019**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

**Graphique 16. Indice d'impact des publications en recherche clinique sur le cancer, 2000-2019**



Source : Base OST, Web of Science, calculs OST. Compte fractionnaire - 2020 : année complète à 95 %

## Annexe 1 : Données et Méthodologie

### A) Base de données bibliométriques

Les publications dans les revues scientifiques constituent pour les chercheurs un des principaux modes de validation et de diffusion de leurs travaux. Les notices qui décrivent ces publications sont enregistrées dans des bases de données bibliométriques qui comportent des informations sur les sources (revue, auteurs, laboratoires et affiliations institutionnelles), les contenus scientifiques (titre, mots-clés, résumé) et les liens de contenu scientifique avec d'autres publications (références, citations).

Les indicateurs bibliométriques de la présente étude sont calculés à partir de l'information contenue dans les notices de la base de publications de l'OST. Cette dernière est une version *off-line* du *Web of Science*® (WoS) de Clarivate Analytics, qui est une base de référence mondiale en bibliométrie [6]. Le *Web of Science* recense les revues scientifiques et les actes de colloques les plus influents au niveau international.

Elle privilégie les publications académiques, particulièrement les articles de recherche. Elle est plus représentative pour les disciplines bien internationalisées que pour les disciplines appliquées, de « terrain » ou à forte tradition nationale. Elle est ainsi moins représentative pour certains domaines comme les sciences pour l'ingénieur, les sciences humaines et les sciences sociales. Néanmoins, la couverture de la base évolue et de nouvelles revues y sont intégrées chaque année suivant le processus de sélection mis en place par Clarivate Analytics. Ainsi, le *Web of Science*® ne doit pas être considéré comme un échantillon représentatif de la production scientifique mondiale mais plutôt comme un recensement raisonné des meilleures revues de niveau international.

L'OST enrichit la base *Web of Science* avec des données géographiques et institutionnelles. La version de la base OST mobilisée dans ce rapport correspond à celle issue de la dernière actualisation qui contient les données WoS arrêtées à la 18<sup>ème</sup> semaine de l'année 2022.

### B) Disciplines et spécialités de la base WoS

La classification disciplinaire dans la base OST correspond aux domaines de recherche établis par Clarivate Analytics. Tous les domaines sont couverts par 254 « subject categories »<sup>3</sup>. L'affectation des articles aux onze disciplines scientifiques/académiques dans la base OST est basée sur une agrégation de ces catégories ou spécialités. Les onze disciplines sont :

- Biologie appliquée-écologie,
- Biologie fondamentale,
- Chimie,
- Informatique,
- Mathématiques,
- Physique,
- Recherche médicale,
- Sciences de l'univers,
- Sciences humaines,
- Sciences pour l'ingénieur,
- Sciences sociales.

---

<sup>3</sup> la liste des 254 subject categories, est détaillée dans le lien suivant [http://images.webofknowledge.com/WOKRS5251R3/help/WOS/hp\\_subject\\_category\\_terms\\_tasca.html](http://images.webofknowledge.com/WOKRS5251R3/help/WOS/hp_subject_category_terms_tasca.html)

Les revues peuvent être rattachées à plusieurs domaines de recherche et donc, par agrégation, à plusieurs disciplines OST. Les articles publiés dans des journaux multidisciplinaires, comme *PNAS*, *Nature et Science* notamment, sont réaffectés aux disciplines sur la base de l'appartenance disciplinaire des références citées dans ces articles. Toutefois, si les références citées sont majoritairement multidisciplinaires ou s'il n'y a pas une dominance disciplinaire dans les références (les références citées sont réparties équitablement entre plusieurs disciplines), les articles ne sont pas réaffectés. Ce reliquat est regroupé dans une « douzième discipline », appelée « Multidisciplinaire », qui a un poids très faible.

Le tableau suivant fournit la correspondance pour le domaine biomédical uniquement.

Disciplines	Catégories	
<b>BIOLOGIE FONDAMENTALE</b>	ANATOMY & MORPHOLOGY BIOCHEMISTRY & MOLECULAR BIOLOGY MEDICAL INFORMATICS MATHEMATICAL & COMPUTATIONAL BIOLOGY EVOLUTIONARY BIOLOGY CELL BIOLOGY MATERIALS SCIENCE, BIOMATERIALS BIOCHEMICAL RESEARCH METHODS BIOPHYSICS BIOTECHNOLOGY & APPLIED MICROBIOLOGY DEVELOPMENTAL BIOLOGY GENETICS & HEREDITY ENGINEERING, BIOMEDICAL CELL & TISSUE ENGINEERING	MICROBIOLOGY MICROSCOPY NEUROIMAGING NEUROSCIENCES NUTRITION & DIETETICS PARASITOLOGY PHYSIOLOGY PSYCHOLOGY BEHAVIORAL SCIENCES REPRODUCTIVE BIOLOGY MEDICAL LABORATORY TECHNOLOGY VIROLOGY
<b>RECHERCHE MEDICALE</b>	ALLERGY ANDROLOGY ANESTHESIOLOGY AUDIOLOGY & SPEAK-LANGUAGE PATHOLOGY ONCOLOGY  CHEMISTRY, MEDICINAL  SURGERY DERMATOLOGY ENDOCRINOLOGY & METABOLISM MEDICAL ETHICS GASTROENTEROLOGY & HEPATOLOGY GERIATRICS & GERONTOLOGY OBSTETRICS & GYNECOLOGY HEMATOLOGY  IMMUNOLOGY  INFECTIOUS DISEASES CARDIAC & CARDIOVASCULAR SYSTEMS PERIPHERAL VASCULAR DISEASE  PRIMARY HEALTH CARE  SUBSTANCE ABUSE SPORT SCIENCES EMERGENCY MEDICINE MEDICINE, RESEARCH & EXPERIMENTAL INTEGRATIVE & COMPLEMENTARY MEDICINE	MEDICINE, GENERAL & INTERNAL MEDICINE, LEGAL TROPICAL MEDICINE VETERINARY SCIENCES CLINICAL NEUROLOGY DENTISTRY, ORAL SURGERY & MEDICINE OPHTHALMOLOGY ORTHOPEDICS OTORHINOLARYNGOLOGY PATHOLOGY PEDIATRICS PHARMACOLOGY & PHARMACY RESPIRATORY SYSTEM PSYCHIATRY RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING REHABILITATION RHEUMATOLOGY HEALTH CARE SCIENCES & SERVICES PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH NURSING CRITICAL CARE MEDICINE TOXICOLOGY TRANSPLANTATION UROLOGY & NEPHROLOGY

### C) Périmètre de calcul des indicateurs

Les indicateurs sont calculés en ne retenant que certains types de documents : les articles originaux (y compris ceux issus des actes de colloques), les articles de synthèse (*reviews*) et les *proceedings papers*. Par ailleurs la procédure standard de traitement des données bibliométriques appliquée à l'OST écarte les documents pour lesquels certaines informations sont manquantes. En particulier les documents pour lesquels le code pays, les liens auteur-adresse ou la catégorie disciplinaire manquent ne sont pas pris en compte.

## D) Les méthodes de décompte

Une publication scientifique peut comporter plusieurs lignes d'adresses de laboratoires, étant produite par des chercheurs de laboratoires différents. Par ailleurs la publication peut hériter de plusieurs disciplines scientifiques associées à son support éditorial. La question du décompte de la publication pour chacun des acteurs contributeurs (laboratoire, institution, région, pays) se pose logiquement. Deux types de comptage sont généralement considérés : le compte entier (ou compte de présence) et le compte fractionnaire.

Le compte entier consiste à créditer une publication chacune des entités signataires. De même, si la publication est indexée dans deux domaines de recherche, elle comptera pour 1 dans chacun des domaines. Le compte entier renvoie à une logique de participation d'une entité à la publication ou encore de présence dans un domaine de recherche. Dans la mesure où chaque publication est comptée autant de fois qu'il y a de signataires, le compte entier n'est pas additif.

Le compte fractionnaire reflète une logique de contribution des signataires à la publication scientifique. Une fraction de la publication est attribuée à chaque entité signataire de manière à avoir une somme unitaire. Le compte fractionnaire est additif à toutes les échelles et pour tous les niveaux de nomenclature.

La mise en œuvre du compte fractionnaire peut être illustrée en considérant une publication co-signée par deux auteurs affiliés à deux institutions, l'une en France et l'autre en Allemagne. Cette publication peut par ailleurs être affectée dans la base de données à trois domaines de recherche et deux disciplines, par exemple, « Biochimie, biologie moléculaire » et « Génétique, hérédité » au sein de la discipline biologie fondamentale et « Biotechnologie et microbiologie appliquée » au sein de la discipline biologie appliquée.

Dans cet exemple, le fractionnement de la publication va porter à la fois sur la dimension géographique et sur la dimension disciplinaire. Chacun des deux pays se voit attribuer une moitié (1/2) de la publication. D'un point de vue disciplinaire, la publication compte pour 2/3 pour la biologie fondamentale et 1/3 pour la biologie appliquée.

Le fractionnement total combine les deux fractions établies précédemment pour tenir compte des combinaisons pays/domaine ou pays/discipline. La part qui revient à la paire pays/discipline est le produit des fractions géographiques et disciplinaires. Dans l'exemple, le document va compter 1/6 pour chacune des deux paires pays/ biologie appliquée et 2/6 pour chacune des deux paires pays/ biologie fondamentale. Ainsi, la France comme l'Allemagne se verront attribuer  $1/6 + 2/6 = 1/2$ .

## E) Les indicateurs bibliométriques

### Nombre de publications

Cet indicateur donne le volume de publications pour une entité donnée à un niveau disciplinaire donné et pour une période donnée.

### Part de publications

Pour un pays et une année de publication P donnés, la part de publications dans une discipline ou un domaine de recherche est définie par son nombre de publications rapporté au nombre de publications parues pour une référence dans la même discipline ou le même domaine de recherche. Cet indicateur représente le poids de du pays dans le total de la référence (monde). L'indicateur « *part de publications* » est défini par

$$\text{Part de publications (\%)} = \frac{\text{Nombre de publications de du pays, année P}}{\text{Nombre de publications du monde, année P}} \times 100$$

### Spécialisation scientifique

L'indice de spécialisation scientifique d'un pays dans une discipline est défini par la part de la discipline dans l'ensemble de ses publications, normalisée par cette même part dans les publications mondiales.

Plus l'indice de spécialisation est supérieur à 1, plus le pays est dit « spécialisé » dans la discipline considérée. L'indice est défini par

$$\text{Indice de spécialisation} = \frac{\text{Part des publications du pays dans une discipline}}{\text{Part des publications du monde dans la même discipline}}$$

### Impact scientifique

Les indicateurs d'impact s'appuient sur les citations que les articles scientifiques font de publications antérieures. Les pratiques de citation étant différentes entre domaines de recherche et entre disciplines, une normalisation est nécessaire pour corriger cet effet discipline et pour pouvoir faire des comparaisons entre disciplines. L'indicateur doit aussi tenir compte de la durée entre la date de publication de l'article et celles de ses citations.

#### *Impact moyen*

L'impact moyen des publications d'un pays est défini par le nombre moyen des citations normalisées de ses publications dans un domaine, rapporté à la moyenne des citations normalisées des publications mondiales de ce domaine. La normalisation est faite par catégorie ou spécialité WoS et permet de tenir compte de la structure disciplinaire de la production de chaque pays.

Un indice d'impact de 1 signifie que l'impact moyen des publications dans une discipline est égal à celui de la moyenne des publications mondiales dans la discipline. Lorsque l'indice est supérieur à 1, les publications de l'entité ont en moyenne un impact plus élevé. A contrario, un indice d'impact inférieur à 1 implique que les publications de l'entité ont en moyenne un impact plus faible que la moyenne mondiale.

### Indice d'activité

La distribution des publications se fait dans les classes de citations définies au niveau mondial par spécialité WoS. Elles correspondent à des découpages de l'ensemble des publications en percentiles décroissants en fonction du nombre de citations reçues au niveau mondial. On distingue la classe des 1 % des publications les plus citées au monde, la classe des 5 %, la classe des 10 %, etc.

L'indice d'activité de chaque classe de citations est égal au ratio entre la part des publications d'un pays dans la classe et la part des publications mondiales dans cette classe. Par construction, la valeur de l'indice d'activité est égale à 1 dans chaque classe pour le monde. Un indice d'activité supérieur à 1 signifie que le pays a une proportion plus importante de publications que celle du monde dans la classe considérée.

### Indicateurs relatifs aux publications en accès ouvert

La publication en accès ouvert est diversement pratiquée selon les domaines scientifiques. C'est pourquoi, outre la part de publications en accès ouvert dans le total, un indice normalisé est fourni. Il permet de comparer les pays entre eux malgré des profils disciplinaires différents.

L'indice d'accès ouvert est défini par la part de publications d'un pays en accès ouvert dans un domaine rapporté à la même part pour le total des publications mondiales. La valeur neutre de l'indice d'accès ouvert est 1. Lorsque l'indice est supérieur à 1, la production scientifique du pays dans le domaine est plus ouverte que la moyenne de celle du monde dans le même domaine.

### Part de co-publications internationales

Une co-publication est une publication ayant au moins deux adresses d'affiliation différentes. Une publication sans collaboration ne comporte qu'une seule adresse d'affiliation.

Une co-publication internationale est une publication dont au moins une adresse d'affiliation est située à l'étranger. La part des co-publications internationales est définie par le nombre de publications du pays qui sont signées par au moins un auteur ayant une adresse à l'étranger, rapporté au nombre total des publications du pays.

## Annexe 2 : Méthode détaillée de constitution du corpus Cancer

La sélection de documents couvrant la thématique du cancer est effectuée en plusieurs étapes.

### **Étape 1. Construction et analyse du premier corpus cancer (V0)**

L'OST procède par interrogation du contenu sémantique des documents de sa base de publications scientifiques, à partir d'une liste établie de mots-requêtes. Ces derniers sont des termes ou des combinaisons de termes utilisés pour réaliser une requête lexicale.

Trois sources de requêtes établies antérieurement ont été mobilisées :

- une requête OST utilisée dans l'étude réalisée en 2017 pour l'INCa ;
- une requête mobilisée dans le rapport global du CNCR [3] sur le cancer, paru en février 2021 ;
- une requête PubMed [7] limitée au MeSH Term [8] « Neoplasms » et aux « subheadings » associés.

Chaque requête a été utilisée séparément pour constituer un ensemble de publications scientifiques liées à la thématique cancer limité à la période 2000-2020. Toutes les publications comportant au moins un des mots-requêtes de la thématique dans le titre et/ou le résumé et/ou les mots-clés auteurs, constituent le corpus de publications sur le cancer.

Les trois corpus de documents résultant ont été fusionnés pour constituer un corpus cancer global (V0) de 2 771 013 documents. Ce corpus global (V0) résulte donc d'une liste de 406 mots-clés/mots-requêtes.

Les experts associés au projet ont été sollicités pour analyser 3 échantillons de publications scientifiques du corpus V0 : sur les 300 publications analysées par les experts, 20 publications ont été considérées comme hors du périmètre cancer, ce qui correspond à un taux de précision de 93 % et qui est très satisfaisant au vu de la littérature sur le sujet. Ce bon taux de précision résulte du fait que la requête s'appuie notamment sur deux requêtes validées par des experts de domaines (la requête du CNCR [1] et la requête de PubMed). Le « bruit » de cet échantillon est d'ailleurs pour l'essentiel rapporté par la requête OST de 2017.

Cette première analyse montre donc que les mots-clés utilisés permettent dans l'ensemble de sélectionner des documents pertinents.

Par ailleurs les experts associés ont aussi été sollicités sur les mots-requêtes utilisés. Les remarques des experts sur les mots-requêtes ont amené à modifier la liste initiale.

Six mots-requêtes jugés non pertinents ont été supprimés :

- ASCO – FUNGI ;
- AACR ;
- AJCC ;
- ICD O ;
- UICC ;
- IARC.

Et les mots-clés redondants ont été supprimés également. La liste ainsi modifiée comporte 371 mots-requêtes.

### **Étape 2. Premier ajustement du corpus cancer et analyse du silence**

Le corpus brut a été complété par les notices de la catégorie WoS ONCOLOGY qui ne sont pas ramenées par les mots-clés. Sur les 734 684 documents de cette catégorie, pour la période 2000-2020, la plupart figure déjà dans le corpus global. Le complément porte seulement sur 40 760 publications scientifiques. Le corpus ajusté (V1) comporte alors 2 811 773 publications.

Pour juger la pertinence de ce corpus, le silence (soit les documents pertinents qui n'y sont pas) a été analysé. À cette fin, un ensemble de publications scientifiques, sur la période de l'étude, associées à des projets financés par l'INCa, a été identifié dans la base WoS. Cet ensemble de publications a été constitué en utilisant les 'grant numbers' des projets INCa fournis par l'INCa, ainsi que le champ remerciement du WoS. Un total de 8 031 documents issus de projets financés ou cofinancés par l'INCa a été ainsi identifié ; 1 876 de ces 8 031 publications ne figurent pas dans le corpus ajusté, ce qui correspond à un silence de 23 %.

Pour déterminer enfin si ces 1 876 publications concernent bien la thématique du cancer et les mots clés avec lesquels il est possible de les récupérer dans le corpus global, trois analyses ont été faites :

- analyse d'un échantillon de publications issues de ces 1 876 publications afin de déterminer si elles portent bien sur le cancer et concernent bien des projets financés par l'INCa ;

- analyse des termes utilisés dans les titres des publications pour voir si les sujets abordés ont un lien avec le cancer ;
- analyse des mots clés auteurs des publications pour voir si les sujets abordés ont un lien avec le cancer.

L'examen par les experts des cartographies et des échantillons de publications a permis d'identifier des catégories de publications qui sont liées à des projets financés par l'INCa et qui ne figurent pas dans le corpus V1. Il s'agit d'articles de biologie fondamentale, de méthodologie (biostatistiques en particulier), d'épidémiologie et de prévention.

### **Étape 3. Deuxième ajustement du corpus cancer : extension pour intégrer des publications manquantes**

- **Ajustement de la liste des mots clés**

Tout d'abord, les requêtes ont été complétées pour les mots-requêtes caractéristiques suivants, fournis par les experts

- (CHIP AND CLONAL) OR (CLONAL HEMATOPOIESIS OF INDETERMINATE POTENTIAL) OR CTC-CHIP
- (CLONAL HEMATOPOIESIS OR CLONAL HAEMATOPOIESIS) NOT (CARDIO% OR ATHERO%)
- (ARCH AND CLONAL) OR AGE\ -RELATED CLONAL HEMATOPOIESIS
- PRELEUKEMIC
- POST-REMISSION
- SOMATIC MUTATION
- MOSAIC LOSS
- MOSAICISM AND (IDH1 OR PAC OR TP53 OR RB1 OR MLH1 OR MEN1 OR HIF2A, NF1 OR PRKAR1A)
- AGE\ -RELATED MOSAIC CHROMOSOMAL ALTERATIONS
- SENOLYSIS NOT (ATHEROSCLEROSIS OR CARDIOVASCULAR OR DIABETES OR NEURODEGENERATIVE)
- MGUS OR (MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE)
- PETO'S PARADOX
- (ICUS AND CYTOPENIA) OR (IDIOPATHIC CYTOPENIA OF UNDETERMINED SIGNIFICANCE)
- (CCUS AND (CYTOPENIA OR CLONAL)) OR (CLONAL CYTOPENIA OF UNDETERMINED SIGNIFICANCE)
- VEXAS NOT VISTA
- SOMATIC EVOLUTION.

Ces 16 mots-requêtes ont permis d'identifier 9 200 publications dont 2 882 n'étaient pas dans le corpus et ont été ajoutées.

- **Extension à la biologie fondamentale**

Le corpus a été ensuite enrichi avec les articles de biologie fondamentale à nouveau suite à un retour des experts. Les publications de 4 grandes familles d'articles de la biologie fondamentale ont été ajoutées :

1. Les articles de la famille 3R en lien avec le cancer par l'ajout de 6 mots requêtes (DNA INSTABILITY, DNA DAMAGE, DNA REPAIR, GENOME INSTABILITY, GENOME SURVEILLANCE et (CELL CYCLE) AND ((GATE KEEPER) OR (GATE KEEPERS))).  
Ces six requêtes ont ramené 103 280 publications dont 52 002, qui n'étaient pas dans le corpus V1, ont été ajoutées.
2. Les articles de biologie cellulaire fondamentale, dont les travaux importants sur les supresseurs de tumeurs et les oncogènes, ont été ajoutés en utilisant une requête MeSH utilisant les deux termes MeSH: « GENES, TUMOR SUPPRESSOR », « CARCINOGENS ».  
Ces deux termes ont permis d'identifier respectivement 28 022 et 17 059 publications dont 3 262 et 4 144 publications qui n'étaient pas dans le corpus V1 et qui y ont été ajoutées.
3. Les articles en immunologie relatifs aux applications récentes des aspects fondamentaux de cette discipline dans le domaine du cancer, ont été ajoutés par une requête MeSH utilisant les termes « ANTINEOPLASTIC AGENTS » restreint au domaine IMMUNOLOGY, « IMMUNOTHERAPY, ADOPTIVE », et « IMMUNE CHECKPOINT ».  
1 546 publications (dont 224 qui n'étaient pas dans le corpus) ont été ajoutées par la requête « ANTINEOPLASTIC AGENTS », 7233 (dont 1 190 qui n'y étaient pas) ont été ajoutées par la requête « IMMUNOTHERAPY, ADOPTIVE » et 2189 (dont 722 qui n'y étaient pas) ont été ajoutées par la requête « IMMUNE CHECKPOINT ».
4. Les publications d'épigénétique sur le cancer ont été prises en compte par les requêtes relatives aux supresseurs de tumeur et oncogènes.

- **Des difficultés à définir un périmètre étendu dans le domaine de la santé publique**

Concernant les articles de prévention, méthodologie / épidémiologie, la définition du périmètre à intégrer pose à la fois des questions méthodologique (par exemple, comment identifier les publications liées au cancer parmi l'ensemble des publications liées à l'obésité ?) et politique (quels aspects de ces domaines l'INCa souhaite intégrer dans le corpus cancer ?) [9].

Les échanges avec les experts et la Direction du pôle recherche de l'INCa n'ont pas permis d'aboutir à un consensus. Nous n'avons pas ajouté de publications en prévention, méthodologie, épidémiologie dans un premier temps pour la construction du corpus cancer.

Le périmètre des publications en prévention, épidémiologie et SHS liées au cancer sera précisé lors des travaux spécifiques sur les corpus Santé Publique. Ces résultats seront publiés en fin d'année.

#### **Étape 4. Troisième ajustement : exploration thématique (topic modeling) pour analyser et réduire le bruit éventuel**

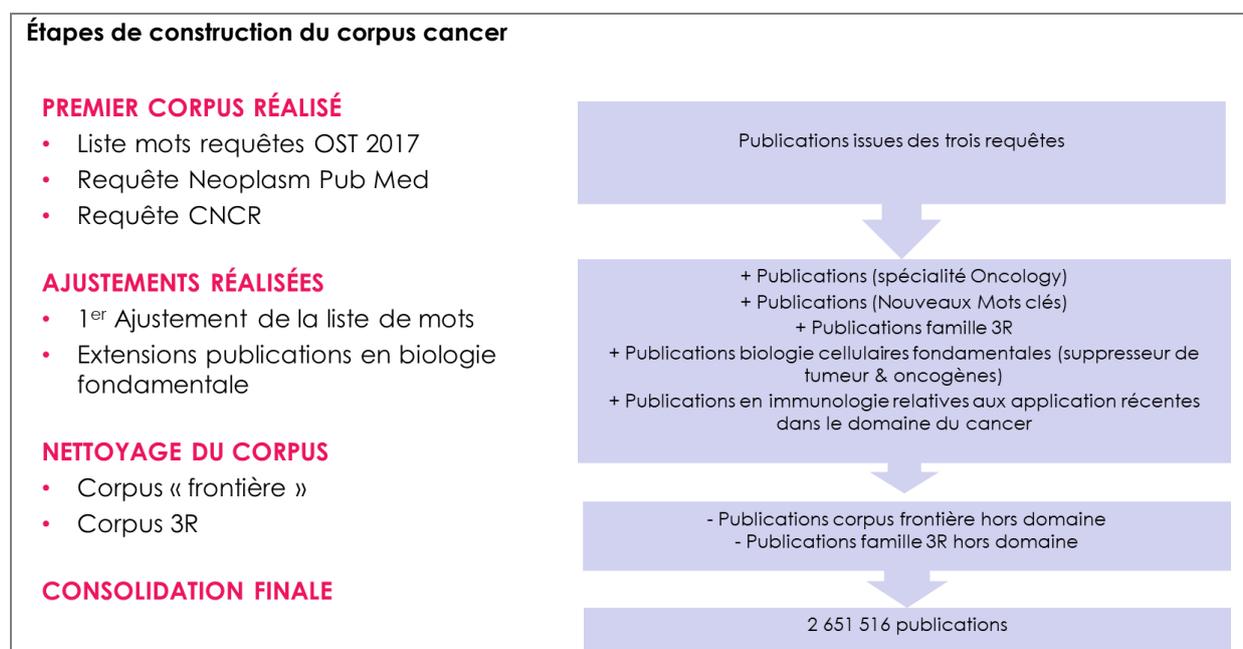
Le Topic Modeling (TM) est une méthode d'analyse qui permet d'identifier les thématiques sous-jacentes au sein d'un corpus de documents. Une thématique correspond à un vocabulaire spécifique constitué d'un ensemble de termes du corpus ayant des co-occurrences fréquentes. Le nombre de thématiques retenues dépend de la taille du corpus et du grain d'analyse que l'on souhaite effectuer. Dans le cas de cette étude, cette méthode a servi à identifier d'éventuelles thématiques hors sujet (ne concernant pas ou très peu le cancer). La maille d'analyse retenue a donc été relativement fine. Chaque thématique (ou topic) peut être interprétée à partir de ses termes caractéristiques (termes assez fréquents et assez spécifiques de la thématique), ou si cela ne suffit pas, à partir des publications les plus significatives.

L'OST a exploré plus finement par TM deux sous-corpus qui sont susceptibles de ramener des publications hors domaine du cancer :

- les documents « frontière » du corpus V1 ne figurant pas dans le corpus PubMed, le corpus CNCR, la "subject category" DM (ONCOLOGY) ;
- les documents de la famille 3R ne figurant pas dans le corpus V1.

Ces deux sous-ensembles de publications ont fait l'objet d'une modélisation thématique avec 150 topics sur le premier ensemble et 30 topics pour le second. Sur les 150 topics du premier ensemble, les experts ont identifié 62 topics hors domaine cancer. Sur les 30 topics du second ensemble, 16 topic hors domaine ont été identifiés. Grâce au retour des experts, l'OST a construit des anti-requêtes à partir de termes spécifiques des topics hors domaine. Ces anti-requêtes ont été appliquées au corpus frontière et au corpus 3R, afin de supprimer les publications relatives aux topics hors domaine. Cette étape a amené à supprimer du corpus cancer 172 000 publications du corpus frontière et 42 269 publications du corpus famille 3R.

Le corpus final de publication scientifiques relatives au cancer et utilisée pour le calcul des indicateurs regroupe 2 651 516 publications.



## Annexe 3 : Requêtes utilisées pour construire les sous corpus par type de Recherche

### Recherche fondamentale

#### **Approche disciplinaire**

Approximation avec le domaine disciplinaire « Biologie fondamentale »

➔ **571 848 publications (environ 22 % du corpus cancer)**

### Recherche clinique

#### **Approche par mots clés + descripteur MESH**

- Descripteurs MeSH : "Clinical Trial" [Publication Type] et terme MeSH « Clinical Trials as Topic »
- Mots requêtes
  - PATIENT\* COHORT,
  - CLINICAL TRIAL\*,
  - CONTROL TRIAL, CONTROLLED TRIAL,
  - RANDOMIZED TRIAL\*, RANDOMISED TRIAL\*, RANDOMLY
  - "PHASE III" , "PHASE-III" "PHASE3", "PHASE 3", "PHASE-3"
  - "PHASE II" ; "PHASE-II", "PHASE2", "PHASE 2", "PHASE-2" ,
  - "PHASE I", "PHASE-I", "PHASE1", "PHASE 1", "PHASE-1"
  - CASE REPORT, REPORT OF A CASE
  - CLINICAL ENTITY, CLINICAL CASE(S)
  - SYSTEMATIC REVIEW(S), UMBRELLA REVIEW, COMPARATIVE REVIEW
  - CANCER SURGERY, TUMOR SURGERY, SURGICAL TREATMENT, SURGICAL PROCEDURE, SURGICAL INTERVENTION
  - THERAPEUTIC DRUG MONITORING, DRUG THERAPY
  - PLACEBO, DOUBLE BLIND
  - PROGNOSIS

➔ **476 403, publications (environ 18 % du corpus cancer)**

### Recherche translationnelle

Pas de définition qui fasse consensus

- Test d'une requête textuelle pré-clinique
  - PRECLINICAL,
  - BIOBANK,
  - ANCILLARY,
  - TUMOR SAMPLE, TUMOR SAMPLES,
  - PLASMA SAMPLE, PLASMA SAMPLES,
  - BLOOD SAMPLE, BLOOD SAMPLES,
  - TISSUES SAMPLE, TISSUES SAMPLES,
  - PRIMARY TUMOR,
  - IN VIVO TRIAL, IN VIVO TRIALS, IN VIVO CLINICAL TRIAL, IN VIVO CLINICAL TRIALS.

➔ **113 144 publications (environ 4 % du corpus cancer)**

## Références Bibliographiques

- [1] La position scientifique de la France dans le monde et en Europe, 2005-2018, Observatoire des Sciences et Techniques, Hcéres.  
[https://www.hceres.fr/sites/default/files/media/downloads/hceres\\_ost\\_positionnement\\_scientifique\\_france\\_edition\\_2021\\_1.pdf](https://www.hceres.fr/sites/default/files/media/downloads/hceres_ost_positionnement_scientifique_france_edition_2021_1.pdf)
- [2] La position scientifique de la France dans le monde et en France, 2000-2015, Observatoire des Sciences et Techniques, Hcéres.  
[https://www.hceres.fr/sites/default/files/media/downloads/hceres\\_ost\\_position\\_scientifique\\_france\\_mars\\_2018\\_0.pdf](https://www.hceres.fr/sites/default/files/media/downloads/hceres_ost_position_scientifique_france_mars_2018_0.pdf)
- [3] CNCR – FHF , Quelle est la place de la France en recherche en cancérologie ? Analyse de la production scientifique 2010-2019 France & Monde –  
<https://cncr.fr/qui-sommes-nous/dossier-recherche-cancerologie-france>
- [4] Supporting France Universities's reflections on the state of biomedical research in France, SIRIS  
<https://franceuniversites.fr/wp-content/uploads/2023/04/Report-State-of-Biomedical-Research-in-FR-France-Universites.pdf>
- [5] •Methods for identifying biomedical translation: a systematic review  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9091120/>
- [6] Web of Science Core Collection –  
<https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection>
- [7] PubMed <https://pubmed.ncbi.nlm.nih.gov/>
- [8 ] MesH <https://meshb.nlm.nih.gov/>
- [9] European Groundshot—addressing Europe's cancer research challenges: a Lancet Oncology Commission - <https://www.thelancet.com/commissions/European-Groundshot-cancer-research>





2 rue Albert Einstein  
75013 Paris, France  
T. 33 (0)1 55 55 60 10  
hceres.fr

Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur