

Les cohortes pour les études et la recherche en santé

Tome 2 : Annexes

Mireille ELBAUM
Dr Aquilino MORELLE

Membres de l'Inspection générale des
affaires sociales



INSPECTION GÉNÉRALE
DES AFFAIRES SOCIALES

N°2019-045R

Bertrand MINAULT
Henri RIBIERAS

Membre de l'Inspection générale de
l'Éducation, du Sport et de la Recherche



N°2020-025

Février 2020

SOMMAIRE

ANNEXE 1 :	COHORTES, REGISTRES, ENQUETES, PANELS : DES DEFINITIONS AUX UTILISATIONS.....	5
1	LES COHORTES : UN INSTRUMENT POUR LA RECHERCHE ET L'EXPERTISE EN EPIDEMIOLOGIE ET EN SANTE PUBLIQUE, DONT LES OBJECTIFS ET LES CARACTERISTIQUES SONT DE FAIT TRES DIVERS.....	5
1.1	LES PRINCIPES DES COHORTES EPIDEMIOLOGIQUES.....	5
1.2	LES PROBLEMES ET LES CARACTERISTIQUES METHODOLOGIQUES DES COHORTES	6
1.3	UNE DIVERSITE D'OBJECTIFS ET D'UTILISATIONS QUI ESTOMPENT CERTAINES SPECIFICITES DE L'OUTIL « COHORTE ».....	7
2	À COTE DES « COHORTES », D'AUTRES INSTRUMENTS D'OBSERVATION ET DE RECHERCHE A DIMENSION LONGITUDINALE, AVEC LESQUELS LA DISTINCTION ET L'ARTICULATION NE SONT PAS TOUJOURS AISEES	8
2.1	LES BASES DE DONNEES D'ORIGINE ADMINISTRATIVE	8
2.2	LES REGISTRES DE MORBIDITE OU DE PATIENTS.....	9
2.3	LES ENQUETES ET PANELS STATISTIQUES.....	10
ANNEXE 2 :	LE FINANCEMENT DES COHORTES PAR LE PROGRAMME DES INVESTISSEMENTS D'AVENIR	13
1	PRESENTATION GENERALE DU DISPOSITIF	13
1.1	LE CADRE FINANCIER GLOBAL.....	13
1.2	LES APPELS A PROJETS	14
1.3	UNE ANALYSE FINANCIERE DIFFICILE A CONDUIRE.....	15
2	FINANCEMENTS ATTRIBUES	16
2.1	PERIODE 2011 - 2018.....	16
2.2	PERIODE 2020-2025 (SUITE A L'EVALUATION DE SEPTEMBRE 2019)	20
ANNEXE 3 :	PANORAMA GENERAL DES COHORTES SOUTENUES DANS LE CADRE DU PIA.....	23
ANNEXE 4 :	PRESENTATION DES 14 COHORTES SOUTENUES DANS LE CADRE DU PIA.....	27
LES COHORTES EN POPULATION		
	LA COHORTE CONSTANCES	29
	LA COHORTE E3N-E4N.....	38
	LA PLATEFORME RECONAI : LES COHORTES ELFE ET EPIPAGE 2	47
LES COHORTES DE PATIENTS		
	LA COHORTE CANTO	57
	LA COHORTE CKD REIN	62
	LA COHORTE COBLANCE	72
	LA COHORTE CRYOSTEM.....	78
	LA COHORTE HEPATHER	84
	LE PROJET HOPE-EPI.....	91
	LA COHORTE MEMENTO	98
	LA COHORTE OFSEP.....	105
	LA COHORTE PSY-COH	110
	LA PLATEFORME DE COHORTES RADICO.....	117
	ANNEXE LES 13 COHORTES RADICO	124

ANNEXE 5 :	QUELQUES COHORTES OU DISPOSITIFS DE SUIVIS LONGITUDINAUX SOUTENUS PAR D'AUTRES CANAUX QUE LE PIA.....	129
1	LES ENQUETES OU BASES D'ETUDES EN POPULATION GENERALE.....	129
1.1	LES OPERATIONS DU SYSTEME STATISTIQUE : L'ENQUETE EHIS ET LE PROJET D'EDP SANTE.....	129
1.1.1	<i>L'enquête EHIS.....</i>	129
1.1.2	<i>Le projet d'EDP santé.....</i>	130
1.2	UNE COHORTE CONSACREE A LA NUTRITION: L'ETUDE NUTRINET	131
2	LES ETUDES SUR LES LIENS ENTRE SANTE ET TRAVAIL	131
2.1	LES ENQUETES STATISTIQUES SUR LES CONDITIONS DE TRAVAIL	132
2.2	LES COHORTES COSET SUR LES ACTIFS INDEPENDANTS ET AGRICOLES : UN COMPLEMENT DIRECT A LA COHORTE CONSTANCES 132	
2.3	D'AUTRES DISPOSITIFS SPECIFIQUES DANS LE DOMAINE DE LA SURVEILLANCE : LES ETUDES ESPRI ET EPINANO.....	133
2.4	LA COHORTE AGRICAN SUR LES RISQUES DE CANCERS LIES AUX ACTIVITES AGRICOLES	133
3	LES ETUDES SUR LE VIEILLISSEMENT ET LA DEPENDANCE	133
3.1	L'ENQUETE SHARE	134
3.2	LE DISPOSITIF STATISTIQUE PUBLIC DE SUIVI DU VIEILLISSEMENT ET DE LA DEPENDANCE	134
4	LES DISPOSITIFS DE SUIVI LONGITUDINAL DES PATIENTS	135
4.1	DES COHORTES DE PATIENTS VIVANT AVEC LE VIH OU UNE HEPATITE VIRALE SOUS L'EGIDE DIRECTE DE L'ANRS.....	135
4.1.1	<i>Des cohortes d'adultes vivant avec le VIH.....</i>	136
4.1.2	<i>Des cohortes pédiatriques.....</i>	136
4.1.3	<i>Une cohorte de patients co-infectés par le VIH et le VHC.....</i>	136
4.2	DES COHORTES OU PROJETS DE PLATEFORMES SUR LES MALADIES CHRONIQUES.....	136
4.2.1	<i>L'étude ENTRED sur les patients diabétiques réalisée par Santé publique France.....</i>	136
4.2.2	<i>Le projet d'infrastructure de suivi de e-cohortes de patients atteints d'une maladie chronique porté par l'AP-HP.....</i>	137
4.3	DES DISPOSITIFS DE SUIVI DES PATIENTS TRAITES POUR UN CANCER	137
4.3.1	<i>La plateforme de données longitudinales en cours de développement à l'Inca</i>	138
4.3.2	<i>Le programme ESMÉ développé par Unicancer pour l'analyse des données de prise en charge dans les CLCC</i>	138
4.4	DES COHORTES DE CONSOMMATEURS DE SOINS DEVELOPPEES PAR LA CNAM A PARTIR DU SNDS, EN PARTENARIAT EVENTUEL AVEC L'ANSM ET LA HAS	139
ANNEXE 6 :	LES POTENTIALITES LIEES AU SYSTEME NATIONAL DES DONNEES DE SANTE (SNDS) ET AUX PLATEFORMES DE DONNEES DE SANTE EN CONSTITUTION	141
1	LE CADRE ACTUEL OFFERT PAR LE SNDS : DES POSSIBILITES D'EXPLOITATION ET D'APPARIEMENTS, DES CONTRAINTES EN TERMES DE STOCKAGE SECURISE DES DONNEES.....	141
1.1	PRESENTATION DU SYSTEME NATIONAL DES DONNEES DE SANTE (SNDS)	141
1.1.1	<i>Le développement du SNDS.....</i>	141
1.1.2	<i>Le contenu du SNDS.....</i>	142
1.2	L'ACCES DES COHORTES AU SNDS ET LE STOCKAGE SECURISE DES DONNEES APPARIEES.....	144
1.2.1	<i>Les procédures d'accès au SNDS.....</i>	144
1.2.2	<i>Les implications et les pratiques en termes de stockage sécurisé des données.....</i>	145
2	LES PERSPECTIVES ET QUESTIONS LIEES AUX AUTRES PLATEFORMES EN COURS DE CONSTITUTION.....	147
2.1	LA PLATEFORME DES DONNEES DE SANTE, DITE HEALTH DATA HUB	147
2.1.1	<i>Le nouveau cadre mis en place en 2019</i>	147
2.1.2	<i>L'avancement du projet.....</i>	148
2.2	LA PLATEFORME DEVELOPEE PAR L'INCA.....	149
2.3	LA PLATEFORME DU PROJET FRANCE COHORTES PORTE PAR L'INSERM	150

ANNEXE 7 :	LES GRANDES COHORTES ÉPIDÉMIOLOGIQUES EN POPULATION GÉNÉRALE : HISTORIQUE ET PANAROMA INTERNATIONAL	153
1	UNE PREDOMINANCE ANGLO-SAXONNE ANCIENNE ET SOLIDEMENT ETABLIE.....	153
1.1	LES COHORTES EPIDEMIOLOGIQUES FONDATRICES	153
1.2	LE DEVELOPPEMENT DES COHORTES DANS LES ANNEES 1970.....	154
2	UNE DIFFUSION INTERNATIONALE DESORMAIS CONSIDERABLE D’UN INSTRUMENT A LA PUISSANCE RENOUVELEE ET AUX OBJECTIFS EVOLUTIFS	155
2.1	L’EMERGENCE INTERNATIONALE DES « MEGA-COHORTES »	155
2.2	LE RATTRAPAGE EUROPEEN	157
ANNEXE 8 :	L’EXPÉRIENCE DES GRANDES COHORTES ÉPIDÉMIOLOGIQUES AU ROYAUME-UNI	159
1	UNE LARGE ET RICHE PALETTE DE « GRANDES COHORTES » EPIDEMIOLOGIQUES	159
1.1	ÉLÉMENTS DE PRESENTATION DES COHORTES EPIDEMIOLOGIQUES LES PLUS IMPORTANTES	159
	<i>Présentation des 34 cohortes les plus importantes.....</i>	<i>159</i>
	<i>Classification thématique des 34 principales cohortes</i>	<i>163</i>
1.2	UN FOCUS SUR TROIS COHORTES PARTICULIEREMENT INTERESSANTES.....	165
1.2.1	<i>The Million Women Study</i>	<i>165</i>
1.2.2	<i>UK BioBank.....</i>	<i>165</i>
1.2.3	<i>The Determinants of Adolescent Social well-being and Health (DASH).....</i>	<i>165</i>
2	LES COHORTES BRITANNIQUES : DES CARACTERISTIQUES LIEES A UN CONTEXTE FAVORABLE MAIS DES ENJEUX ET DES STRATEGIES EN DEBAT	166
2.1	DES CARACTERISTIQUES UN CONTEXTE FAVORABLE	166
2.1.1	<i>La coopération entre trois grands financeurs.....</i>	<i>166</i>
2.1.2	<i>Un environnement propice permettant de tirer parti de bases médico-administratives très riche</i>	<i>166</i>
2.1.3	<i>Un accent fort mis sur les liens entre caractéristiques socio-économiques et données de santé</i>	<i>167</i>
2.2	DES ENJEUX ET DES STRATEGIES EN DEBAT.....	168
3	QUELQUES ENSEIGNEMENTS OU SUJETS DE REFLEXION POUR LA FRANCE	168

ANNEXE 1 : Cohortes, registres, enquêtes, panels : des définitions aux utilisations

1 Les cohortes : un instrument pour la recherche et l'expertise en épidémiologie et en santé publique, dont les objectifs et les caractéristiques sont de fait très divers

Les cohortes, sont, avec les études de cas-témoins¹, l'un des instruments reconnus et revendiqués par l'épidémiologie, discipline dont l'objet, est, selon l'Organisation mondiale de la santé (OMS,) « l'étude de la distribution des maladies et des invalidités dans les populations humaines, ainsi que des influences qui déterminent cette distribution ».

1.1 Les principes des cohortes épidémiologiques

Le principe des cohortes est de « sélectionner un ensemble de sujets partageant un certain nombre de caractéristiques communes, et de les suivre dans le temps à l'échelle individuelle afin d'identifier la survenue d'événements de santé d'intérêt. Une étude de cohorte repose sur le respect d'un protocole pré-établi, concernant d'une part la collecte d'informations et d'autre part, les modalités du suivi des sujets permettant d'identifier sans biais, la survenue des événements de santé². » Le portail Épidémiologie France recensait en juillet 2019 251 études de cohorte de source française répondant à cette définition.

Il s'agit ainsi d'instruments qui visent un suivi continu, le plus souvent prospectif, des mêmes individus sur une période de temps significative, avec comme objectifs d'une part « de décrire les circonstances, la survenue et l'évolution des maladies, », d'autre part, grâce à « la quantification des risques et à l'analyse de la cohérence des résultats observés », « d'argumenter l'éventuelle causalité des liens mesurés »³ entre des déterminants et une problématique de santé, et enfin, dans certains cas, d'apprécier l'efficacité dans la durée d'interventions préventives ou réparatrices.

L'ambition première des cohortes est donc centrée sur la recherche d'associations, et si possible de relations de causalité, entre différents facteurs de risque et des événements de santé, en identifiant les enchaînements et les interactions entre l'état de santé des personnes incluses et les facteurs qui ont pu y contribuer. Elles privilégient ainsi par principe, dans les recueils d'informations auxquelles elles procèdent, la recherche d'une « validité interne »⁴ fondée sur la description précise de l'état de santé des personnes incluses (« phénotypage ») et la catégorisation fine des types de pathologies et facteurs de risque, plutôt que la représentativité statistique des populations d'intérêt (« validité externe »). Cet objectif implique le plus souvent de collecter, généralement sur la base du volontariat, une diversité de données détaillées à différents moments du temps (examens cliniques, réponses à

¹ Études d'observation dans lesquelles les caractéristiques et expositions de sujets atteints d'une pathologie définie (les cas) sont comparées à celles de sujets indemnes de cette pathologie (les témoins)

² Définition retenue par le Portail Épidémiologie France développé sous l'égide de l'Alliance nationale pour les sciences de la vie et de santé (Aviesan).

³ Source : Institut de recherche en santé publique (Iresp).

⁴ C'est-à-dire la fiabilité des résultats sur la population étudiée.

des questionnaires ou auto-questionnaires, prélèvements biologiques de diverse nature, et parfois imagerie...) en prêtant une forte attention aux conditions et à la qualité de ces recueils.

1.2 Les problèmes et les caractéristiques méthodologiques des cohortes

Les problèmes de méthode auxquels sont traditionnellement confrontées les cohortes épidémiologiques sont de plusieurs types⁵ :

- la mise en place de tailles d'échantillons suffisants pour permettre l'estimation avec la précision souhaitée des relations entre les événements de santé identifiés et des facteurs de risque spécifiques (« puissance statistique » de la cohorte) ;
- les biais de sélection liés aux modes de recrutement des participants (par exemple à partir d'un contact avec le système de soins) et à leur volontariat, qui, même si l'étude ne vise pas une représentativité statistique de la population cible, doivent éviter d'affecter l'estimation des relations mises en évidence ; cela peut rendre souhaitable de procéder à des redressements pour estimer les variables d'intérêt, redressements qui ne sont pas toujours possibles faute de données plus générales et ne peuvent souvent tenir compte des biais dus « aux caractéristiques inobservées » des participants (par exemple, le fait qu'accepter d'entrer dans un protocole de suivi peut traduire, à caractéristiques sociodémographiques identiques, des attitudes particulières vis-à-vis de la santé susceptibles d'influer sur les résultats de l'étude) ;
- l'attrition de l'échantillon suivi au cours du temps (participants devenus « non-répondants » ou « perdus de vue ») qui, outre le fait qu'elle accroît les biais de sélection, est susceptible de nuire à l'objectif recherché d'un suivi cohérent des individus pendant la période de temps jugée pertinente ;
- l'arbitrage entre qualité, variété et champ des informations recueillies, qui doit à la fois permettre de répondre à des problématiques de recherche précisément définies *ex ante*, mais aussi parfois, pour certaines cohortes, ouvrir la possibilité de traiter des questions évolutives ou (re)formulées ultérieurement⁶.

Eu égard à ces questions de méthode, les cohortes peuvent avoir des caractéristiques différentes qui rétroagissent sur leurs utilisations potentielles. Une distinction de base existe entre les cohortes en population, centrées sur les déterminants (biologiques, génétiques, sociaux, professionnels environnementaux...) des événements et/ou problèmes de santé et les cohortes de patients diagnostiqués ou traités pour une maladie, dont on suit ensuite l'évolution de la pathologie et l'efficacité des prises en charge et des traitements. Au-delà de cette distinction principale, une analyse réalisée en 2010 par l'Institut national de veille sanitaire (INVS) identifiait plusieurs types de caractéristiques méthodologiques, susceptibles de se combiner différemment selon les cohortes⁷ :

- les cohortes spécialisées, centrées sur une problématique spécifique (par exemple les effets du tabagisme), les cohortes multi-objectifs et les cohortes généralistes, qui couvrent un large éventail de problèmes de santé, voire de disciplines ;

⁵ Marcel Goldberg et Marie Zins, « Les études de cohortes : principes et méthode », *ADSP*, n°78, mars 2012

⁶ Agence nationale de recherche sur le sida et les hépatites virales

⁷ Georges Salines et Catherine de Launay, « Les cohortes : intérêt, rôle et position de l'INVS », *Document de travail*, 2010.

- les cohortes prospectives qui concentrent la période de suivi après l'inclusion, et les cohortes rétrospectives ou historico-prospectives, qui intègrent dans le suivi des événements sanitaires et/ou facteurs d'exposition observés par le passé ;
- les cohortes fermées, qui suivent les mêmes sujets tout au long de l'étude, et les cohortes ouvertes, qui incluent de nouveaux participants en cours de suivi, soit pour limiter les problèmes liés à l'attrition, soit pour élargir la représentativité de l'étude ;
- les cohortes donnant lieu à des recueils de données actifs (questionnaires, prélèvements biologiques..) ou privilégiant un suivi « passif » à partir des informations figurant dans des bases de données externes.

Les cohortes se distinguent enfin par la durée du suivi prévue, qui peut aller de quelques années pour une pathologie déterminée à une observation sur longue période, non limitée dans le temps.

1.3 Une diversité d'objectifs et d'utilisations qui estompent certaines spécificités de l'outil « cohorte »

Les objectifs premiers des cohortes, centrés sur la mesure d'associations causales entre facteurs de risque et maladies à partir de questions de recherche identifiées, se sont ainsi diversifiés et ont perdu de leur spécificité, suite à plusieurs phénomènes.

- Les demandes des pouvoirs publics de pouvoir disposer d'instruments susceptibles de contribuer à la surveillance épidémiologique et d'éclairer les politiques de santé publique ont conduit certaines cohortes, notamment en population, à rechercher davantage la représentativité statistique, que ce soit par des réouvertures périodiques (cohortes COSET de Santé publique France) ou par la recherche de sources « témoins » en population générale, notamment à travers les bases de données médico-administratives.
- Certaines cohortes, notamment de patients, prennent appui sur des participants recrutés et suivis à l'occasion de la fréquentation de centres de diagnostic ou de soins, et ont pour objectif connexe, non seulement de collecter, mais aussi d'harmoniser et d'améliorer les pratiques des professionnels travaillant dans ces centres (cas de cohortes OFSEP et CRYOSTEM).
- La mobilisation de moyens importants pour suivre sur le long terme l'état de santé des personnes incluses peut conduire à une évolution des questions de recherche au fur et à mesure de l'avancée en âge des participants, amenant à l'utilisation de certaines cohortes pour des problématiques autres que celles pour lesquelles elles ont été initialement construites (cas de la cohorte E3N).
- Le développement des maladies chroniques, et, avec le vieillissement de la population, des poly-pathologies chez les sujets âgés, implique des approches multi-factorielles, qui, d'une part peuvent rendre plus difficile l'établissement de causalités, et d'autre part induisent des modélisations sophistiquées dont le côté « boîte noire » est souvent critiqué⁸ ; ces problèmes de « discrimination » entre facteurs de risque sont accrus lorsque ceux-ci sont en interaction

⁸ Bernard-Simon Leclerc, « L'épidémiologie contemporaine en crise de paradigme », *Ruptures, revue transdisciplinaire en santé*, vol 10, n°2, 2005.

complexe les uns avec les autres, et que leurs effets sont liés aux contextes économiques et sociaux qui caractérisent l'environnement des personnes^{9 10}.

- Certaines cohortes affichent enfin l'objectif de contribuer au développement d'une médecine « prédictive », c'est-à-dire d'identifier une série de « marqueurs » biologiques ou génétiques permettant de repérer des individus « à risque », principalement en vue de politiques de prévention ou de traitement précoce, mais avec aussi de possibles implications socio-économiques non encore complètement maîtrisées aujourd'hui.

Cette diversité d'objectifs et d'utilisations potentielles met en évidence des incertitudes et un débat autour de la « nature scientifique » des cohortes, qui peuvent être considérées soit comme un instrument associé à des projets de recherche spécifiques, soit comme des « infrastructures » à vocation plus générale, destinées à des communautés élargies de chercheurs et d'experts, et ayant vocation répondre à des problématiques pas toujours identifiées au départ.

2 À côté des « cohortes », d'autres instruments d'observation et de recherche à dimension longitudinale, avec lesquels la distinction et l'articulation ne sont pas toujours aisées

Le portail Épidémiologie-France propose un catalogue des « principales bases de données individuelles en santé de source française qui peuvent être utiles au développement de la recherche et de l'expertise en santé publique » (hors essais cliniques). Il distingue, à côté des cohortes, d'autres études longitudinales « d'observation » permettant de décrire les caractéristiques et expositions d'un ensemble de sujets à l'occasion de contacts successifs avec le système de santé (177 occurrences recensées), ainsi que diverses sources d'information comme les études de cas-témoins, les enquêtes, les registres ou les bases de données administratives.

L'instrument « cohortes » s'inscrit à cet égard dans un paysage complexe et multiple d'instruments d'observation des problèmes de santé à l'échelle individuelle, potentiellement utilisables pour la recherche et les études en matière de santé. Ces divers instruments ont en principe des propriétés méthodologiques différentes, mais, développant de plus en plus souvent une dimension longitudinale, ils peuvent questionner l'intérêt spécifique de certaines cohortes, ainsi que les articulations à envisager entre les différentes sources.

Parmi ces sources de données potentiellement « concurrentes » ou « articulables » avec les cohortes, quelques-unes méritent un retour sur leurs définitions et leurs principes d'utilisation.

2.1 Les bases de données d'origine administrative

Ces bases de données, constituées principalement à des fins de gestion, peuvent avoir des origines diverses : principalement issues des systèmes d'information des hôpitaux et des organismes de protection sociale (CNAM, CNAV, CNSA), elles incluent aussi les données issues de l'état-civil et des recensements de la population (échantillon démographique permanent-EDP), les informations issues des sources fiscales et sociales (revenus, prestations...) ou des collectes d'informations

⁹ Marcel Goldberg, Maria Melchior, Annette Leclerc et France Lert, « Les déterminants sociaux de la santé : apports récents de l'épidémiologie sociale et des sciences sociales de la santé », *Sciences sociales et santé*, vo.20, n°4, décembre2002.

¹⁰ Arnaud Fontanet, « L'épidémiologie arriverait à bout de souffle », *Interview à UP Magazine*, janvier 2019.

spécifiques comme les accidents de la circulation, les certificats de santé du jeune enfant, les déclarations d'accidents du travail ou de maladies professionnelles,

Le principal apport en France est celui du système national des données de santé (SNDS) qui regroupe, suite à des opérations d'anonymisation et de chaînage, les données issues du système national inter-régimes de l'assurance maladie (SNIIRAM), du programme de médicalisation des systèmes d'information (PMSI) et de l'analyse des causes de décès, et ce pour toute la population française (cf. annexe 6). Ce système, que les pouvoirs publics veulent enrichir en intégrant dans une Plateforme des données de santé (*Health data hub*) un ensemble élargi d'informations (dossiers médicaux, résultats d'analyses biologiques et d'imagerie...), offre une richesse importante et évolutive pour les études et la recherche en santé¹¹, suscitant des questions sur son articulation, présente et future, avec les cohortes épidémiologiques. Il peut en effet :

- permettre la constitution de cohortes *ad hoc*, le cas échéant de grande taille et donnant lieu à un suivi « passif » à partir des bases de données existantes (cas de la plupart des études réalisées par la CNAM);
- servir de base d'échantillonnage ou de « témoin » à des études de cohorte plus approfondies ;
- dispenser les cohortes de certains recueils d'informations, qui seront au fil du temps rassemblées *via* les entrepôts de données en cours de constitution ;
- enrichir les informations qu'elles collectent, d'une part en étendant leurs champ et périodes d'observation, d'autre part en fournissant une meilleure appréciation de leur représentativité (grâce à des informations sur les « perdus de vue » et/ou les non participants de la sous-population d'intérêt).

Les recoupements et interactions potentiels sont donc multiples, évolutifs et propres à chaque étude ou problématique scientifique. Utiliser la pleine mesure de ces potentialités implique en tout état de cause qu'un effort soit réalisé pour permettre une identification pertinente et robuste des pathologies dans les bases médico-administratives et que la qualité des informations médicales recueillies y soit suivie et attestée.

Il importe aussi de prendre conscience des limites explicatives que peuvent avoir les utilisations de ce type de données massives, qui, à côté des « signaux faibles » qu'elles permettent de percevoir, ne permettent pas toujours de répondre, en l'absence de recueil d'informations spécifiquement pensé à cet effet, à des questions donnant lieu à des causalités ou à des interactions complexes¹².

2.2 Les registres de morbidité ou de patients

À côté des bases de données constituées à des fins administratives ou de gestion, les registres de morbidité recueillent également avec une visée exhaustive, des données individuelles intéressant un ou plusieurs événements de santé. Selon la définition adoptée par le Comité stratégique des registres en 2016, il s'agit d'un « un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de surveillance, de recherche et d'évaluation en santé publique, par une équipe ayant les compétences appropriées ».

¹¹ Marcel Goldberg, Mireille Coeuret-Pellicer, Céline Ribet et Marie Zins, « Cohortes épidémiologique et bases de données d'origine administrative », *Med Sci*, vol. 28, n°4, avril 2012.

¹² Rodolfo Saracci, « Epidemiology in Wonderland : Big Data and precision medicine », *European journal of Epidemiology*, avril 2018.

Ces registres populationnels se distinguent d'une autre catégorie de registres constitués à partir d'une patientèle, qui peuvent recenser les patients consultant, pour des troubles donnés, certains établissements hospitaliers ou centres de soins, sachant que, dans la littérature scientifique, le terme de registre peut désigner aussi bien l'un que l'autre¹³.

Il s'agit dans les deux cas de bases de données en vie réelle constituées en dehors d'un contexte expérimental formalisé. S'agissant des registres de morbidité, 61 étaient répertoriés en France en 2017, dont au moins 55 évalués par le Comité d'évaluation des registres (CER). Il s'agit pour environ la moitié de registres de cancers, dont douze sont spécialisés, c'est à dire qu'ils recueillent les cas présentant un type de cancer particulier, alors que les 17 registres généraux incluent tous les types de cancer dans une zone géographique donnée, et couvrent près de 24 % de la population française.

Ces registres, dont la couverture et la qualité sont extrêmement diverses, forment un ensemble hétérogène, qui pose le même type de problèmes de complémentarité et d'articulation avec les bases de données médico-administratives que les cohortes. Cela a conduit la Direction générale de la santé (DGS) à saisir en juin 2019 le Haut comité de la santé publique (HCSP) d'une demande de réflexion sur la politique nationale en matière de registres, à l'horizon de la fin 2020.

La distinction et l'articulation entre registres et cohortes suscite par ailleurs des questions spécifiques, dans la mesure où nombre de ces registres ont eux-mêmes une dimension longitudinale et des recouvrements divers avec les cohortes. Ainsi,

- certaines cohortes par pathologies sont directement ancrées, voire se confondent pour partie avec des registres, notamment ceux de patients diagnostiqués ou traités dans des centres spécialisés : c'est par exemple le cas pour CRYOSTEM qui recrute 82 % des patients ayant reçu une greffe de cellules souches hématopoïétiques, en adjoignant à leur suivi une plateforme de données biologiques, ou pour HOPE-EPI qui inclut dans son suivi l'ensemble des personnes recensées depuis 2000 par le registre national des cancers de l'enfant, avec une collecte de données enrichie (traitements détaillés, prélèvements...) pour une partie d'entre eux ;
- d'autres cohortes viennent compléter des registres à dimension longitudinale, comme le projet de cohorte dite « haute définition » de l'Observatoire Français de la Sclérose en Plaques (OFSEP) partant d'un recueil d'informations systématique sur cette maladie, ou, avec un autre angle, la cohorte CKD REIN qui éclaire l'évolution de l'insuffisance rénale chronique avant le stade terminal, tandis que le registre REIN suit l'ensemble des patients dialysés ou greffés parvenus à ce stade,.

Ces recouvrements rendent là encore parfois la distinction entre registres et cohortes peu aisée, hormis sur des critères de non exhaustivité ou de volontariat, et rendent nécessaire, pour chaque problématique d'étude, une articulation pensée à partir de l'ensemble des données disponibles.

2.3 Les enquêtes et panels statistiques

La troisième source de données à mettre en regard des cohortes épidémiologiques sont les enquêtes statistiques, qui peuvent être conduites sous l'égide du système statistique public (Insee, Drees, Dares), des agences sanitaires, ou d'organismes ou équipes de recherche.

¹³ Florence Francis-Oliviero, « les registres de morbidité en France : état des lieux, enjeux et perspective », *Médecine humaine et pathologie*, 2018.

Ces enquêtes se différencient des cohortes par une recherche systématique de représentativité statistique des populations d'intérêt : prêtant une attention première aux questions d'échantillonnage, de redressement et de non-réponse, elles ont pour principal objectif d'estimer des indicateurs statistiques en population, et de procéder à une description analytique de divers aspects de l'état de santé des individus en fonction de leurs caractéristiques et comportements, sans chercher à inférer de relations de causalité à partir des corrélations mises en évidence.

Ces enquêtes peuvent être soit ponctuelles (comme l'enquête Santé et itinéraires professionnels), soit répétées à un rythme régulier, comme les enquêtes européennes sur la santé (*European Health Interview Survey* -EHIS) ou les enquêtes Conditions de travail (*cf.* annexe 5), ou enfin les enquêtes sur la nutrition de Santé publique France et de l'Anses.

Ces enquêtes réalisées à partir d'un échantillon représentatif diffèrent donc dans leur principe des cohortes, mais, même lorsqu'elles ont une visée instantanée, une dimension longitudinale peut leur être donnée, et l'est de fait de plus en plus souvent, au travers de plusieurs voies :

- l'introduction dans les questionnaires de parties rétrospectives sur les parcours de vie et de santé des personnes ;
- l'appariement avec des bases de données médico-administratives, qui permet de les assortir d'un suivi à la fois rétrospectif et prospectif des situations socio-sanitaires et des consommations de soins : c'est par exemple le cas des enquêtes EHIS, dont l'appariement prévu avec le SNDS permettra une vision de 5 ans en rétrospectif et 5 ans en prospectif, et sera complété par un appariement avec les données socio-fiscales ; plus généralement, la Drees privilégie désormais, tant en matière de santé que de vieillissement, l'utilisation de données administratives organisées dans une perspective de suivi longitudinal, et complétées en tant que de besoin par des enquêtes statistiques (*cf.* annexe 5).
- la réalisation possible de ces enquêtes, totalement ou partiellement, sous forme de « panels », c'est-à-dire d'échantillons aléatoires suivis eux-mêmes de façon longitudinale et réalimentés de façon à conserver leur représentativité à chaque étape ; c'est le cas par exemple de l'enquête Share qui recueille tous les deux ans depuis 2004 des informations sur la santé et les conditions de vie d'un panel de personnes de plus de 50 ans ; un autre exemple en est l'enquête Conditions de travail de la Dares, qui, réalisée tous les trois ans auprès de 35 000 actifs, procède à trois interrogations successives des actifs en emploi et à au moins deux de ceux entrés dans la vie professionnelle ou ayant pris leur retraite pendant la période (*cf.* annexe 5).

Les apports respectifs et les complémentarités entre enquêtes et cohortes méritent donc, comme pour les autres instruments, d'être appréciés au cas par cas pour chaque thématique d'intérêt.

Il faut enfin noter qu'au-delà de leurs propriétés méthodologiques, des critères de coût entrent aussi en ligne de compte dans le choix de tel ou tel instrument : d'un côté, cohortes et enquêtes engendrent des coûts liés à la collecte d'informations « non routinisées » (contacts personnalisés avec les enquêtés, passation de questionnaires, vérification des réponses, relance des non-répondants ...), ces coûts étant notablement majorés lorsque, comme c'est le cas pour la plupart des cohortes, les recueils d'information s'étendent à des prélèvements biologiques ou des examens cliniques *ad hoc*. D'un autre côté, les bases de données administratives induisent des coûts de constitution et de maintenance qui sont liés à la fois au besoin de capacités massives de stockage et de calcul et aux travaux techniques nécessaires pour en documenter et en faciliter l'accès : pour l'essentiel assumés par les organismes publics détenteurs de ces bases, ces coûts peuvent parfois être refacturés aux utilisateurs qui souhaitent accéder à ces informations, surtout lorsque ces accès appellent des prestations de service spécifiques (appariements...).

ANNEXE 2 : Le financement des cohortes par le programme des investissements d'avenir

1 Présentation générale du dispositif

1.1 Le cadre financier global

Le programme d'investissement d'avenir (PIA 1) est créé en 2010 et doté de 35 milliards d'euros. S'y ajoutent en 2013 un deuxième programme (PIA2) doté de 12 Milliards d'euros et, en 2017, un troisième programme (PIA3) doté de 10 milliards d'euros.

L'Agence nationale de la recherche (ANR) gère pour le compte du Commissariat aux investissements (CGI), devenu en 2017 secrétariat général pour l'investissement (SGPI), sur la base d'appels à projets, les crédits destinés à financer les actions de formation, recherche et équipements. Ces crédits représentent 22,5 Mi€ sur le PIA1, 4,5 Mi€ sur le PIA2 et 2,85 Mi€ sur le PIA3.

Les cohortes entrant dans le champ de la mission ont été financées dans le cadre du PIA1.

Les modalités de financement des cohortes sont décrites dans la convention Etat – ANR du 20 juillet 2010 pour l'action « *Santé et biotechnologies* » qui porte notamment sur le financement des cohortes et des infrastructures en biologie-santé. Cette convention est modifiée par avenant en janvier 2011 et mai 2012, et, plus récemment, en avril 2017.

Par ailleurs la convention Etat – ANR du 16 juin 2010 a fixé les modalités de financement des « équipements d'excellence » (Equipex) qui correspondent à des équipements de recherche de valeur intermédiaire dont le financement ne peut être assuré ni dans le cadre des très grands équipements couverts par des feuilles de routes (Très grandes infrastructures de recherche – TGIR) ou des accords internationaux, ni par les organismes et établissements de recherche sur leur budget récurrent.

Le principe général, commun à l'ensemble du PIA, est celui du co-financement de chaque projet, de manière à générer un « effet de levier » des sommes investies.

Le montant initial de financement de l'action Santé et biotechnologies est le suivant :

- Cohortes : 200 M€ de dotation non consommable, dont seuls les intérêts sont utilisés pour financer les projets retenus, auxquels s'ajoutent 10 M€ de dotation consommable comme suite à l'avenant de 2012 ;
- Infrastructures : dotation de 1,35 Mi€, non consommable à hauteur de 800 M€.

Pour la période post-évaluation finale de 2019, l'avenant d'avril 2017 de la convention Etat – ANR prévoit un montant de financement reposant sur la réallocation d'une partie de la dotation non consommable de l'action : « *un maximum de 500 M€ de dotation non consommable de cette action sera concerné par cette décision de prorogation partielle. Cette prorogation couvre la période du 30 juin 2020 au 30 juin 2025* ».

Les crédits du PIA gérés par l'ANR sont soumis aux règles de gestion présentées dans le cadre d'un règlement financier adopté, pour le financement des cohortes, le 14 décembre 2012¹⁴ et modifié en 2016 et 2017. Il stipule notamment que l'aide est conditionnelle (nécessité de signer un accord de consortium entre partenaires du projet dans les 12 mois à compter de la signature de la convention attributive d'aide) et ne couvre pas tous les coûts du projet (frais de gestion initialement limités à 4 %). Un *reporting* financier annuel permet de calculer le montant des dépenses éligibles sur la base duquel est ensuite versé le montant annuel de la subvention.

1.2 Les appels à projets

Treize cohortes ont été retenues dans le cadre de trois appels à projets distincts.

1.2.1 Appel à projets Cohortes

Doté de 200 M€ de dotation non consommable et 10 M€ de dotation consommable, l'appel à projets Cohortes est lancé le 28 juin 2010 pour financer entre cinq et dix projets de cohortes, soit « *existantes, pour en assurer la pérennité et élargir leur champ d'investigation, soit destinées à couvrir des champs de recherche publique encore peu étudiés* ».

Date de lancement	Montant global de financement	Nombre de projets déposés	Nombre de projets retenus	Cohortes retenues
Juin 2010	200 M€ de dotation non consommable auxquels s'ajoutent 10 M€ de dotation consommable	44	10	CANTO CKD-REIN COBLANCE CRYOSTEM E4N HOPE-EPI I-SHARE ¹⁵ OFSEP PSY-COH RADICO

¹⁴ <https://anr.fr/fileadmin/documents/2012/ANR-REGLEMENT-COHORTES.pdf>

¹⁵ Projet arrêté le 31 janvier 2018.

1.2.2 Appel à projets Equipex deuxième vague

Le deuxième appel à projets (Equipex vague 2), lancé en 2011, vise à financer des « équipements de recherche de valeur intermédiaires » (inférieur à 20 M€). Doté de 210,3 M€, il a permis de retenir 36 projets sur un total de 270 projets soumis.

Date de lancement	Montant global de financement	Nombre de projets déposés	Nombre de projets retenus	Cohortes retenues
Juin 2011	210,3 M€	270	36	RE-CO-NAI HEPATHER

1.2.3 Appel à projets Infrastructures nationales en Biologie santé

L'appel à projets Infrastructures Nationales en Biologie et Santé (INBS) lancé en 2011 vise à financer, selon les termes de l'appel, « des ensembles d'équipement et de personnel qui soutiendront les projets de recherche de l'ensemble de la communauté française dans les divers domaines des sciences de la vie. Il doit doter la France de plusieurs grandes infrastructures d'envergure nationale et très compétitives internationalement ».

Date de lancement	Montant global de financement	Nombre de projets déposés	Nombre de projets retenus	Cohortes retenues
Juillet 2010	220 M€	37	9	CONSTANCES

CONSTANCES est financée sur les crédits de cette sous-action bien que la cohorte n'ait pas déposé de projet dans le cadre de l'appel. En revanche, le projet BIOBANQUE, doté de 17 M€, figurait parmi les lauréats de l'appel à projet, mais n'a pas été mené à terme.

La cohorte MEMENTO n'a pas été financée dans le cadre de ces appels à projet, mais principalement par la Fondation Alzheimer grâce aux crédits mobilisés au titre du Plan Alzheimer et le ministère chargé de l'enseignement supérieur et de la recherche dans le cadre des plans successifs consacrés à la maladie.

1.3 Une analyse financière difficile à conduire

1.3.1 Les sources utilisées

Les données présentées dans la présente annexe proviennent de deux sources :

- les données financières contenues dans le dossier d'évaluation finale des projets, qui indiquent notamment le montant des financements privés obtenus et la demande de l'aide pour la période post-évaluation ;
- les données issues du suivi effectué par l'ANR sur la base du reporting financier annuel transmis par les projets.

1.3.2 Des données très fragiles

Une première fragilité tient aux incohérences constatées entre les sources. Ainsi, par exemple, le dossier d'évaluation de CONSTANCES avance un montant de financements privés de 4,5 M€ quand le fichier de suivi de l'ANR ne mentionne pas de financement privé.

Une seconde fragilité tient à la faible robustesse des montants déclarés. Trois difficultés se conjuguent :

- les données financières de l'année 2019 sont traitées de manière différentes selon les dossiers : parfois ajoutées à la période de *reporting* stricto sensu (2011-2018), parfois intégrées à la demande de ressources pour 2020-2024, avec des risques possibles de double comptes ou d'omissions ;
- le périmètre des apports des partenaires n'est pas précisé et supposerait une comptabilité analytique (par exemple pour ventiler les charges indirectes) qui n'existe en général pas ;
- les modes de valorisation des apports varient d'un dossier à l'autre, allant d'une mesure de temps/homme à une référence à des salaires moyens ;
- le recensement des effectifs s'effectue parfois en emplois, parfois en équivalent temps-plein - ETP (mais jamais en équivalent temps plein travaillé ETPT). Certains emplois (attachés de recherche clinique) sont dans certains cas décomptés au titre des frais de personnels, mais dans d'autres cas financés en tant que prestations de service.

Les données financières présentent donc un manque flagrant de fiabilité et de comparabilité. Dans ce contexte, les tableaux qui suivent, qui combinent les données de ces différentes sources, doivent être appréciés davantage comme donnant des ordres de grandeur plutôt qu'une image fidèle de la réalité économique des cohortes.

2 Financements attribués

2.1 Période 2011 - 2018

2.1.1 Part de la subvention ANR par rapport à la demande initiale de financement

Ce tableau rapporte les financements attribués au titre du PIA par l'ANR sur la période considérée (après les éventuels réaménagements ou étalements qui ont pu concerner certaines cohortes) à leur demande de financement initiale au titre du PIA, telle qu'elle figure dans la réponse à l'appel à projets de 2010.

	CANTO	CKD-REIN	COBLANCE	CRYOSTEM	E4N	HOPE-EPI	OFSEP
Aide accordée (K€)	13 100	5 861	3 945	5 238	7 507	6 274	9 768
en % de la demande d'aide	62 %	52 %	73 %	53 %	78 %	40 %	54 %

	PSY-COH	RADICO	HEPATHER	RECONAI	CONSTANCES
Aide accordée (K€)	3 846	10 513	10 010	13 018	35 000
en % de la demande d'aide	54 %	53 %	35 %	87 %	60 %

Le niveau de financement au titre du PIA des projets retenus, rapporté à la demande de leurs porteurs, varie dans des proportions très importantes (de 35 à 87 %). La mission n'a pas eu connaissance d'une grille de critères, en dehors de l'avis du jury qui n'abordait pas directement les aspects financiers, justifiant ces variations.

2.1.2 Structure de l'ensemble des financements obtenus sur la période 2011 - 2018

Le tableau suivant resitue les financements obtenus au titre du PIA dans l'ensemble des ressources dont les cohortes ont pu disposer au cours de la période, grâce à des concours financiers publics ou privés.

en K€	CANTO	COBLANCE	CKD-REIN	CRYOSTEM	E4N ¹⁶	HOPE-EPI	OFSEP
PIA	13 100	3 945	5 861	5 238	7 507	6 274	9 768
Autres financements publics	552	694	217	482	8 300	5 690	671
Financements privés	1 466	307	2 864	0	380	194	3 646
Part des financements privés	10 %	6 %	32 %	0 %	2 %	2 %	26 %

¹⁶ Aux données financières communiquées par l'ANR ont été ajoutées les diverses contributions publiques ou privés recensées dans le dossier d'évaluation finale du projet

En K€	PSY-COH	RADICO	HEPATHER ¹⁷	RECONAI	CONSTANCES ¹⁸	<i>MEMENTO</i>
PIA	3 846	10 513	10 010	13 018	35 000	0
Autres financements publics	0	0	6 030	13 746	119 640	21 108
Financements privés	750	2 383	12 500	431	4 500	2 407
Part des financements privés	16 %	18 %	44 %	2 %	3 %	10 %

Les « autres financements publics » correspondent ici aux contributions financières de diverses natures versées par des structures publiques autres que l'ANR dans son rôle d'agence de financement au titre du PIA. Il peut s'agir de contributions au financement des dépenses d'investissement ou de fonctionnement (y compris dépenses de personnels) des cohortes elles-mêmes, versées notamment par :

- le ministère chargé de l'enseignement supérieur et de la recherche (CONSTANCES et RECONAI) ;
- la Cnam (CONSTANCES) ;
- des agences, comme l'agence de la biomédecine – ABM (CKDRein), l'Inca (HOPE-EPI et E4N), l'ANRS (HEPATHER) ;
- des établissements ou organismes de recherche (l'Inserm ; le GIP Genopole pour HOPE-EPI)
- par extension, à des fondations ou organismes faisant appel à la générosité publique (la Ligue nationale contre le cancer pour E4N et HOPE-EPI ; la Fondation Alzheimer pour MEMENTO).

Il peut par ailleurs s'agir de financements fléchés correspondant à des projets de recherche particuliers, résultant ou non d'appels à projet, versés dans ce cadre par les structures précitées ainsi que le ministère chargé de la santé ou l'ANR. Cette ligne intègre enfin les financements reçus au titre de projets de recherche financés dans le cadre d'appels à projet européens.

La part des financements privés reste relativement modeste pour la plupart des projets. Elle est cependant significative (supérieure à 25 %) pour les cohortes de patients CKDREIN, HEPATHER et OFSEP.

¹⁷ Aux données financières communiquées par l'ANR est ajoutée la contribution de l'ANRS

¹⁸ Les données relatives à Constances retracent l'ensemble des contributions financières mentionnées dans le dossier d'évaluation finale pour la période 2011 - 2019

2.1.3 Structure des dépenses sur la période 2011-2018

Le tableau suivant présente la structure des dépenses telle qu'elle ressort des relevés annuels de dépenses de l'ANR jusqu'en 2028.

	CANTO	COBLANCE	CKD-REIN	CRYOSTEM	E4N	HOPE-EPI	OFSEP
Personnel	54 %	62,7 %	37 %	39,9 %	66 %	57,3 %	42 %
Autres dépenses de fonctionnement et investissement	46 %	37,3 %	63 %	60,1 %	34 %	42,7 %	58 %

	PSY-COH	RADICO	HEPATHER	RECONAI	CONSTANCES
Personnel	40 %	50 %	60 %	26,4%	26,6 %
Autres dépenses de fonctionnement et investissement	60 %	50 %	40 %	73,6 %	73,4 %

Les variations entre cohortes qui ressortent de ces données doivent cependant être analysées avec prudence, dans la mesure notamment où le périmètre des dépenses de personnel prises en compte peut varier d'un projet à l'autre, et par exemple ne pas intégrer, ou intégrer en partie seulement et selon des quotités qui peuvent varier en fonction des relations financières entre les établissements contribuant au fonctionnement d'une cohorte, les dépenses de rémunération des attachés de recherche clinique chargés du recueil des données en phase d'inclusion et de suivi.

Ainsi par exemple CRYOSTEM, organisée autour d'une petite équipe de 5 personnes, fait apparaître une part des dépenses de personnel inférieure à 40 % sachant que le recueil des données donne lieu à des reversements financiers aux centres qui en sont chargés.

Pour l'ensemble des cohortes, et hors dépenses de personnel, l'essentiel porte sur les dépenses de fonctionnement courant, les investissements (1,9 % du total) et les frais de gestion (2,6 % du total) étant marginaux.

2.1.4 Taux d'exécution des dépenses à partir des financements du PIA en 2018

Le tableau ci-après retrace par cohorte, à partir des données issues du suivi de l'ANR, le taux d'engagement des dépenses à fin 2018 sur les financements reçus au titre du PIA. Ce taux qui atteint en moyenne 69 %, varie de 55,5 % pour CONSTANCES à 87 % pour CKD-REIN.

Cohortes	Subvention PIA ¹⁹ (en k€)	Total dépenses (en k€)	Taux exécution
CANTO	13 100	10 563	80,6 %
CKD-REIN	5 861	5 103	87 %
COBLANCE	3 945	2 598	65,8 %
CRYOSTEM	5 238	4 069	77,7 %
E4N	7 507	5 188	69,1 %
HOPE EPI	6 274	4 390	69,9 %
OFSEP	9 768	6 630	67,8 %
PSY-COH	3 846	2 206	57,3 %
RADICO	10 513	7 563	71,9 %
HEPATHER	10 010	8 316	83 %
RECONAI	13 018	9 076	69,7 %
CONSTANCES	35 000	20 877	59,6 %
Total	124 080	85 579	69 %

Compte tenu des délais de mise en place et de construction, le niveau d'exécution, qui ne prend en compte que les dépenses constatées fin 2018, s'avère globalement assez élevé. À la connaissance de la mission, il n'a pas servi à l'analyse des demandes de financement supplémentaire pour la période 2020-2025.

2.2 Période 2020-2025 (suite à l'évaluation de septembre 2019)

Le tableau suivant retrace les décisions financières qui viennent d'être proposées par le comité de pilotage de l'action Santé et biotechnologies du PIA à la suite de la réunion du jury scientifique qui s'est tenue en septembre 2019 pour procéder à l'évaluation finale des cohortes.

Ces montants vont donner lieu à une nouvelle subvention de la part de l'ANR, rapportée ici à la demande de financement présentée pour la période Juin 2020 – Juin 2025

En K€	CANTO	CKD-REIN	COBLANCE	CRYOSTEM	E4N	HOPE-EPI	OFSEP
Aide accordée	3 900	1 800	900	900	3 000	900	0
en % de la demande d'aide	82 %	82 %	55 %	68 %	40 %	29 %	0 %

¹⁹ Site ANR (sauf HEPATHER)

En K€	PSY-COH	RADICO	HEPATHER	RECONAI	CONSTANCES
Aide accordée	800	1 800	1 500	2 700	8 000
en % de la demande	53 %	51 %	52 %	100 %	45 %

Il est à noter que les financements de la cohorte OFSEP ne seront pas prolongés sur période 2020-2025, et que, a contrario, 8 M€ supplémentaires ont été dégagés pour le financement de l'infrastructure CONSTANCES. Pour les autres cohortes, les montants notifiés varient très fortement selon les projets (de 0,8 à 3,9 M€, soit de 29 à 100% des demandes financières présentées). Là encore, la mission n'a pas eu connaissance d'une grille de critères, en dehors de l'avis du jury rendu à l'issue de l'évaluation finale, expliquant ces variations, qui n'apparaissent pas aisément corrélables au contenu des demandes ou aux perspectives de cofinancement présentées par les porteurs de projets.

Par ailleurs, le ministère chargé de l'enseignement supérieur et de la recherche a proposé dans le même cadre de compléter pour 2020 sur crédits budgétaires ces financements, en faveur de la cohorte MEMENTO (0,6 M€) et des cohortes en population générale CONSTANCES (1 M€) et E3N/E4N (2,3 M€).

ANNEXE 3 : PANORAMA GENERAL DES COHORTES SOUTENUES DANS LE CADRE DU PIA

1 Cohortes en population

Nom du projet Responsable scientifique et technique Site	Objet et enjeux	Mode de fonctionnement / statut	Effectifs suivis	Établissement coordinateur	Aide PIA (2011-2019 en K€)
CONSTANCES Marie Zins https://www.constances.fr/	Construire une infrastructure de recherche reposant sur une vaste cohorte en population générale et comportant des données individuelles à spectre très large, issues de plusieurs sources (cliniques, médico-administratives,...)	Unité mixte de service Inserm / UVSQ	200 000 âgés de 18 à 69 ans suivis à partir de centres d'examen de santé de l'assurance maladie	Université de Versailles Saint Quentin (UVSQ)	35 000
E3N/ E4N Gianluca Severi https://www.e4n.fr/	Identifier les facteurs favorisant les principales maladies chroniques en suivant des familles sur trois générations	Equipe interne au centre de recherche en épidémiologie et santé des populations (CESP – unité mixte de recherche Inserm / UVSQ)	80 000 femmes adhérentes à la MGEN suivies depuis 1990 18 000 conjoints et 10 000 enfants	Université Paris Sud	7 507
RECONAI Marie-Aline Charles (ELFE) Pierre-Yves Ancel (Epiage2) https://www.elfe-france.fr/ http://epiage2.Inserm.fr/index.php/fr/	Constituer une plateforme de suivi sur longue périodes de deux cohortes d'enfants à la naissance, dont l'une de prématurés	Infrastructure mutualisée accueillant deux cohortes d'enfants suivis jusqu'à 18 ans : ELFE et Epiage2.	18 000 enfants à la naissance 7 800 enfants nés grands prématurés	Ined	13 018

2 Cohortes de patients

Nom du projet Responsable scientifique et technique Site	Objet et enjeux	Mode de fonctionnement / statut	Effectifs suivis	Établissement coordinateur	Aide PIA (2010-2018 en K€)
CANTO CANCER Fabrice André http://www.etudecanto.org/	Apprécier la toxicité des traitements chez les patientes porteuses d'un cancer du sein localisé	Etude portée par la fédération Unicancer, réseau hospitalier fédérant 18 centres de lutte anticancer (CLCC)	12 000 femmes atteintes d'un cancer du sein de stade I à III	Fédération Unicancer	13 100
CKD REIN Bénédict Stengel http://ckdrein Inserm.fr/index.php	Appréhender les facteurs d'évolution des maladies rénales chroniques	Equipe interne au centre de recherche en épidémiologie et santé des populations (CESP – unité mixte de recherche Inserm / UVSQ)	3 000 adultes atteints par une maladie rénale chronique non terminale	Université Paris Sud	5 861
COBLANCE Simone Benhamou https://www.coblance.fr/francais/l-etude/	Catégoriser et identifier les causes d'évolution du cancer de la vessie	Equipe interne au centre de recherche en épidémiologie et santé des populations (CESP – unité mixte de recherche Inserm / UVSQ)	1 120 adultes traités pour un cancer de la vessie confirmé histologiquement	Inserm	3 945
CRYOSTEM Régis Peffault de la Tour http://www.cryostem.org/en/	Comprendre et prévenir les complications de la greffe de moelle	Mise en réseau de l'ensemble des unités de greffe et de 28 centres de ressources biologiques	5 750 patients adultes et pédiatriques ayant eu une allogreffe de cellules souches hématopoïétiques	Société Française de Greffe de Moelle et de Thérapie Cellulaire (SFGM-TC) puis association CRYOSTEM	5 238
HEPATHER Fabrice Carrat	Évaluer les prises en charge des patients infectés par le virus des hépatites B et C	Recueil prospectif de données et la constitution de collections biologiques au sein de 38 centres experts en hépatologie.	21 000 patients atteints d'une hépatite B (6 000) ou C (15 000) au stade chronique de l'infection ou guéris	ANRS	10 010

RAPPORT IGAS N°2019-045R/IGESR N°2020-025

HOPE EPI Jacqueline Clavel	Mutualiser la collecte d'informations sur les cancers de l'enfant et progresser sur l'identification des risques environnementaux et génétiques	Projet adossé au Centre de recherche en épidémiologie et statistiques (CRESS), UMR Inserm / université Paris Descartes	25 000 personnes ayant été atteintes par un cancer avant l'âge de 18 ans	Inserm	6 274
OFSEP Sandra Vukusic http://www.ofsep.org/fr/	Construire un outil épidémiologique de suivi de la sclérose en plaques	Recueil de données dans centres EDMUS (logiciel permettant de recueillir les données cliniques des malades	62 000 patients dans la base de données 1600 dans la cohorte dite « haute définition » recueillant des observations plus détaillées	Université Claude Bernard – Lyon 1	9 768
PSY-COH Marion Leboyer https://www.psy-cohorte.fr/	Suivre les patients atteints de schizophrénie et de troubles bipolaires	Mise en réseau des centres experts FondaMental (11 sur les troubles bipolaires et 10 sur la schizophrénie)	800 sujets atteints de troubles bipolaires et 452 de schizophrénie	Fondation FondaMental	3 846
RADICO Serge Amselem https://www.radico.fr/fr/	Construire une plateforme mutualisée permettant de développer des cohortes de patients atteints de maladies rares	Infrastructure mutualisée support pour 13 cohortes	13 000 patients atteints de maladies rares (de 60 à 2600 selon la maladie)	Inserm	10 513

ANNEXE 4 : Présentation des 14 cohortes soutenues dans le cadre du PIA

La présente annexe propose une présentation synthétique des 14 cohortes soutenues dans le cadre du PIA, et qui relèvent, comme le montre le tableau synoptique figurant en annexe 3, de plusieurs catégories :

- les projets retenus au titre de l'appel à projet « Cohortes » de 2010-2011 ;
- les deux projets retenus au titre de la deuxième vague des Equipex lancée en 2011-2012 : HEPATHER et RECONAI ;
- la cohorte Constances retenue au titre des « Infrastructures nationales en biologie et santé » (INBS) également sélectionnées en 2012 ;
- la cohorte Memento portant sur des patients atteints de troubles de la mémoire, dont la DGRI a pris le relais du financement à la suite de la Fondation Alzheimer.

Ces cohortes sont ici présentées en distinguant les cohortes en population (CONSTANCES, E3N/E4N, et RECONAI) et les cohortes de patients diagnostiqués ou traités, classées par ordre alphabétique. Chacune d'elle fait l'objet d'une fiche distincte, réalisée à partir d'une grille commune et reposant sur trois principales sources d'information :

- le dossier demandé par la mission aux responsables du dispositif et portant sur les modalités de constitution de la cohorte, les modes de recueil et d'exploitation des informations, les projets et résultats de recherche associés, la gouvernance du dispositif, les ressources et moyens mobilisés et attendus, et les règles d'accès aux données constituées ;
- des sessions de travail d'une demi à une journée réalisées par la mission avec les équipes promotrices de chacune de ces cohortes : organisées autour d'un schéma-type, elles ont permis des rencontres avec les responsables scientifiques des cohortes et leurs collaborateurs (y compris en charge des aspects informatiques, administratifs et financiers), avec des chercheurs appartenant ou non à leur équipe et associés à l'élaboration ou à l'utilisation du dispositif, ainsi qu'avec certains des partenaires (institutionnels, associatifs, privés) les plus directement impliqués ou intéressés par l'infrastructure de recherche ;
- la série d'entretiens conduits par la mission auprès des différents acteurs ou partenaires concernés des mondes de la santé et de la recherche, qui ont permis de recueillir leur point de vue sur le bilan, les apports et l'utilisation de chacun de ces projets.

L'existence de rapports de synthèse élaborés en vue de l'évaluation scientifique programmée en septembre 2019, et leur diffusion préalable au jury comme aux autorités de tutelle, a toutefois conduit la mission à ne reprendre que de façon rapide et succincte dans ces fiches les informations relatives à la description et à l'avancement des projets, en les enrichissant par quelques éléments d'observation et d'appréciation recueillis à l'occasion de ses travaux.

Un commentaire final, dans une rubrique intitulée « Appréciations portées par la mission », y ajoute un avis d'ensemble sur les perspectives de la cohorte en termes d'horizon, de pérennité, de réorientation ou d'enrichissement scientifiques, sur les besoins en ressources et en compétences associés, ainsi que les problèmes institutionnels ou de gouvernance demeurant selon la mission à résoudre à un horizon de court ou moyen terme.

Cet avis est issu de l'examen des dossiers et des discussions que la mission a pu avoir avec les responsables des cohortes lors des journées de travail organisées avec eux-mêmes et leurs équipes, sachant que, concernant l'avancement et les apports scientifiques des projets, elle a aussi pu bénéficier de l'appréciation des jurys d'évaluation (de 2017 et 2019), dont elle a eu communication.

LA COHORTE CONSTANCES

1 Présentation générale du dispositif

1.1 Objectifs

- Constituer une infrastructure de recherche reposant sur une vaste cohorte en population générale (200 000 personnes âgées de 18 à 69 ans)
- Mettre à disposition des chercheurs des données individuelles à spectre très large, issues à la fois d'examens de santé périodiques et de questionnaires sur l'état et les comportements de santé, les habitudes et la qualité de vie, l'environnement social et professionnel, appariées avec des données administratives sur les carrières et sur les soins remboursés par l'assurance maladie.

1.2 Principales caractéristiques de la cohorte

- Cohorte généraliste couvrant des sujets d'âge adulte et réunissant des informations sur de multiples aspects de leur état et de leurs problèmes de santé, ainsi que sur leur situation sociale et professionnelle
- Inclusion et suivi de volontaires à partir de 21 centres d'examens de santé de l'assurance maladie
- Constitution d'une base d'études prospective visant à la fois à permettre d'évaluer l'impact de d'une large série déterminants et facteurs d'exposition sur un vaste ensemble d'individus et d'assurer une représentativité en population générale pour estimer la prévalence de difficultés problèmes de santé publique et les modalités de leur prise en charge
- Poursuite du suivi des problèmes de santé au travail, suite à l'expérience historique de la cohorte Gazel, centrée sur les relations entre travail, environnement social et santé physique ou psychologique chez 16 000 salariés d'EDF-GDF, et en lien avec les cohortes COSET mises en place par Santé publique France pour les travailleurs agricoles et indépendants (*cf. infra*)
- Infrastructure de recherche ayant vocation à être ouverte à des chercheurs sur la base de projets présentés à un conseil scientifique

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

Après une phase pilote, l'inclusion dans la cohorte Constances a démarré de façon graduelle à partir de 2012 : le nombre de participants était de 174 000 en mai 2018 et a atteint l'objectif fixé de 200 000 au printemps 2019.

La perspective est celle d'un suivi sur longue période sans terme fixé au départ.

2.2 Modalités d'inclusion et d'échantillonnage

Les sujets éligibles sont tirés au sort dans le répertoire des bénéficiaires de l'assurance maladie des travailleurs salariés ; le sondage stratifié vise à permettre une représentativité (en termes d'âge, de sexe, de statut d'activité et de PCS) des assurés du régime général de sécurité sociale (c'est-à-dire hors travailleurs indépendants et assurés relevant du régime agricole, qui sont suivis via des cohortes spécifiques mises en place sous l'égide de l'agence Santé publique France).

Les participants sont des volontaires acceptant de remplir un ensemble de questionnaires à l'inclusion, ainsi que de venir dans un centre d'examen de santé de l'assurance maladie pour un examen de santé permettant le recueil de diverses données physiologiques, biologiques et cliniques. Ils consentent ensuite à participer à un suivi sous forme d'auto-questionnaires annuels, d'examens de santé périodiques tous les cinq ans, et de recueil de leurs événements professionnels ou de santé enregistrés sur les fichiers de la CNAV et par le système national des données de santé (SNDS).

Le taux d'acceptation étant d'environ 8 % de la population totale échantillonnée, une cohorte de 400 000 non-participants a par ailleurs été constituée à partir des seules données des bases administratives (suivi passif) pour procéder à des redressements permettant réduire les biais de sélection dans l'estimation de prévalences en population générale. Le comité du label de la statistique publique a souligné en 2014 l'importance de cette question pour la validité du dispositif.

Environ 75 % des participants initiaux ont rempli le questionnaire de suivi annuel, et entre 45 et 50% se sont déplacés pour les premières vagues d'examens afférents au suivi périodique. Cette attrition appelle comme les biais de sélection un suivi et des redressements particuliers, sachant que le recueil continu des données administratives et la possibilité pour un sujet de « revenir » ultérieurement dans la cohorte peuvent en réduire la portée.

La cohorte Constances est pour l'instant une cohorte fermée, sachant que les cohortes COSET qui en sont le pendant pour les travailleurs indépendants et les travailleurs agricoles sont de leur côté conçues comme des cohortes ouvertes, devant régulièrement inclure de nouveaux participants.

2.3 Contenu, collecte et stockage des différents types d'informations

Les données recueillies par questionnaire sont destinées à décrire et suivre dans le temps :

- des caractéristiques sociodémographiques, de statut et de situation sociale : situation et activité professionnelle, niveau d'études, niveau de revenus, situation matrimoniale, composition du ménage, statut socio-économique des parents et du conjoint, conditions de vie matérielles... ;
- des informations relatives à la santé : antécédents personnels et familiaux, échelles de santé auto-déclarée, liste de pathologies déclarées et traitées, problèmes de santé propres aux femmes ;
- des informations sur les comportements : consommations de tabac et d'alcool, habitudes alimentaires et activité physique, usage du cannabis, orientation sexuelle ;
- des facteurs professionnels : histoire professionnelle, expositions aux risques professionnels et psycho-sociaux.

Dans les centres de santé, sont recueillies des données du type :

- examen clinique (poids, taille, tension artérielle) et paraclinique (vision, audition, électrocardiogramme...),
- investigations biologiques (glycémie, bilans lipidique et hépatique, numération sanguine, examen urinaire ...);
- tests des capacités physiques et cognitives pour les personnes âgées de 45 ans et plus.

Les données recueillies auprès des individus participant à Constances font en outre l'objet d'appariements avec les bases de données administratives et médico-administratives nationales :

- bases de la Cnav pour les données socioprofessionnelles (épisodes professionnels et données de carrière pour les salariés du régime général, trimestres validés au titre du chômage, de la maladie et des périodes assimilées);
- bases du CépiDC-Inserm pour les causes de décès ;
- accès au SNIIRAM et au PMSI pour les événements de santé pris en charge par l'assurance maladie et/ou les établissements de santé.

Compte tenu des exigences de respect d'un référentiel de sécurité défini par la loi pour l'accès aux données de santé, les données de la cohorte Constances sont depuis 2019 hébergées par le centre d'accès sécurisé aux données (CASD) qui offre l'accès à une bulle sécurisée où sont entreposées ces données, en l'attente de la mise en place par l'Inserm de dispositifs spécifiques en la matière.

Une biobanque incluant des échantillons de sang et d'urine a par ailleurs été récemment constituée pour un objectif de 85 000 individus ; des projets en matière de génotypage et de séquençage du génome sont envisagés pour une part d'entre eux à l'avenir.

Cette biobanque a fait l'objet d'un marché européen, conclu avec l'institut IBBL (*Integrated BioBank of Luxembourg*), situé au Luxembourg.

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

La principale caractéristique de Constances est l'appariement d'informations recueillies sur un large éventail de sujets ayant trait à la santé et aux modes de vie au travers de questionnaires et d'examen cliniques avec les données du SNDS sur les parcours de soins et de la CNAV sur les parcours professionnels.

Les appariements avec le SNIIR-AM sont enrichis par un ensemble de variables construites, de tables et d'algorithmes simplifiés. L'équipe de Constances anime un réseau, REDSIAM, qui mobilise plusieurs groupes thématiques pour améliorer l'identification et le repérage des pathologies à partir du SNIIR-AM.

Les données collectées sur les revenus des ménages sont *a contrario* recueillies sans appariement ou rapprochement de celles des bases sociales et fiscales gérées par l'Insee.

La cohorte Constances, centrée sur les salariés du régime général, a pour pendant deux cohortes COSET portées par Santé publique France, et constituées à partir d'assurés de la MSA et de l'ex-RSI ; elles font l'objet des mêmes appariements et du même type de repérage des déterminants professionnels des problèmes de santé que Constances.

Les articulations ou intégrations envisagées à l'origine avec d'autres cohortes en population générale comme Nutrinet ne sont par contre pas entrées dans les faits. La coopération se fait de façon plus indirecte sous forme de confrontation des réponses ou de répartition des sujets d'investigation concernant les questionnements relatifs aux comportements alimentaires.

Les diverses collaborations en cours ou en projet avec d'autres cohortes visent enfin soit à leur fournir une « base témoin » en population générale, soit à leur permettre de mettre en place des recueils complémentaires de données sur des sous-populations d'intérêt spécifique. Constances a par exemple été utilisée pour mettre en place certaines cohortes portées par l'AP-HP ou l'étude Cosmos-France, soutenue par l'ANSES, sur les risques associés aux téléphones mobiles et aux technologies sans fil.

2.5 Horizon et perspectives envisagés pour le dispositif après 2020

La cohorte Constances a été conçue comme un dispositif permanent, susceptible de servir à des utilisateurs et des projets de recherche diversifiés, sans limite de temps fixée à l'avance.

Ses promoteurs envisagent un suivi longitudinal poursuivi sur longue durée et enrichi au fur et à mesure par de nouvelles informations, concernant par exemple les expositions professionnelles, le contexte environnemental ou les analyses génétiques.

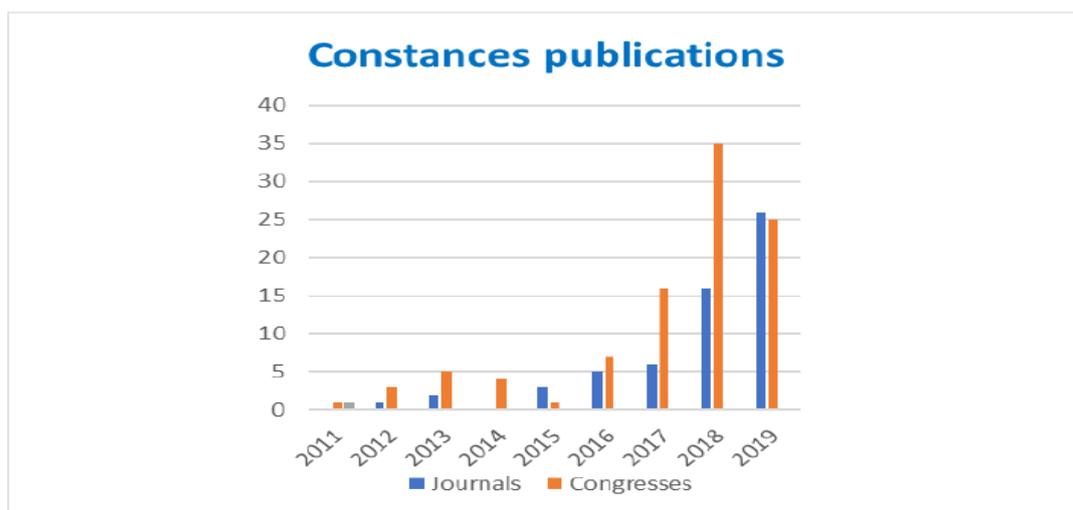
La question de la réouverture à terme de la cohorte à de nouvelles inclusions (pour la compensation de l'attrition ou le recrutement de jeunes adultes) n'a toutefois pas été à ce stade envisagée.

3 Principaux résultats et retombées en termes de recherche et de santé publique

Les responsables de la cohorte ne portent pas eux-mêmes de projets de recherche qui leur soient propres, et apprécient les résultats du dispositif à l'aune du nombre, de la variété et de la qualité des projets de recherche qui sont proposés et commencent à être menés à bien par une diversité d'équipes de recherche.

Les projets de recherche approuvés par les comités scientifiques et de pilotage de la cohorte étaient début 2019 au nombre de 80, mobilisant environ 190 utilisateurs académiques. Les domaines couverts sont extrêmement divers, allant des problèmes de santé engendrés par les risques professionnels et environnementaux, aux déterminants et conséquences des pathologies chroniques (notamment les maladies cardio-vasculaires ou l'asthme), aux déterminants sociaux de la santé, au vieillissement cognitif ou aux trajectoires de prévention, de dépistage et de soins.

L'activité de publications dérivée de ces recherches est fortement croissante. Elle passe progressivement d'articles méthodologiques décrivant les fondements scientifiques de la cohorte à la publication de travaux décrivant les relations entre des facteurs de risque et maladies ou les marqueurs précoces de ces dernières.



Des journées scientifiques sont en outre régulièrement organisées (par exemple en 2013, 2017 et 2019) autour de projets et résultats de recherche associés à la cohorte.

La cohorte Constances participe à une série de *consortia* internationaux, dont cinq en Europe et, au niveau mondial, le *100K Cohort Consortium* (IHCC) qui regroupe un ensemble de très grandes cohortes populationnelles.

Les évaluations des jurys scientifiques réunis par l'ANR en 2016 et 2019 ont été positives, saluant le caractère stratégique de l'infrastructure. En 2016, le jury a insisté sur la mise en place du projet de biobanque, la procédure de gestion des données personnelles et l'association des participants. En septembre 2019, il a souligné la nécessité d'assurer une poursuite et une soutenabilité à long terme à la cohorte, appelant de la part du comité de pilotage du PIA des demandes concernant une gouvernance consolidée et un renforcement de l'engagement des tutelles, au-delà du soutien financier complémentaire envisagé jusqu'en 2025.

Les partenaires et utilisateurs institutionnels rencontrés par la mission ont une vision assez disparate des retombées et des utilisations potentielles de la cohorte :

- la Direction générale de la santé (DGS) la subventionne pour des éclairages, par exemple dans le domaine des maladies cardio-vasculaires ;
- Santé publique France s'y intéresse pour valider certaines estimations de prévalence (BPCO, diabète) ainsi qu'en matière de santé environnementale et de santé au travail, en couplage avec ses cohortes COSET ;
- la Cnam met surtout l'accent sur l'apport de Constances à la structuration de l'activité des centres d'examen de santé et à l'identification des pathologies dans le SNIIR-AM (projet REDSIAM), mais assez peu sur ses utilisations directes. Elle, comme l'Irdes, indiquent que peu de focus existent à partir de Constances sur les sujets de régulation des soins, qui les intéressent le plus directement ;
- la Drees utilise quant à elle peu les cohortes épidémiologiques : elle a seulement utilisé Constances comme base de sondage dans le cadre d'une enquête spécifique sur les délais d'attente ;
- la Dares, tout en regardant les résultats des recherches issues de Constances, n'en utilise pas non plus directement les données, se limitant à ses propres enquêtes. Dares et Drees ont toutefois financé projet de recherche prenant appui sur Constances dans le cadre d'un appel à

projets sur la santé mentale. La Dares considère néanmoins qu'en matière de santé au travail, les enquêtes statistiques et une cohorte comme Constances sont fortement complémentaire, avec des qualités et des défauts qui sont symétriques les uns des autres (représentativité versus richesse des données médicales et biologiques recueillies).

4 Gouvernance du dispositif

La gouvernance de la cohorte est classiquement organisée à partir d'un comité de pilotage institutionnel, présidé par la Cnam, et d'un comité scientifique international.

Elle prend place dans le cadre d'une unité mixte de services (UMS) associant l'Inserm et l'Université de Versailles Saint-Quentin (UVSQ).

L'accord de consortium initial de 2015 associe à ces deux opérateurs la Cnav (tirage au sort des éligibles, apport de données) et la Cnam (6 équivalents-temps plein portés par le CETAF, participation des centres d'examens de santé aux examens cliniques et au suivi des participants, fourniture des données issues du SNDS). L'UVSQ porte la gestion des financements issus de l'ANR, notamment celle des personnels recrutés et de la masse salariale, tandis que l'Inserm est juridiquement responsable du traitement des données et de la gestion de certains financements sur projets.

La responsable scientifique du programme ayant rejoint l'université Paris-Descartes, un avenant est en cours de discussion visant à associer cette dernière (ainsi que la CMUE Paris Saclay) au consortium, tandis que les engagements demandés par l'ANR aux différents partenaires sur leurs apports et leur implication respective dans la poursuite terme du projet à moyen terme doivent donner lieu à discussion.

La cohorte a par ailleurs d'autres partenaires institutionnels comme la caisse d'assurance maladies des industries électriques et gazières (CAMIEG), qui a une longue histoire de collaboration avec la précédente expérience GAZEL, et une série de mutuelles (dont la MGEN, la MFP et la LMDE).

Les entreprises apportant un financement à la cohorte peuvent en outre exprimer leurs demandes dans le cadre d'un comité des partenaires industriels, tandis qu'un « club d'ambassadeurs » est en projet pour assurer des contacts plus étroits avec les participants.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

Les financements attribués à la cohorte CONSTANCES dans le cadre du programme relatif aux INBS du PIA ont été de 35 M€ sur la période 2012-2019. S'y sont ajoutés 2,4 M€ de financements ministériels pour la consolidation et la sécurisation de l'hébergement et du stockage des données.

Les diverses contributions des partenaires institutionnels ont par ailleurs été estimées à un total de 115 M€, dont environ 112 M€ en provenance de la Cnam pour la réalisation des examens de santé.

Par ailleurs, CONSTANCES a bénéficié de 2 M€ de redevances versées par ses utilisateurs, de 0,2 M€ de fonds européens et de 4,5 M€ au titre de partenariats industriels.

Un poste de dépenses important concerne la rémunération des personnels recrutés au titre de l'infrastructure, pour des coûts annuels d'environ 1,3 M€. Celle-ci ne bénéficie en effet que de deux postes statutaires (1 ingénieur Inserm, 1 poste de PU-PH pour l'investigateur principal au sein l'Université Paris-Descartes), en plus des 9 ETP mis à disposition par la Cnam par l'entremise du Centre technique d'appui et de formation des centres d'examen de santé (CETAF). Les personnels recrutés en CDD au titre du projet sont au nombre de 18 ETP, et sont administrativement « portés » par l'UVSQ.

5.2 Perspectives budgétaires pour la période 2020-2024

Les financements supplémentaires demandés par la cohorte à l'ANR pour la période 2020-2024 sont de 18 M€, pour des dépenses totales prévues de 29 M€ et d'autres ressources attendues à hauteur d'environ 11 M€ (550 000 € par an au titre de redevances pour l'utilisation de l'infrastructure, 450 000 € par an de subventions diverses et 1,1 M€ par an en provenance de financements industriels-*cf. infra*).

Les CDD portés par l'UVSQ ont une échéance finale prévue entre 2020 et 2024, impliquant des décisions à prendre par la tutelle quant à la consolidation de ces emplois.

Les responsables de l'infrastructure disent en outre avoir besoin de renforts en compétences tant en matière d'épidémiologie statistique que d'ingénierie informatique et statistique (*data manager*, exploitation statistique du SNDS). Même si certains d'entre eux pourraient être mutualisés dans le cadre du projet France Cohortes envisagé par l'Inserm (en plus d'une éventuelle assistance en matière juridique et budgétaire), leurs demandes budgétaires incorporent six postes supplémentaires à ce titre pour la période 2020-2024.

Comme suite à l'évaluation finale, le comité de pilotage chargé d'instruire la demande de financement complémentaire a proposé en novembre 2019 un complément de financement du projet à hauteur de 8 M€ jusqu'au 30 juin 2025. Dans ce cadre le ministère chargé de l'enseignement supérieur et de la recherche a proposé de compléter cette enveloppe sur crédits budgétaires à hauteur d'1 M€ pour 2020.

5.3 Les partenariats industriels

Les partenariats industriels de l'infrastructure restent limités en nombre et en volume, ce qui est pour partie lié aux caractéristiques généralistes de la cohorte qui limite ses potentialités de valorisation économique, ainsi que, dans une moindre mesure, à l'intention de ses responsables de lui conserver à ses utilisations un caractère principalement public.

Pour le moment, seules deux grandes entreprises, MSD et L'Oréal, ont un partenariat pluriannuel avec l'infrastructure, principalement pour l'accès à des tableaux de bord et la possibilité de conduire des projets de recherche sur des populations d'intérêt spécifique. Deux autres entreprises, Astra Zeneca et Lundbeck, ont aussi eu ce type de partenariat entre 2014 et 2018, mais ne l'ont pas renouvelé.

Ces partenariats privés sont négociés et gérés par Inserm-Transfert ; ils donnent lieu à la réunion d'un « comité des partenaires industriels » destiné à faire le point régulièrement et de façon transparente sur les projets et les besoins.

6 Accès aux données

CONSTANCES est une infrastructure dont l'objet même est de permettre un accès aux utilisateurs institutionnels et aux chercheurs, le cas échéant en leur apportant un appui pour la définition de leurs projets et l'utilisation du SNDS. Un appel à projets permanent a à cet égard été lancé depuis 2013, qui est depuis permanent.

L'accès aux données donne lieu à une charte, qui prévoit que les projets de recherche soient expertisés par l'équipe de la cohorte, soumis à son Conseil scientifique, puis approuvés par son comité institutionnel. Des participations financières sont demandées aux promoteurs des projets, à hauteur de 15% des subventions publiques reçues pour les projets académiques, et au travers de contributions pluriannuelles pour les organismes publics, caritatifs ou les entreprises privées.

Le conseil d'éthique de l'Inserm est le cas échéant consulté en cas de besoin.

7 Appréciations portées par la mission

La cohorte CONSTANCES a été conçue comme une infrastructure épidémiologique de grande taille et à spectre large destinée à fournir un outil de long terme aux chercheurs et aux institutions intervenant dans le domaine de la santé publique. Ces objectifs sont globalement bien avancés au stade actuel de développement du projet, et l'apport de la cohorte est scientifiquement et institutionnellement reconnu. Toutefois, le dispositif présente un certain nombre de fragilités et d'incomplétudes qui posent question quant à sa capacité à assurer ces objectifs dans la durée.

À court-moyen terme, les problèmes concernent :

- en matière de ressources humaines, la consolidation des postes à durée déterminée et le renforcement de certaines compétences en *data management* ou bio-statistique ;
- l'engagement des tutelles et l'inscription du projet dans les politiques universitaires de sites, tous deux conditions du bon fonctionnement de la gouvernance du dispositif (clarification de l'engagement de l'Inserm, introduction de l'Université de Paris Descartes dans le consortium des partenaires institutionnels et dialogue régulier entre eux).

Contreparties à la poursuite des financements de l'ANR, leur solution nécessite un accord rapide entre partenaires (ou des arbitrages au niveau ministériel) sur leurs conditions de coopération et leurs engagements respectifs de soutien à la cohorte. Dans cette perspective, le maintien de l'hébergement des données de Constances au sein du CASD ou son transfert vers la future plateforme envisagée par le projet France Cohortes feront partie des décisions importantes à prendre.

À moyen-long terme, et en dépit des enrichissements envisagés entre 2020 et 2024 (diagnostics détaillés, histoire des expositions professionnelles et des situations résidentielles, données environnementales sur la zone de résidence), Constances n'a pas dans l'état actuel des choses l'ensemble des caractéristiques permettant d'en réaliser toutes les potentialités, non plus que d'assurer la totalité des fonctions d'une grande cohorte généraliste que la France pourrait souhaiter dans le long terme. On peut noter par exemple que :

- sa biobanque est juste en voie de constitution et a été sous-traitée à un organisme situé hors du territoire français ;

- les analyses génomiques permettant mieux éclairer les combinaisons et interactions entre facteurs de risque n'ont pu être intégrées aux projets actuels de développement de la cohorte, alors qu'il s'agit d'un élément clé dans certains dispositifs étrangers et d'un atout dont disposent d'autres cohortes généralistes comme E3N/E4N ;
- les revenus et le niveau de vie des ménages auxquels appartiennent les participants à la cohorte sont recueillis et calculés à partir des réponses individuelles aux questionnaires, sans rapprochement avec les données socio-fiscales par ailleurs collectées et traitées par l'Insee, ce qui soulève des questions de fiabilité et de comparabilité dans l'appréhension des inégalités sociales ;
- la réouverture de la cohorte à de nouveaux participants n'est pas prévue à ce stade ; en son absence, le dispositif ne pourrait plus à terme viser une représentativité en population générale, et ce alors même que les cohortes COSET qui lui sont couplées pour les travailleurs agricoles et indépendants ont, elles, prévu d'être ré-ouvertes.

La tension est à cet égard manifeste entre l'ampleur du projet, son caractère encore incomplet et la taille et le statut fragiles de son équipe.

Plus fondamentalement, les coopérations avec les autres grandes cohortes ou dispositifs longitudinaux généralistes (notamment les cohortes Elfe et NUTRINET, mais aussi l'extension de la cohorte E3N/E4N aux conjoints et enfants des femmes suivies) sont aujourd'hui envisagées *ex-post* et ciblées sur certaines thématiques de recherche ou questions méthodologiques particulières (techniques de fidélisation des participants, par exemple), sans que les recueils d'informations soient articulés en amont (à l'exception des appariements avec le SNDS). La question se pose donc de l'évolution et de la convergence à envisager sur le long terme pour ces dispositifs, qui, en dépit des intentions initiales, ont développés dans le cadre de projets séparés et avec des temporalités différentes.

LA COHORTE E3N-E4N

1 Présentation générale du dispositif

1.1 Objectifs

- Compléter la cohorte E3N, lancée en 1990 et incluant près de 99 000 femmes relevant de la Mutuelle générale de l'éducation nationale (MGEN), par des cohortes complémentaires concernant leurs conjoints, leurs enfants et leurs petits enfants
- Élargir, en poursuivant le suivi des participantes, le champ d'étude initial de l'étude E3N, à savoir les facteurs nutritionnels et hormonaux en lien avec les cancers féminins, à d'autres pathologies chroniques majeures et étendre ces analyses aux hommes d'une part, aux effets familiaux et intergénérationnels susceptibles d'influer sur ces pathologies d'autre part.
- Constituer une base de données « familiale » comportant, en termes d'objectif final, le suivi de jusqu'à 20 000 familles et 200 000 individus.

1.2 Principales caractéristiques de la cohorte

1.2.1 Une base préexistante : la cohorte E3N :

- Cohorte généraliste constituée en 1990 à partir des femmes adhérentes à la MGEN nées entre 1925 et 1950.
- Suivi par auto-questionnaires de leur état de santé et leur mode de vie tous les deux à trois ans, complété par une biobanque intégrant des échantillons sanguins (pour 25 000 sujets) et salivaires (pour 47 000 sujets), des données d'imagerie et une tumorotheque.

1.2.2 Trois cohortes additionnelles proposées dans le cadre du PIA

Le projet proposé au PIA portait sur l'adjonction à E3N d'une cohorte dite E4N ciblée sur trois sous-groupes spécifiques, dans la perspective évoquée plus haut d'évolution et d'élargissement du champ de l'étude :

- les pères de leurs enfants, qui constituent aujourd'hui comme leurs conjointes un groupe de sujets âgés ou très âgés (groupe G1) ;
- leurs enfants, dont la moyenne d'âge est aujourd'hui de l'ordre de 55 ans (groupe G2);
- leurs petits-enfants (groupe G3).

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

2.1.1 Pour la cohorte E3N

Le recrutement a débuté en 1990 et a concerné 98 995 femmes adhérentes à la MGEN âgées à l'époque de 40 à 65 ans.

Le suivi continue à concerner trente ans plus tard une part importante de cette population, avec une cible d'environ 80 000 femmes contactables en 2018.

2.1.2 Pour la cohorte E4N

Le recrutement des pères a démarré en octobre 2014. 18 000 d'entre eux ont répondu favorablement à la première invitation.

L'inclusion des enfants a pris du retard, compte tenu des délais de construction d'une plateforme, reposant désormais sur des questionnaires numérisés (*cf. infra*). Leur recrutement a débuté seulement en octobre 2018 et avait concerné en septembre 2019 environ 10 000 « enfants ». L'inclusion doit se poursuivre au dernier trimestre 2019 et au cours de l'année 2020.

Le recrutement des « petits-enfants » a quant à lui été retardé et n'est envisagé qu'à la fin de 2020.

2.2 Modalités d'inclusion et d'échantillonnage

2.2.1 Pour la cohorte E3N

Les femmes de la cohorte E3N ont été recrutées en 1990 parmi les adhérentes à la MGEN âgées de 40 à 65 ans, et ont donc aujourd'hui entre 70 et 95 ans. Les participantes remplissent une série de questionnaires tous les deux à trois ans, et fournissent des échantillons de salive.

Cette affiliation professionnelle implique que la cohorte ne vise pas la représentativité et ait principalement des objectifs de « validité interne » ; elle n'empêche pas une certaine variabilité des situations socio-économiques des participantes, suffisante, selon les responsables de la cohorte, pour appréhender l'articulation de ces facteurs avec les informations recueillies en matière génétique et d'état de santé.

Depuis 1990, 13 % des femmes sont décédées et 4% ont été complètement perdues de vue. 12% des participantes n'avaient en outre pas répondu aux questionnaires de 2011 et 2014, sans être forcément considérées comme « perdues de vue ».

Au bout du compte, le questionnaire de juin 2018 (le 12^e du genre) a recueilli, après une relance, des réponses de la part de 56 000 femmes sur 82 000 contactées, soit un taux de réponse de 68% et un suivi concernant encore 57% des femmes incluses dans E3N vingt-huit ans plus tôt.

D'après les estimations des responsables de la cohorte, 72 000 participantes devraient encore être « contactables » en 2021 et 50 000 en 2027.

2.2.2 Pour la cohorte E4N

Les contacts ont dans un premier temps concerné les pères restés en couple avec les femmes de la cohorte. Les 18 000 d'entre eux ayant répondu à l'invitation de 2014 représentent 57% des hommes contactés (environ 32 000), et ont tous donné un échantillon de salive. 13 000 d'entre eux, soit 71%, ont répondu au questionnaire de suivi de juin 2018.

Les 10 000 acceptations « d'enfants » réunies à l'été 2019 l'ont été à partir de contacts noués par l'entremise des mères et d'adresses e-mails connues. Les contacts prévus à la fin 2019 et en 2020 devaient quant à eux viser plus de 93 000 autres personnes par le biais d'une invitation postale.

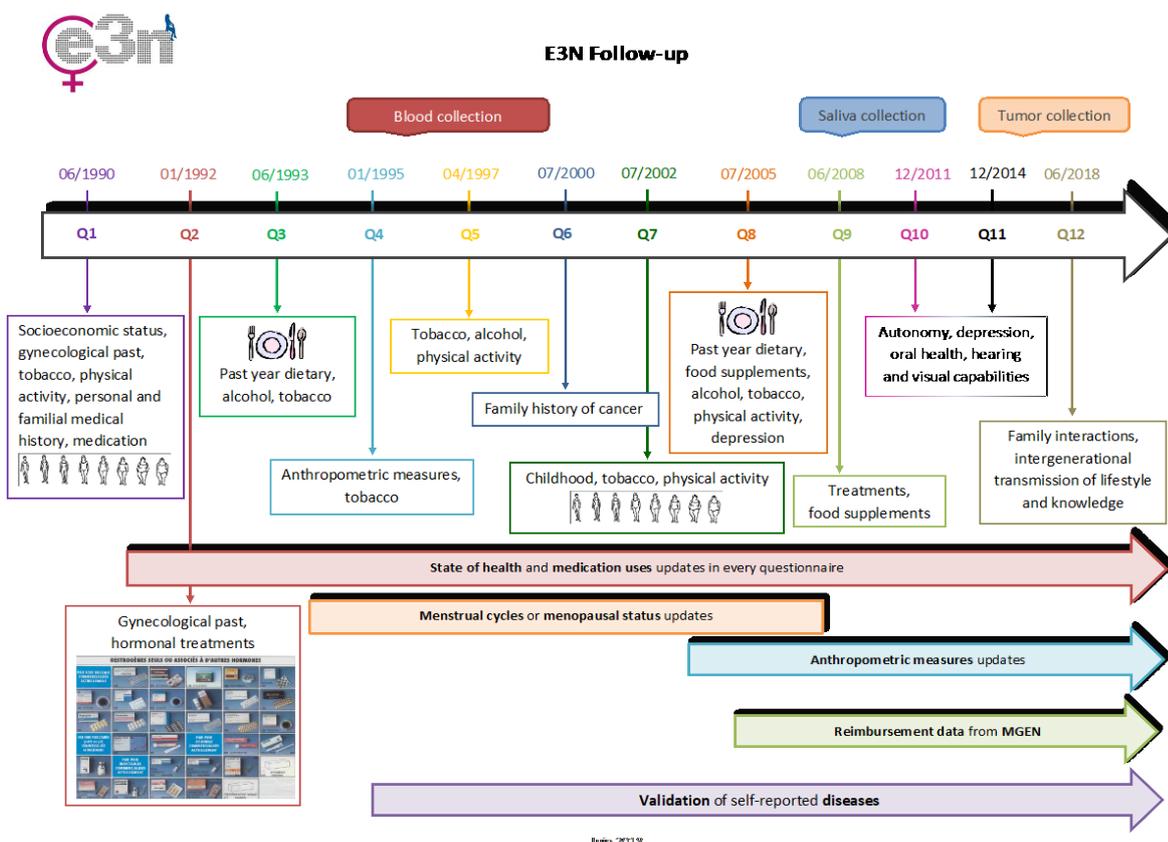
Les modalités de recrutement des « petits-enfants » n'ont à ce stade pas encore été précisées.

Ces modes de sélection illustrent le ciblage tout à fait spécifique de la cohorte E4N, qui vise à analyser les facteurs de transmission familiaux des caractéristiques d'état de santé (qu'elles tiennent à la génétique ou aux modes de vie), sans aucune visée de représentativité. Ils n'excluent néanmoins pas des biais de sélection concernant les répondants parmi les conjoints et les enfants.

2.3 Contenu, collecte et stockage des différents types d'informations

2.3.1 Pour la cohorte E3N

Les questionnaires auxquels répondent les participantes tous les deux à quatre ans dont une partie des thématiques est commune (situation socio-économique et familiale, état de santé, traitement, mesures anthropométriques) et l'autre évolutive, en fonction de sujets d'intérêt particuliers et de problématiques liées à l'avancement en âge (déficiences sensorielles, autonomie).



La cohorte a également donné lieu au recueil d'échantillons de salive (47 000 en 2009-2010), de sang (25 000 entre 1996 et 1999), ainsi que, dans certains cas, de mammographies (pour 9000 femmes) et de tumeurs.

2.3.2 Pour la cohorte E4N

Concernant les pères des enfants, les auto-questionnaires d'inclusion et de suivi portent comme pour E3N sur un ensemble large de thématiques ayant trait à l'état de santé et aux modes de vie (situation socio-économique et familiale, état de santé, traitement, mesures anthropométriques...). Des échantillons de salive sont également systématiquement collectés.

Pour les enfants et petits-enfants, une nouvelle conception des interrogations a été mise en œuvre de façon à proposer aux participants, *via* des questionnaires pluriels et de taille réduite, des interactions moins longues et plus fréquentes et des réponses par la voie de canaux numériques diversifiés. Ces informations sont intégrées dans une plateforme web conçue avec la société EPICconcept, qui sera également susceptible de recevoir des données en provenance d'objets connectés (par exemple des bracelets électroniques ou des patchs pour le suivi de l'activité physique d'une fraction des participants).

2.3.3 Stockage des données

Les bases de données afférentes aux cohortes E3N et E4N sont hébergées sur les serveurs de l'Institut Gustave Roussy.

Les prélèvements sanguins réalisés pour E3N ont été stockés par l'Établissement français du sang (EFS) à Annemasse, et les échantillons salivaires et de tumeurs par le Centre de ressources biologiques (CRB) de l'Institut Gustave Roussy (IGR) ; cet Institut considère néanmoins arriver aujourd'hui à la limite de ses capacités.

Des contacts sont en cours avec la fondation Jean Dausset - Centre d'Etude du Polymorphisme Humain (CEPH) - et le Centre National de Recherche en Génétique Humaine (CNRGH) à Évry pour reconcevoir le stockage de l'ensemble des échantillons biologiques.

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

Les cohortes E3N-E4N sont appariées aux données de mortalité du CepiDC, ainsi que, s'agissant d'E3N et depuis 2004, avec les données de remboursement de la MGEN transmises sur une base trimestrielle. Un appariement avec les données du SNDS est projeté pour l'ensemble de la base de données E3N-E4N, mais est en l'attente de l'appui du dispositif France Cohortes.

E3N est la partie française de l'étude de cohorte EPIC (*European Prospective Investigation into Cancer and Nutrition*), constituée sous l'égide de l'Agence internationale de recherche contre le cancer (IARC), qui réunit plus d'un demi-million de participants dans dix pays européens.

Les relations avec d'autres grandes cohortes comme Elfe, Nutrinet ou Constances n'ont pas été programmées et instituées en amont. Il existe néanmoins des collaborations thématiques informelles, par exemple sur les approches nutritionnelles ou environnementales, ou sur les sujets tenant à l'information et à la participation des patients.

Des confrontations ont surtout lieu *ex post, via* les chercheurs travaillant sur des thématiques spécifiques, concernant la complémentarité des études réalisées et la validation des résultats obtenus (par exemple sur des pathologies comme l'asthme ou la maladie de Parkinson).

2.5 Horizon et perspectives envisagés pour le dispositif après 2020

Les perspectives envisagées dans le cadre du PIA pour le dispositif E3N-E4N n'en sont qu'à leur début de réalisation et comprennent, de façon large et ambitieuse :

- la poursuite du suivi des femmes de la cohorte E3N sur des problématiques de recherche élargies au-delà du cancer et évolutives en fonction de l'âge : étiologie et évolution des maladies chroniques, études sur le vieillissement et la dépendance ;
- la constitution d'une cohorte plus complète de pères (25 000 participants souhaités) permettant l'étude de ces maladies sur des sous-populations masculines (par exemple le cancer de la prostate ou la maladie de Parkinson) ;
- la constitution et le suivi effectifs des cohortes d'enfants et de petits enfants prévues (respectivement 50 000 et 20 000 participants souhaités) pour développer l'étude de la santé chez les sujets d'une même famille ayant un terrain génétique et des modes de vie passés en partie partagés ;
- l'élargissement des approches explicatives à la prise en compte des expositions environnementales (à partir de travaux sur les lieux géographiques de résidence, de scolarisation et de travail des participants) ou du microbiome et l'utilisation de l'intelligence artificielle pour l'identification de biomarqueurs.
- la constitution d'une biobanque élargie à des échantillons de sang, d'urine, de fèces, d'ongles et de cheveux pour un sous-échantillon de 20 à 50 000 participants donnant lieu à un phénotypage approfondi, et qui représente à elle seule plus de la moitié des demandes budgétaires formulées par la cohorte sur la période 2020-2024 (*cf. infra*).

3 Principaux résultats et retombées en termes de recherche et de santé publique

Les projets de recherche conduits sont nombreux et diversifiés. Ils sont menés à la fois par l'équipe « Exposome, hérédité, cancer et santé » du Centre de recherche en épidémiologie et santé des populations (CESP), qui assure le support de la cohorte, par d'autres équipes du même centre de recherche, ou enfin par des équipes extérieures, le plus souvent dans le cadre de collaborations. Des projets sont par exemple conduits :

- en matière de cancers, sur les liens entre facteurs génétiques et cancers de la thyroïde, entre traitements hormonaux et mélanome, entre allergies et cancer colorectal ou sur le cancer de la prostate ;
- concernant d'autres pathologies comme la maladie de Parkinson, l'asthme, le diabète de type 2 ou la polyarthrite rhumatoïde ;
- sur les effets des contaminants alimentaires et des perturbateurs endocriniens ou ceux de la pollution de l'air, grâce au système d'information développé pour géolocaliser les participants.

Certains de ces projets sont engagés dans le cadre de collaboration ou de *consortia* internationaux, comme l'EPIC, l'ESCAPE (*European study of cohorts for air pollution effects*) ou le *Lifepath project* (compréhension des effets socio-économiques sur le vieillissement)

Les publications réalisées à partir de la cohorte sont très nombreuses, et compte tenu de son ancienneté, concernent pour l'essentiel la cohorte E3N : plus de 500 publications ont par exemple été dénombrées entre 2011 et 2019, ce nombre atteignant 50 à 70 par an au cours des trois dernières années.

La cohorte donne en outre lieu à une série de communications publiques, que ce soit en direction des participants (*newsletter*, journée prévue en septembre 2020) ou dans le cadre de manifestations comme le *Cohort Innovation Day*.

Les évaluations des jurys scientifiques appelés à évaluer la cohorte soulignent le caractère exceptionnel, au plan national comme international, du dispositif de suivi familial et intergénérationnel ainsi institué. Le jury de 2017 avait toutefois appelé l'attention sur les problèmes de biais de recrutement des conjoints (dont une partie sont décédés) et enfants des femmes appartenant à la cohorte initiale, sur la mesure imparfaite des expositions professionnelles et des inégalités socio-économiques, sur la nécessité de croiser les données issues de la cohorte avec celles des bases médico-administratives et sur l'intérêt de lui adjoindre la collecte d'échantillons sanguins.

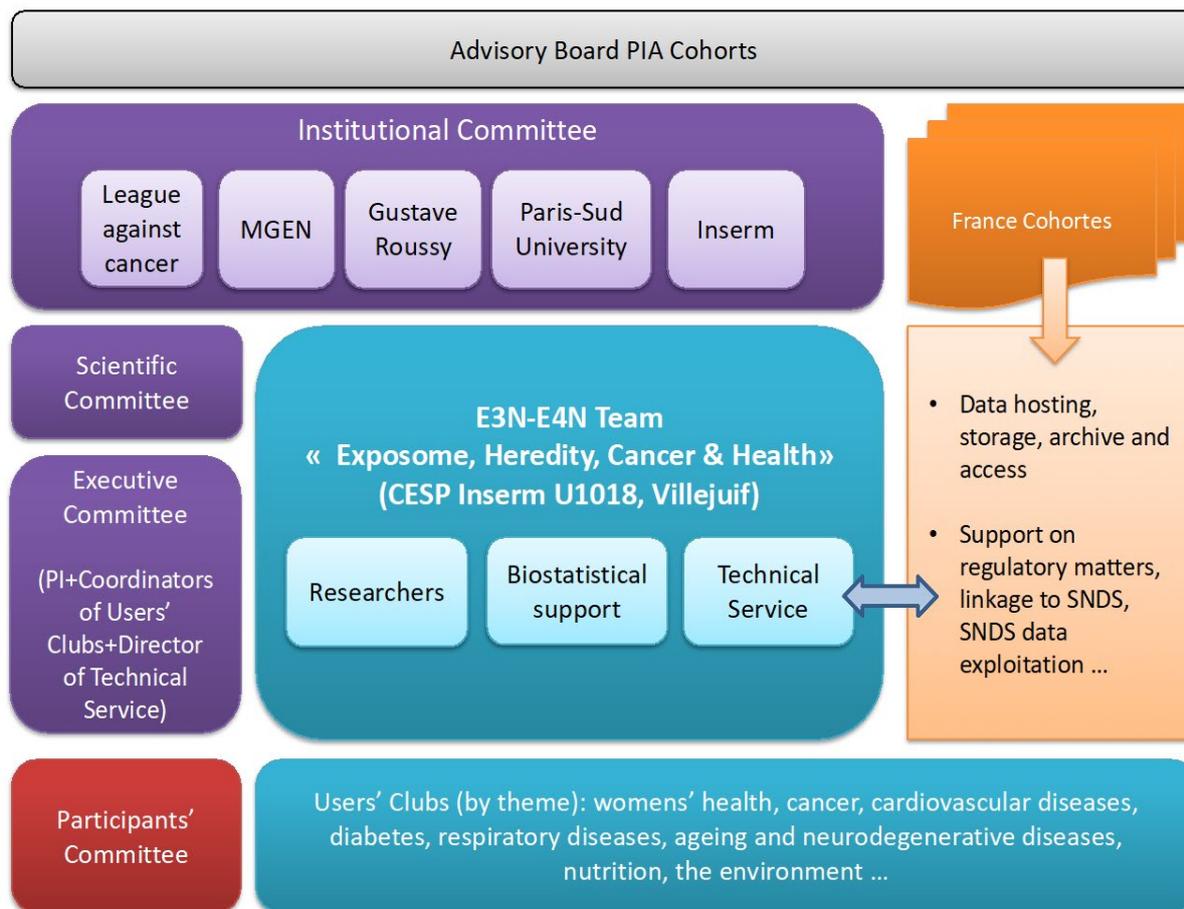
Dans leur notoriété comme dans leurs utilisations, les cohortes E3N-E4N sont avant tout des outils de recherche, avec peu de proximité du ministère de la santé pour les politiques de surveillance sanitaire et de santé publique, mais des apports significatifs de la part de certaines agences sanitaires dans le cadre de leurs politiques de soutien à la recherche (INCA, Anses...).

4 Gouvernance du dispositif

La cohorte E3N résulte d'un partenariat institutionnel historique entre l'Inserm, la MGEN, l'Institut Gustave Roussy (IGR) et la Ligue nationale contre le cancer.

La cohorte E4N a quant à elle donné lieu en 2015 à consortium entre l'Université Paris-Sud, établissement coordinateur, l'Inserm, tutelle du CESP et l'Institut Gustave Roussy.

À la faveur de la réorganisation du CESP et de la création d'une nouvelle équipe, élargie, dirigée par le Pr Gianluca Severi, la gouvernance des deux cohortes sera revue et unifiée en un seul ensemble, doté de comités institutionnel et scientifique uniques, et complété par un comité des participants et des comités d'utilisateurs.



5 Ressources, budgets et financements

5.1 Sur la période 2011-2019

La dotation attribuée à la nouvelle cohorte E4N dans le cadre du PIA a été de 7,5 M€, dont il a été convenu en 2017 que, compte tenu des délais de mise en place des trois sous-cohortes, les reliquats pourraient être utilisés jusqu'à 2025.

Par ailleurs, les deux cohortes E3N et E4N ont reçu entre 2011-2019, hors PIA, des cofinancements de 8,3 M€, dont 3,3 M€ de fonds publics dévolus à la recherche (Inca, Plan cancer, autres agences... ;), 0,84 M€ de financements européens ou internationaux, 3,77 M€ de soutiens de fondations ou organismes faisant appel à la générosité publique et 0,38 M€ de financements industriels.

Les dépenses de personnel ont représenté en 2018 55 % des dépenses annuelles de la cohorte (évaluées à 2,78 M€), et sont pour les deux tiers afférentes au financement de personnels statutaires ou recrutés à durée indéterminée.

L'équipe de la cohorte a à cet égard connu une extension notable, passant de 20 personnes en 2011 à 49 en 2019, dont 10 chercheurs et 20 membres de l'équipe technique.

5.2 Perspectives budgétaires pour la période 2020-2024

Les financements publics supplémentaires demandés par la cohorte au titre des suites du PIA portent sur une nouvelle tranche de 7,5 M€ sur la période 2020-2024, pour un budget prévisionnel total évalué à 24,9 M€, dont 4 M€ dévolus à la maintenance de la cohorte, 12,3 M€ aux frais de personnel et 4,3 M€ à la constitution de la biobanque (*cf. supra*).

Comme suite à l'évaluation finale, le comité de pilotage chargé d'instruire la demande de financement complémentaire a proposé en novembre 2019 un complément de financement du projet à hauteur de 3 M€ jusqu'au 30 juin 2025. Dans ce cadre le ministère chargé de l'enseignement supérieur et de la recherche a proposé de compléter cette enveloppe sur crédits budgétaires.

5.3 Les partenariats industriels

Les partenariats industriels des cohortes E3N-E4N sont demeurés très limités, à l'exception d'un partenariat en cours de conclusion avec la société EPICconcept pour une recherche collaborative autour du traitement des données.

D'après Inserm Transfert qui a conduit une étude exploratoire à ce sujet, son potentiel de valorisation semble encore assez faible aujourd'hui, notamment auprès des laboratoires pharmaceutiques, même s'il pourrait croître à l'avenir.

6 Accès aux données

L'accès aux données des cohortes E3N et E4N est prévu par des chartes similaires et implique :

- l'examen du projet par l'investigateur principal, concernant à la fois sa faisabilité, le risque de chevauchement ou de concurrence avec les projets portés par l'équipe et la vérification d'éventuels problèmes éthiques ;
- l'approbation du comité scientifique de la cohorte ;
- le financement des supports nécessaires à la mise à disposition des données (couverture des coûts directs et contribution aux charges indirectes).

Il se traduit par la signature d'accords qui peuvent être soit de collaboration, soit de transfert des données, seules les équipes académiques et les agences publiques pouvant accéder aux échantillons individuels.

7 Appréciations portées par la mission

La cohorte E3N est un projet emblématique, porté depuis trente ans par la MGEN, l'Inserm et l'Institut Gustave Roussy (IGR) et soutenu par la Ligue contre le cancer. Elle a donné lieu à un ensemble important de projets de recherche débouchant sur de nombreuses publications, et a une forte reconnaissance internationale.

L'intérêt et l'originalité d'une cohorte à plusieurs générations, instituée à travers le projet E4N, sont en outre manifestes et ouvrent des potentialités de recherche prometteuses sur les articulations entre facteurs génétiques, sociaux et environnementaux. Il ne faut par contre en attendre aucune représentativité, dans la perspective des politiques de surveillance et de santé publique.

L'évolution des thématiques de recherche à l'ensemble des maladies chroniques et au vieillissement, au fur et à mesure de l'avancée en âge des participants, change en outre profondément la nature présente et future de la cohorte, qui était précédemment centrée sur les risques de cancer, et dont on peut aujourd'hui questionner la pérennité du positionnement à l'Institut Gustave Roussy.

Les projets spécifiquement financés dans le cadre du PIA, à savoir la constitution des sous-cohortes de pères, d'enfants et de petits enfants des femmes suivies par E3N, ont par ailleurs connu des retards de mise en œuvre importants, et sont encore pour partie, au moins pour le dernier d'entre eux, sujets à questions.

La constitution d'une biobanque permettant le phénotypage approfondi d'un sous-échantillon important de participants impliquera également de résoudre des problèmes de gestion et de stockage importants.

Dans ce nouveau cadre, on peut d'ailleurs regretter que les biais de recrutement et la caractérisation professionnelle et socio-économique des participants, pointés par le jury de 2017, n'aient pas donné lieu à une attention renforcée.

Les moyens mobilisés par la cohorte, notamment en termes de ressources humaines, sont enfin très importants, comparés à ceux d'autres dispositifs du même type, et pourraient encore demander à être accrus à l'avenir avec l'extension du projet et la réalisation des appariements prévus avec le SNDS.

LA PLATEFORME RECONAI : LES COHORTES ELFE ET EPIPAGE 2

1 Présentation générale du dispositif

Le projet « *Recherche sur les Cohortes d'enfants suivis depuis la naissance* » (RECONAI) désigne le regroupement, opéré dans le cadre du deuxième appel à projet « équipement d'excellence », de deux cohortes non retenues dans le cadre de l'appel à projet « Cohortes ». Il s'agit de :

- l'enquête longitudinale française depuis l'enfance, désignée par l'acronyme ELFE ;
- l'étude épidémiologique sur les petits âges gestationnels, EPIPAGE 2.

1.1 Objectifs

1.1.1 La plateforme RECONAI

L'objectif est de mettre en œuvre « une plateforme de recherche » qui « donnera accès à de nombreuses données associées collectées auprès des cohortes ELFE et EPIPAGE2 : données sociales, démographiques, de santé, biologiques et d'imagerie à partir d'enquêtes auprès des familles, des médecins, l'organisation de bilans de santé et de collections biologiques, ou par appareillement sécurisé avec d'autres sources de données. Cet instrument ouvre de nouveaux domaines d'investigation interface entre sciences de la vie et sciences humaines et sociales »²⁰,

Pour la phase 2011-2018, trois objectifs sont retenus :

- le lancement effectif des deux cohortes ELFE et EPIPAGE 2 ;
- le renforcement de la coopération entre les cohortes françaises et internationales ;
- le partage des outils et pratiques en matière de collecte, stockage et accès aux données.²¹

1.1.2 La cohorte ELFE

Le projet ELFE s'inscrit dans l'objectif général des cohortes d'enfants de « *suivi du développement de l'enfant grâce à l'étude du contexte familial, social et économique, et de son état de santé* »²². Retenant une perspective pluridisciplinaire, et non seulement épidémiologique, ELFE se propose « *d'étudier les différents facteurs en interaction tout au long du devenir de l'enfant jusqu'à l'âge adulte (facteurs*

²⁰ Fiche Equipex vague 2, https://www.Ined.fr/fichier/s_rubrique/23/re.co.nai_203075.fr.pdf

²¹ Dossier d'évaluation, 2019.

²² PIRUS C., LERIDON H., *les grandes cohortes d'enfants dans le monde*, Population, 2010, disponible sur <https://www.Ined.fr/fr/publications/editions/population/population-2010-n-4/>

familiaux, sociaux, environnementaux, scolaires, comportementaux, sanitaires, nutritionnels ...) et de comprendre l'impact des situations traversées durant l'enfance ». ²³

Le projet entend répondre à plusieurs questions de recherche - de manière à aider les pouvoirs publics dans « la mise en place de politiques familiales et de santé plus adaptées » :

- *« A quel âge faut-il diversifier l'alimentation ? Quelle influence cela a-t-il sur les préférences alimentaires et la santé ultérieure de l'enfant ?*
- *Quels sont les effets du mode d'accueil du jeune enfant sur ses relations avec les autres enfants, son intégration, à la maternelle et son acquisition du langage ?*
- *Quelle est l'influence de l'usage de l'informatique, du sport ou des activités culturelles sur le développement physique et intellectuel de l'enfant ?*
- *Quel est l'impact des polluants présents dans notre environnement sur la santé et le développement des enfants ?*
- *Quels sont les facteurs familiaux, économiques et socioculturels qui conditionnent la réussite tout au long de la scolarité ?*
- *Les enfants d'aujourd'hui grandissent-ils au même rythme que ceux d'hier ? »* ²⁴

1.1.3 La cohorte EPIPAGE 2

Succédant à une première étude lancée en 2007 sur l'ensemble des enfants nés grands prématurés dans 9 régions de France (EPIPAGE 1), le projet EPIPAGE 2 concerne l'ensemble des régions et retient les objectifs suivants :

- *« Mieux connaître le devenir des enfants grands prématurés et de leurs familles.*
- *Evaluer les effets de l'organisation des soins et des pratiques médicales sur la santé et le développement des enfants prématurés.*
- *Définir les besoins de prise en charge spécifiques pendant l'enfance.*
- *Mieux connaître les parcours des familles.*
- *Améliorer les connaissances sur les causes de la prématurité.*
- *Mieux connaître les conséquences de la prématurité :*
- *En étudiant la valeur pronostique des anomalies cérébrales que peuvent présenter certains enfants sur le développement neurologique et les troubles des fonctions cognitives.*
- *En analysant l'impact des apports nutritionnels et des anomalies de l'installation de la flore intestinale sur la santé et le développement des enfants.*
- *En recherchant des marqueurs biologiques des complications en relation avec la prématurité.*

²³ Note de présentation devant le CNIS, https://www.cnis.fr/wp-content/uploads/2018/03/DPR_2007_STATISTIQUES_SOCIALES_presentation_elfe.pdf

²⁴ Site internet de la cohorte, <https://www.elfe-france.fr/fr/l-etude/objectifs-de-letude/>.

- *Appréhender les questions éthiques posées par la prise en charge des enfants extrêmement prématurés au moment de la naissance et dans les services de soins intensifs »²⁵.*

1.2 Principales caractéristiques

RECONAI est un regroupement de deux cohortes, au départ conçues séparément, qui permet une double mutualisation :

- sur le plan de l'animation scientifique, en comparant les populations respectives d'EPIPAGE 2 (population spécifique) et d'ELFE (population générale), les périodes de collecte étant les mêmes²⁶ ;
- sur le plan de la mutualisation des fonctions de support et de soutien.

2 Caractéristiques scientifiques du projet

S'inspirant des exemples pionniers des cohortes d'enfants britanniques, le projet RECONAI a pour objet de permettre l'étude globale et multidisciplinaire des grands enjeux de santé, de développement et de socialisation de l'enfant, l'accent étant mis sur l'analyse des déterminants précoces des résultats des enfants et le lien entre développement précoce et résultats antérieurs.

2.1 La cohorte ELFE

2.1.1 Date de démarrage du recueil des informations

Recrutement à la naissance à partir de mars 2011.

2.1.2 Nombre de participants prévus et effectifs, modalités d'inclusion et d'échantillonnage

- Prévus : 20 000 enfants
- Réalisés : 18 329

ELFE n'inclut pas de grands prématurés mais est utilisé pour constituer le groupe témoin d'EPIPAGE2.

Le taux d'attrition, de l'ordre de 20 % pour la première année de collecte, est descendu à 7 % de participations en moins pour les enquêtes à 3,5 et 5,5 ans.

L'appariement avec les données médico-administratives permettra un suivi passif des enfants perdus de vue.

²⁵ Site internet de la cohorte, <https://epipage2.inserm.fr/index.php/fr>

²⁶ Par exemple pour comprendre les situations de handicap. Cf. https://www.persee.fr/doc/spira_0994-3722_2016_num_57_1_994

2.1.3 Contenu et mode de recueil des différents types d'informations :

Les informations recueillies comportent à la fois :

- des données recueillies auprès des parents (entretiens téléphoniques, questionnaires postaux, internet) et auprès des enfants à partir de 3,5 ans (visites à domicile) ;
- des prélèvements biologiques pour un sous-échantillon de familles ;
- des mesures environnementales.

	Naissance	2-10mois	1 an	2 ans	3,5 ans	4,5 ans
Enquêtes				X		X (PMI à l'école)
Entretiens tel.			X	X	X	
Entretiens physiques	X	X			X	
Echantillons biologiques	X				X	
Echantillons poussières	X				X	

	5 ans	5,5 ans	7-8 ans
Enquêtes	X (école)		X + questionnaire alimentation
Entretiens tel.		X (+ jeux internet)	
Entretiens physiques			
Echantillons biologiques			
Echantillons poussières			

2.1.4 Stockage des informations

- Stockage des données sur le serveur sécurisé de l'INSERM
- Stockage des prélèvements par l'Établissement français du sang (EFS) dans trois sites distincts, géré par un système d'information spécifique.
- Stockage des échantillons de poussière au laboratoire d'étude et de recherche en environnement et santé de l'école des hautes études en santé publique (EHESP) de Rennes.

2.1.5 Articulation et /ou appariements avec d'autres dispositifs d'information

L'appariement avec le SNDS a été réalisé en 2018.

La transmission des données du SNIIRAM a commencé début 2019 pour les mères et enfants pour lesquels un consentement écrit a été signé.

2.1.6 Horizon et perspectives envisagés pour le dispositif après 2020 :

- Poursuite des opérations communes de collecte (5 opérations prévues entre le printemps 2020 et 2024) qui permettront de mieux comprendre les processus de différenciation en matière d'apprentissages. En 2021-22 se dérouleront une première interrogation directe des enfants à l'âge de 10,5 ans et une première enquête auprès d'adolescents de 14 ans ;

- Développement de projets de recherche reposant sur les appariements avec les données du SNDS, suite aux premiers projets conduits en 2018, le couplage étant étendu aux données des maisons départementales de personnes handicapées ;
- Développement des recherches reposant sur les techniques « omics » utilisant les échantillons collectés en maternité et à l'âge de 3,5 ans (puis les échantillons à 10,5 ans).

2.2 La cohorte EPIPAGE 2

2.2.1 Dates de démarrage et de fin prévue du recueil des informations

- Début : mars 2011.
- Fin : 2023.

Les enfants seront suivis jusqu'à l'âge de 11-12 ans.

2.2.2 Nombre de participants prévus et effectifs, modalités d'inclusion et d'échantillonnage

- L'étude se concentre sur les enfants nés avant 35 semaines d'aménorrhée SA, répartis en trois groupes (très grands, grands et modérément prématurés).
- Effectifs inclus à la naissance (après information des familles) : 7 499 enfants.

2.2.3 Contenu et mode de recueil des différents types d'informations

Les étapes du suivi se font à l'âge de 1, 2, 5 et 8 ans, puis à l'âge de 11-12 ans (entrée au collège).

- à la naissance : informations sur la grossesse, l'accouchement et la prise en charge immédiate de l'enfant collectées en maternité à partir du dossier médical ;
- à l'issue de l'hospitalisation en néonatalogie : bilan complet recueilli à partir du dossier médical ;
- à 1 an (2012-2013) : questionnaire auprès des parents concernant la santé et la prise en charge médicale de l'enfant ainsi que la santé et les conditions de vie des parents. 3 841 participants sur 4 290 éligibles (90 %) ;
- à 2 ans (2013-2014) : questionnaire aux parents et questionnaire au médecin de l'enfant. 3 689 participants sur 4 199 éligibles pour les parents (88 %), 3 599 sur 4 199 pour les médecins (86 %) ;
- à 5,5 ans (septembre 2016- décembre 2017) : examen médical et tests psychologiques. 2 974 participants sur 3 937 éligibles (68 %).

2.2.4 Stockage des informations

- Stockage des données : serveur sécurisé de l'INSERM

- Stockage des prélèvements au Centre d'étude du polymorphisme humain (Fondation Jean Dausset).

2.2.5 Articulation et /ou appariements avec d'autres dispositifs d'information

L'appariement avec le SNDS réalisé en 2018.

La transmission des données du SNIIRAM a débuté début 2019 pour les mères et enfants pour lesquels un consentement écrit a été signé.

2.2.6 Horizon et perspectives envisagés pour le dispositif après 2020

Une enquête auprès des enseignants est envisagée lorsque les enfants seront en CE2. Une étape de suivi à l'âge de 11 ans devrait également être mise en place.

Le suivi devait initialement s'arrêter à l'âge de 11-12 ans. Cependant, si le nombre de familles participantes est suffisant et sous condition de l'obtention d'un financement complémentaire, un dernier recueil de données auprès de ces enfants devenus adolescents serait souhaitable.

3 Principaux résultats et retombées en termes de recherche et de santé publique

3.1 Résultats et apports scientifiques jugés les plus importants par les responsables des cohortes

3.1.1 Pour la cohorte ELFE

Trois principaux résultats sont mis en avant par le document d'évaluation :

- confirmation de la faible prévalence de la France du début et de la durée de l'allaitement maternel (26 % des nourrissons français passent à l'alimentation solide à l'âge de 4 mois, âge recommandé par l'OMS) ;
- meilleure connaissance des surexpositions, d'une part aux risques pyréthroides des femmes françaises enceintes par rapport aux autres pays (risque accru de malformations congénitales), d'autre part aux risques professionnels (retard de croissance fœtale documenté chez 5 % des femmes ayant des expositions élevées aux nanoparticules) ;
- meilleure connaissance du développement des compétences langagières : rôle des interactions parents / enfants, poids des différences socio-économiques, rôle des modes de garde.

3.1.2 Pour la cohorte EPIPAGE 2

Quatre résultats sont particulièrement mis en avant par le document d'évaluation :

- meilleure connaissance des risques de troubles du développement pour les enfants modérément prématurés (plus d'un tiers présente un risque), alors qu'aucun suivi spécialisé (réservé aux enfants très prématurés) ne leur est proposé ;

- nouvelles recommandations émises en 2019 concernant les soins dispensés aux enfants très prématurés pour diminuer un taux de survie plus faible en France que dans d'autres pays européens ;
- évaluation de l'impact des pratiques médicales (dépistage par échocardiographie avant le jour 3 de la vie ; traitement hypotenseur au cours de trois premiers jours de vie) sur la mortalité ;
- meilleure appréhension des causes d'accouchement prématuré sur le risque de décès, en plus du rôle de l'âge gestationnel à la naissance.

3.2 Activité en termes de publications, de communication et de diffusion

2011-juin 2019	Nb d'articles	Dont articles en anglais	Dont revues de rang A
ELFE	64	43	15
EPIPAGE 2	48	48	24
TOTAL	112	91	39

Les articles en anglais représentent plus de 80 % des publications. Plus du tiers ont été publiés dans le *top 10* des meilleures revues spécialisées.

3.3 Projets de recherche en cours à partir des cohortes

2011- juin 2019	Nb de projets	Laboratoires de recherche concernés
ELFE	99	26
EPIPAGE 2	87	27
TOTAL	186	53

3.4 Insertion dans des coopérations ou des *consortia* européens ou internationaux

3.4.1 Pour la cohorte ELFE

Insertion dans le projet européen LIFECYCLE, visant la mise en place d'un réseau européen de cohortes d'enfants.

3.4.2 Pour la cohorte EPIPAGE 2

Participations à deux grands projets européens :

- projet européen EPICE (*Effective Perinatal Intensive Care in Europe*) conduit dans 17 régions d'Europe²⁷ ;
- projet RECAP (*Research on European Children and Adults Born Preterm*²⁸) lancé en 2016 dans le cadre de H 2020.

²⁷ <https://www.epiceproject.eu/en/cohort/epice-project.html>

²⁸ <https://recap-preterm.eu/>

3.5 Appréciations portées par les jurys internationaux d'évaluation

Avis positif du jury à mi-parcours, reposant sur la reconnaissance de la très grande qualité méthodologique (représentativité statistique, qualité et fiabilité du système de données), scientifique (intégration étroite des sciences sociales) et de la gouvernance de la plateforme.

L'unique recommandation du jury de mener des études pilotes sur la viabilité et la qualité des échantillons biologiques

4 Gouvernance du dispositif

Définie comme une plateforme commune aux deux cohortes, RECONAI fonctionne comme une instance de coordination entre les deux équipes d'ELFE et d'EPIPAGE 2, afin de mutualiser les outils et processus de collecte, de stockage et de libre accès aux données.

La gouvernance s'appuie sur :

- un comité stratégique : validation des orientations stratégiques et budgétaires dans le cadre de réunions annuelles ;
- un comité de pilotage : gestion des aspects techniques, administratifs et scientifiques communs aux deux cohortes dans le cadre de réunions mensuelles ;
- un comité scientifique : réunions annuelles.

La cohorte ELFE est portée par une unité mixte de service conjointe à l'INED, à l'INSERM et à l'EFS.

La cohorte EPIPAGE 2 a le statut d'équipe (équipe de recherche en épidémiologie obstétrique, périnatale et pédiatrique – EPOPEE) au sein du CRESS, laboratoire de recherche commun à l'INSERM et à l'université de Paris.

Les deux cohortes associent régulièrement les familles à l'avancée des travaux, *via* une communication dédiée (*newsletter*, « espace famille » sur le site de ELFE, « coté parents » sur le site d'EPIPAGE 2).

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

Les recettes du projet ce sont globalement élevées sur la période 2012-2019 à 27,20 M€ dont 13,01 M€ au titre du PIA. Les dépenses au titre du seul projet ELFE avaient atteint à l'été 2019 15,22 M€ ; les dépenses pour EPIPAGE 2 avaient atteint 6,02 M€.

En 2019, le projet RECONAI mobilise 17,6 ETP (13,7 pour ELFE et 3,9 pour EPIPAGE 2) ; 10 ETP correspondent à des emplois permanents (personnels statutaires et contrats à durée indéterminée) et 7,6 à des contrats à durée déterminée, principalement positionnés sur des fonctions de data manager et de statisticien.

Outre le financement au titre du PIA (13,01 M€), le projet RECONAI a obtenu de 2012 à 2018 les très importants financements publics suivants (en M€) :

Financements du ministère chargé de la recherche (dont TGIR)	Autres financements publics	Fondation de France	Appels à projet gérés par l'ANR	Financements européens (H2020 notamment)
10,06	2,94	0,15	0,30	0,35

5.2 Perspectives budgétaires pour la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale porte une demande de financement complémentaire de 2,70 M€. Le budget prévisionnel global cumulé du projet sur la période 2020-2024 s'élève à 11,89 M€ (dont 2,64 M€ pour l'étude des 10,5 ans), en intégrant le financement des postes supporté par les organismes parties à l'accord de consortium (4,83 M€), ainsi que 2,41 M€ de produits constatés d'avance au 1^{er} janvier 2020.

Au titre de la période 2020 – 2024, 3,5 ETP supplémentaires sont demandés dont 1,5 data manager en priorité.

Comme suite à l'évaluation finale le comité de pilotage chargé d'instruire la demande de financement complémentaire a proposé en novembre 2019 de compléter le financement du projet jusqu'au 30 juin 2025 à hauteur de la demande des porteurs (2,70 M€).

5.3 Les partenariats industriels

La part des financements correspondant à des partenariats public / privé (0,38 M€ pour l'entreprise ABBVIE) reste marginale par rapport aux financements publics.

6 Accès aux données

Une politique de libre accès aux données est définie pour les chercheurs, sur la base d'un projet de recherche et dans le respect des conditions fixées par une charte.

Le conseil d'administration de l'Ined s'est prononcé contre la facturation aux équipes de chercheurs de tout ou partie des coûts de l'infrastructure.

Les demandes d'accès aux données se font par une plateforme sécurisée Internet, puis sont examinées par des comités propres à chacune des cohortes.

Un temps d'exclusivité de 18 mois est consenti aux équipes contribuant à l'enquête.

Pour les équipes privées, l'accès est conditionné à l'existence d'un projet commun avec une équipe de chercheurs publics.

Le Comité éthique de l'Ined est sollicité en tant que de besoin.

7 Appréciations portées par la mission

Les deux cohortes permettent de répondre à l'intérêt croissant sur les origines développementales des maladies (*Developmental Origins of Health*) permettant de mettre en relation les facteurs épigénétiques et environnementaux²⁹.

Il s'agit à cet égard du seul dispositif français de suivi des enfants à la naissance, qui a été institué dans le cadre d'un partenariat institutionnel très large et assis sur un organisme de recherche reconnu (Ined).

Ce dispositif est considéré, dans les rapports du Haut Conseil de la famille, de l'enfance et de l'âge (HCFEA) et du Haut conseil de la santé publique (HCSP) comme une pièce maîtresse du suivi statistique de la santé des enfants³⁰, de par l'accent simultané mis sur le développement physique et le bien-être de l'enfant et l'association explicite des problématiques sanitaires, sociales et environnementales.

Une bonne intégration a à cet égard été mise en œuvre entre la cohorte maîtresse et la cohorte d'enfants prématurés, renforçant l'intérêt de la poursuite du dispositif pendant les années d'adolescence des enfants suivis.

Le prolongement des financements devrait permettre, outre la prolongation de ce suivi, le passage à la phase « *omics* » envisagée par l'équipe, mais non encore réalisée, et que rendent envisageable les tests de qualité effectués. De même, l'élargissement des appariements avec les données fiscales-sociales de l'Insee et celles de la CNAF (suggérée par le HCFEA) permettrait d'accroître le potentiel du projet.

L'absence de perspectives significatives de financements privés rend en outre nécessaire que les financements publics continuent à jouer un rôle majeur.

La plateforme, fortement soutenue par les tutelles Inserm et Ined, s'inscrit parfaitement dans les évolutions en cours marquées par le développement des laboratoires de recherche en épidémiologie (CRESS) et le démarrage du projet France Cohortes.

²⁹ <https://hal.archives-ouvertes.fr/hal-01004446>

³⁰ Conseil de l'enfance et de l'adolescence, *Des données et des études publiques mieux centrées sur les enfants*, juin 2019 et Haut Conseil de la santé publique, *Le dispositif statistique de surveillance et de recherche sur la santé de l'enfant*, octobre 2019

LA COHORTE CANTO

1 Présentation générale du dispositif

Le cancer du sein touche chaque année plus de 50 000 femmes en France. L'amélioration des traitements permet à 80 % d'entre elles de vivre plus de 10 ans après leur cancer, mais souvent au prix d'une qualité de vie dégradée en raison de la toxicité des traitements administrés. 30 000 patientes souffriraient ainsi d'une altération de leur qualité de vie (fatigue, difficulté d'insertion professionnelle, baisse du niveau social, ...), liée aux toxicités chroniques dont leur étude est l'objet de la cohorte CANTO (CANcer TOxicities).

1.1 Objectifs

Le projet a pour objet de :

- créer une base de données cliniques et biologiques prospective sur les toxicités chroniques permettant de les décrire (incidence, présentation clinique, caractéristiques biologiques) et d'apprécier leur impact social ;
- développer des prédicteurs moléculaires de l'apparition des toxicités chroniques ;

Sur le plan scientifique, la cohorte met en avant deux objectifs :

- identifier les toxicités à moyen et long terme des traitements anticancéreux ;
- évaluer leur impact grâce à des marqueurs moléculaires.

1.2 Principales caractéristiques de la cohorte

Portée par un consortium réunissant la fondation de coopération scientifique CANTO Cancer, l'institut Gustave Roussy (IGR), le centre Léon Bérard (CLB), le centre George François Leclerc (CGFL) et le centre François Baclesse (CFB), CANTO est une cohorte prospective portant sur l'étude de la survie après un cancer du sein. Il s'agit de la plus grande étude au monde sur le sujet.

Par ailleurs, il s'agit de structurer une communauté de recherche dédiée à la question de la survie après un cancer (120 professionnels environ).

2 Caractéristiques scientifiques du projet

2.1 Dates de démarrage, nombre de participants

Le recueil des informations a débuté en février 2012 et s'est achevé en décembre 2018. Ont été incluses dans la cohorte 12 000 patientes atteintes d'un cancer du sein de stade I-III inscrites dans 26 centres de recrutement sur le territoire métropolitain. Les informations recueillies sont :

- des données relatives à la qualité de vie, aux trajectoires personnelles, à l'état psychologique, collectées par questionnaires (40 questionnaires au total par patiente) ;
- des données de radiothérapie pour 4 000 patientes dans 12 centres, collectées de manière automatique et anonyme par logiciel ;
- 107 925 échantillons issus de prélèvements biologiques se décomposant ainsi : 25 283 échantillons de sang (permettant l'extraction de l'ADN chez les 8 475 premières patientes) ; 41 357 échantillons de sérum et 41 285 échantillons de plasma ;
- des échantillons de tumeurs sont prélevés sur 3 000 patientes présentant un risque élevé.

2.2 Contenu, collecte et stockage des différents types d'informations

Les données sont collectées par des professionnels experts en acquisition de données et leur qualité fait l'objet d'un plan de suivi. Elles sont hébergées par le centre Georges François Leclerc de Dijon (centre qui tient le registre cancer de la Côte d'or dédié au cancer du sein). Les échantillons sont stockés dans les biobanques de l'institut Gustave Roussy et du centre Léon Bérard, partenaires du projet.

L'appariement avec les données du SNDS n'est pas encore effectif. Par ailleurs, CANTO n'a pas été retenu parmi les projets pilotes du *Health Data Hub*.

La gestion des données devrait évoluer sur la base d'une nouvelle répartition des tâches entre Unicancer, qui centraliserait (notamment pour mieux répondre aux exigences du RGPD) la base de données de la cohorte, et le centre de Dijon, qui superviserait l'ensemble des analyses statistiques.

3 Principaux résultats et retombées en termes de recherche et de santé publique

Parmi les résultats mis en avant par les responsables de la cohorte, on relève :

- l'analyse différentielle de l'impact de la chimiothérapie et de la thérapie endocrinienne sur la dégradation de la qualité de vie deux ans après le diagnostic montrant que c'est l'hormonothérapie, et non la chimiothérapie, qui est associée à une dégradation de l'état de santé global des femmes post-ménopausées ;
- la mesure de l'impact de la prise de tamoxifène (traitement le plus vendu) montrant que la non-observation du traitement diminue le risque de survie sans maladie à distance ;
- la définition de nouvelles méthodologies en matière d'analyse moléculaire permettant le profilage des échantillons à l'aide d'approches à haut débit dans le but de générer des prédicteurs multimarqueurs ou encore le recours à l'intelligence artificielle pour identifier les polymorphismes nucléotidiques (SNP) associés à une fatigue ou une obésité accrue ;
- la structuration d'une communauté de recherche (environ 120 chercheurs) autour de la survie du cancer à partir de l'identification préalable de six domaines de recherche (impact des traitements systémiques, hétérogénéité des soins, neuropsychologie, impact social, impact de la radiothérapie, génomique). Sept groupes de recherche interdisciplinaires travaillent sur chaque toxicité.

Sur un plan quantitatif, CANTO a permis :

- 37 projets de recherche assortis de demandes d'accès aux données (sur 42 demandes) ;
- 18 demandes d'informations de la part de médecins ;
- 5 partenariats avec d'autres projets PIA (le Labex GEMNED, les RHU My Probe, Lumière et ONCOBIOTA, et l'IHU MMO)
- 9 partenariats internationaux.

Le jury international a souligné l'excellence d'une cohorte unique au monde, et dont les développements à venir devraient encore accroître l'intérêt scientifique :

- analyse des toxicités liées à la radiothérapie ;
- identification de nouvelles toxicités (gynécologiques, osseuses, cardiaques ...) ;
- développement des analyses protéomiques et métabolomiques et intégration dans une base de données moléculaires unique permettant le recours à l'analyse mathématique pour générer des prédicteurs moléculaires.

Suite à l'évaluation à mi-parcours, les responsables de CANTO ont rédigé un article de présentation de la cohorte (sous presse) pour accroître sa visibilité et renforcé leur stratégie de publication, centrée sur les revues de haut niveau (9 articles en cours). Ont également été mise en place la biobanque tumorale et précisée l'utilisation des prélèvements sanguins. Dans son évaluation finale, le jury a néanmoins réitéré ses questionnements sur la perspective de prélèvements de nouveaux échantillons et le moment de réaliser un second prélèvement sanguin.

4 Gouvernance du dispositif

CANTO dispose d'une gouvernance simple reposant principalement sur :

- une instance décisionnelle : le comité exécutif, comprenant le chef de projet, le coordinateur scientifique et les représentants de chaque membre du consortium. Le CE se réunit fois par an et prend l'ensemble des décisions stratégiques ;
- une instance consultative : le conseil scientifique consultatif (SAB), qui se réunit une fois par an.

Des comités de pilotage interdisciplinaires pour chacun des six domaines de recherche identifiés prennent en charge la collecte des données, l'examen des projets de recherche

Le support et le soutien à l'activité de la cohorte (aspects réglementaires, éthiques et financiers) sont fournis par Unicancer.

5 Ressources, budgets et financements

5.1 Sur la période 2011-2019

Selon les données communiquées par les porteurs, les recettes du projet CANTO se sont globalement élevées sur la période 2011-2018 à 15,27 M€, dont 10,25 M€ au titre du PIA (le financement total au titre du PIA s'élève à 13,10 M€, les fonds correspondants étant utilisables jusqu'au 31 décembre 2020). Les dépenses au titre du projet CANTO avaient atteint à fin 2018 15,17 M€, dont 9,05 M€ au titre des dépenses de personnel.

Outre le financement au titre du PIA (13,10 M€), le projet CANTO a obtenu de 2012 à 2018 les principaux financements publics suivants (en M€) :

Financements ministériels et d'établissements publics (dont INSERM)	CNAF	Appels à projet gérés par l'ANR	Financements européens
4,71	0,25	0,32	0,28

En 2019, le projet s'appuie sur une équipe de 16 membres qui, sauf exception, contribuent au projet pour une partie seulement de leur temps de travail ; pour l'essentiel, il s'agit d'agents en contrats à durée indéterminée de droit privé, salariés d'Unicancer ou d'employeurs hospitaliers. Sur la période 2011-2018, le projet CANTO a mobilisé environ 25 ETP financés à hauteur d'environ 75 % par le PIA et pour le reste par les membres du consortium.

5.2 Perspectives budgétaires sur la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale portait une demande de financement complémentaire de 4,76 M€. Le projet de budget prévisionnel global cumulé sur la période 2020-2024 s'élevait à 14,96 M€.

Des renforts en statisticiens étaient demandés au titre de l'analyse des données résultant de l'appariement du SNDS.

Comme suite à l'évaluation finale le comité de pilotage chargé d'instruire la demande de financement complémentaire a proposé en novembre 2019 un complément de financement du projet à hauteur de 3,9 M€ jusqu'au 30 juin 2025.

5.3 Les partenariats industriels

La part des financements correspondant à des partenariats public / privé (0,59 M€, dont 0,39 M€ pour l'entreprise ABBVIE) est relativement marginale par rapport aux financements publics.

6 Accès aux données

Les données sont en accès libre pour les chercheurs sous réserve du paiement d'une contribution financière et d'un avis favorable du CE sur proposition du comité de pilotage concerné, sur la base des critères suivants : excellence scientifique, faisabilité, plan statistique, cohérence avec la stratégie scientifique de CANTO. Une fois accepté, le projet fait l'objet d'un contrat incluant les clauses de

sécurité en matière de données. Le CGFL de Dijon prend en charge le codage et la diffusion des données et assure le plan d'analyse statistique.

7 Appréciations portées par la mission

Les problèmes liés à la toxicité des traitements, qui surviennent parfois au-delà de cinq ans après le cancer, et qui sont plus fréquents chez les femmes jeunes (sous représentées dans la cohorte initiale), l'apparition de nouveaux médicaments et de nouveaux traitements et l'apparition de nouvelles sources de données (fèces) montrent tout l'intérêt qu'il y a à prolonger la cohorte et à l'enrichir d'un sous-groupe de femmes de moins de quarante ans présentant un cancer du sein précoce.

Si le recours aux technologies « *omics* » constituent l'un des points forts de CANTO, de même que ses partenariats de recherche avec d'autres infrastructures du PIA (Labex, IHU et RHU), ses préoccupations visent aussi la prise en compte de dimensions diverses, telles la fatigue, la dépression, la réinsertion professionnelle, les aspects médico-économiques, qui sont des aspects importants pour les politiques de santé publique.

CANTO bénéficie dans ce cadre d'un partenariat fort associant la fédération Unicancer, l'INCA et la Ligue contre le cancer, et de l'appui direct, à la fois financier, administratif et opérationnel d'Unicancer, qui lui permet par exemple le recrutement de personnels sous statut permanent.

Une attention doit être néanmoins portée à la qualité statistique de la cohorte (par l'évaluation des biais de sélection liés aux modalités de recrutement dans les centres de lutte contre le cancer) et à l'appariement avec les données du SNDS, encore non réalisé à ce jour.

Il serait à cet égard important que la cohorte aille au bout de ses propres potentialités et objectifs, sans forcément chercher, comme cela a pu être évoqué, à déboucher à terme sur une plateforme plus large, susceptible d'être redondante avec d'autres suivant des patients atteints de cancers.

LA COHORTE CKD REIN

1 Présentation générale du dispositif

Internationalement définie, depuis 2002, par la présence d'anomalies de structure ou de fonction du rein, et cela pendant plus de trois mois au moins, la maladie rénale chronique (MRC) affecte environ 10 % de la population adulte mondiale et plus de 30 % des personnes âgées de plus de 70ans. Le poids de l'insuffisance rénale terminale (IRT), requérant la dialyse ou une greffe rénale, ne cesse d'augmenter. Il s'agit là désormais d'un enjeu de santé publique important, dont l'impact financier est, de surcroît, substantiel. Or, prévenir l'IRT nécessite de mieux connaître la MRC avant le stade terminal.

1.1 Principales caractéristiques de la cohorte

L'étude *Chronic Kidney Disease-Réseau Epidémiologie et Information en Néphrologie*, ou CKD REIN, a pour but de mettre en place une plateforme de recherche pour répondre à des questions clés concernant les déterminants et les conséquences de la MRC et identifier les meilleures prises en charge. Elle est coordonnée par l'Université Paris-Sud et soutenue par l'Inserm. Y participent 4 CHU (Lyon, Bordeaux, Amiens et Nancy), 2 agences sanitaires (ABM et EFS), 2 unités mixtes de recherche INSERM-Universités, la Biobanque de Picardie, le Centre national de génotypage (CNG) et une institution américaine (*Arbor research collaborative for health*).

Cette cohorte inclut le volet français de l'étude internationale CKDopps (*Chronic Kidney Disease Outcomes and Practice Patterns Study*) ciblant la MRC avancée.

1.2 Objectifs

La cohorte CKD REIN a pour objectif général de développer une plateforme de recherche destinée à répondre à des questions-clés concernant les déterminants et les biomarqueurs associés au pronostic de la MRC, ainsi que d'évaluer l'efficacité de sa prise en charge. Les objectifs spécifiques de l'étude, dans le cadre nosologique de la MRC (stades 3 à 5, définis selon le niveau du débit de filtration glomérulaire -DFG), sont :

- étudier un ensemble de facteurs psycho-sociaux, environnementaux, biologiques, génétiques et leurs interactions en lien avec plusieurs événements rénaux et non rénaux ;
- identifier de nouveaux biomarqueurs prédictifs de la progression et des complications de la MRC ;
- identifier les pratiques et les modes d'organisation des soins associés à un meilleur pronostic, une meilleure prise en charge, moins d'hospitalisations, et une meilleure qualité de vie à l'échelon national et international ;
- quantifier le coût des pratiques de soins et les rapporter à leur efficacité, à l'échelon national et international ;
- évaluer la prévalence de la MRC prise en charge en néphrologie.

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

L'étude CKD REIN est une cohorte prospective de patients. Elle a débuté par un recensement de l'ensemble des patients éligibles vus en consultation entre le 1^{er} janvier et le 31 décembre 2012 dans chaque site. Les informations personnelles des patients, les dates de consultations et le nom du néphrologue sont obtenus à partir des fichiers de gestion des consultations ; le diagnostic de la MRC et le niveau de DFG à partir des dossiers médicaux. Ce recensement a servi de base à la sélection des patients, et à l'estimation de la prévalence de la MRC, ainsi que de la mortalité et du risque d'IRT pour l'ensemble des patients éligibles.

2.2 Modalités d'inclusion et d'échantillonnage

La période d'inclusion s'est étendue du juillet 2013 à mars 2016. L'étude a inclut 3033 patients atteints de MRC, aux stades 3 et 4-5.

Une sélection de 40 sites a été effectuée à partir de la liste exhaustive des 241 consultations externes de néphrologie française ouvertes aux patients MRC ni dialysés, ni transplantés, en respectant leur répartition selon la région et le secteur d'activité.

Les critères d'inclusion ciblent des patients adultes ayant une MRC diagnostiquée (au moins deux mesures de débit de filtration glomérulaire (DFG) < 60 mL/min), ni dialysés, ni transplantés. Sont au contraire exclus les patients âgés de moins de 18 ans, dans l'Incapacité ou refusant de participer, non-résidents en France ou ayant un déménagement prévu.

L'inclusion s'effectue à l'occasion d'une consultation de routine en néphrologie, après recueil du consentement ; les patients sont suivis une fois par an pendant au moins 5 ans. Le recueil des données repose sur plusieurs questionnaires remplis par les ARC, les patients, les médecins et les chefs de service, à l'entrée dans l'étude et durant le suivi.

2.3 Contenu, collecte et stockage des différents types d'informations

Pour tous les patients éligibles recensés, sont recueillis : nom(s) et prénom(s) ; adresse de résidence ; date, commune et pays de naissance. Ce recueil a pour but d'estimer la prévalence de la MRC, de suivre passivement l'ensemble de ces patients pour la mortalité et l'incidence de l'IRT, et d'évaluer la représentativité de l'échantillon des inclus dans la cohorte du point de vue de l'âge, du sexe, de la catégorie sociale évaluée par les données des Ilots Regroupés pour l'Information Statistique (IRIS) de l'INSEE et du suivi. La récupération des données des IRIS nécessite le recueil de l'adresse détaillée des patients, mais celles-ci ne seront pas conservées pour les éligibles non inclus. Le recueil de la date, de la commune et du pays de naissance en sus du nom patronymique et des prénoms est justifié pour le suivi de l'IRT et du statut vital à travers les registres nationaux de santé (REIN) et le RNIPP.

L'origine géographique des parents a aussi été demandée aux patients inclus dans la cohorte, cette donnée étant essentielle à l'estimation du DFG par équation. Concernant les échantillons biologiques potentiellement identifiants, de l'ADN est conservé de façon anonymisée au Centre d'études du polymorphisme humain (CEPH) pour tous les patients ayant donné leur consentement.

Au cours du suivi, les évènements rénaux et non rénaux sont activement recherchés et enregistrés, ainsi que la perception de leur état de santé par les patients, les hospitalisations et leurs causes, et la

consommation de soins. Des mesures anthropométriques, de pression artérielle et de fréquence cardiaque seront réalisées à l'inclusion et aux suivis, comme l'indique le tableau suivant :

Type de mesures	Paramètres mesurés
Anthropométrie	Poids, taille, indice de masse corporelle Tour de taille, tour de hanche
Cardiovasculaire	Fréquence cardiaque, PA* au repos ; Auto-mesure de la PA*
Biologie_Sang	Créatinine (méthode enzymatique) Ionogramme : Na, K, Cl, HCO3- Albumine (pré-albumine) CRP ultrasensible Calcium, Phosphore, PTH, 25 OH vitamine D Phosphatase alcaline Hémogramme, réticulocytes et % de GR hypochromes (à spécifier) Ferritine, fer sérique, transferrine, coefficient de saturation de la transferrine Lipides à jeun: Cholestérol Total, HDL-chol, LDL-chol, Triglycérides Glycémie à jeun, HbA1c Uricémie Ag Hbs
Biologie_Urine	Na, urée (échantillon) Protéinurie, albuminurie, créatininurie (échantillon) Hématurie, leucocyturie

*PA : pression artérielle

De plus, chez tous les participants à l'entrée dans l'étude et lors du suivi à 5 ans, et chez un patient sur trois à 1 an et à 3 ans, les examens biologiques de base recommandés en routine par la Haute Autorité de Santé pour le suivi des patients MRC sont réalisés de façon systématique dans un laboratoire centralisé. Les résultats de ces examens sont transmis aux néphrologues ainsi qu'au responsable scientifique de l'étude. Des échantillons de sang et d'urine sont prélevés aux mêmes temps et à l'inclusion sont aussi prélevés des échantillons à destination du CEPH pour le Centre National de Génotypage (CNG) à des fins d'études génétiques.

L'ensemble des échantillons de la cohorte CKD REIN devait initialement être stocké à l'Établissement français du sang (EFS), mais le changement de politique de ce dernier a conduit à maintenir le stockage des échantillons au niveau décentralisé des centres participants, sous la supervision pour moitié de l'EFS et pour moitié de la Biobanque de Picardie. Seules les extractions d'ADN sont centralisées auprès de cette dernière.

La base de données de CKD REIN est aujourd'hui administrée par l'Agence de la Biomédecine, qui gère également le registre REIN recensant les malades atteints de pathologies rénales chroniques au stade terminal.

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

Le consortium CKD REIN bénéficie d'un partenariat avec trois plateformes technologiques :

- la Biobanque de Picardie, CRB du CHU d'Amiens qui est l'une des premières biobanques d'Europe;
- le centre national du génotypage (CNG) d'Évry ;
- la plateforme technique de l'Institut des maladies cardiovasculaires et métaboliques (I2MC) de Toulouse.

L'appariement envisagé avec les données du SNDS a pris du retard et n'avait pas fin 2019 été encore effectué, bien que l'autorisation obtenue permette d'envisager qu'il soit réalisé prochainement.

3 Principaux résultats et retombées en termes de recherche et de santé publique

CKD REIN a fourni pour la première fois un aperçu des pratiques néphrologiques actuelles en France, sur la base d'un échantillon représentatif de consultations de néphrologie. L'étude a mis en évidence le profil à risque très élevé des patients atteints de maladie rénale chronique, avec une prévalence élevée de l'HTA (91%) et des maladies cardiovasculaires (53%) (cf. tableau), un mauvais contrôle de la pression artérielle, un grand nombre de médicaments prescrits souvent de façon inappropriée, une fréquence élevée d'insuffisance rénale aiguë sur chronique, et un fort impact de la maladie sur la qualité de vie.

Caractéristiques des patients de la cohorte CKD REIN

N=3033	Médiane (IQR) ou moyenne±sd ou %
Age, années	67±13
Hommes	65%
DFG en ml/min/1.73 m ²	32 (23-41)
Diabète	41%
Obésité	35%
Hypertension	91%
Pression artérielle ≥140/90 mmHg	57%
Antécédents cardiovasculaires	53%
Antécédent d'insuffisance rénale aiguë.	22%
Albuminurie ou proteinurie	
Modérée	32%
Sévère	41%

Source : *Stengel et al. Nephrol Dial Transplant 2018;34:277-86*

L'étude CKD REIN a favorisé des relations plus étroites entre les équipes de recherche fondamentale, clinique et en population, au sein et au-delà de son propre consortium.

Ce consortium a contribué à la mise en place du projet PROTEOMARK, soutenu par un financement de l'ANR (2018) ; l'objectif est de valider et de développer des profils moléculaires prédisant les complications cardiovasculaires et la progression de l'IRC. Il a également contribué à la mise en place du projet CKD-MICROBIOTA, avec la participation de deux équipes académiques externes (le CHRU de Nancy et l'Institut National de Recherche Agronomique (INRA), également financé par l'ANR³¹.

CKD REIN permet aussi d'évaluer la mise en œuvre des recommandations actuelles concernant la prise en charge de la MRC et de leur impact sur le pronostic des patients. Les comparaisons internationales à travers l'étude CKDopps favorisent également l'identification des meilleures pratiques de néphrologie grâce à la grande variété de pratiques cliniques dans le monde.

En ce qui concerne la politique de santé publique dans le domaine des maladies chroniques, CKD REIN a apporté des preuves de la nécessité d'inclure les consultations de diététique et l'éducation des patients sur les options de traitement de la défaillance rénale dans le forfait à mettre en œuvre d'ici la fin de 2019 pour le suivi des patients de consultation atteints de maladie rénale chronique avancée. Ces mesures visant à prévenir la progression et les complications de la maladie rénale chronique devraient réduire le poids des traitements de suppléance rénale et améliorer la transition vers l'insuffisance rénale terminale.

La cohorte a également montré un taux d'hospitalisation élevé inattendu. Des analyses sont en cours pour identifier les facteurs potentiellement modifiables qui pourraient réduire le poids de ces hospitalisations.

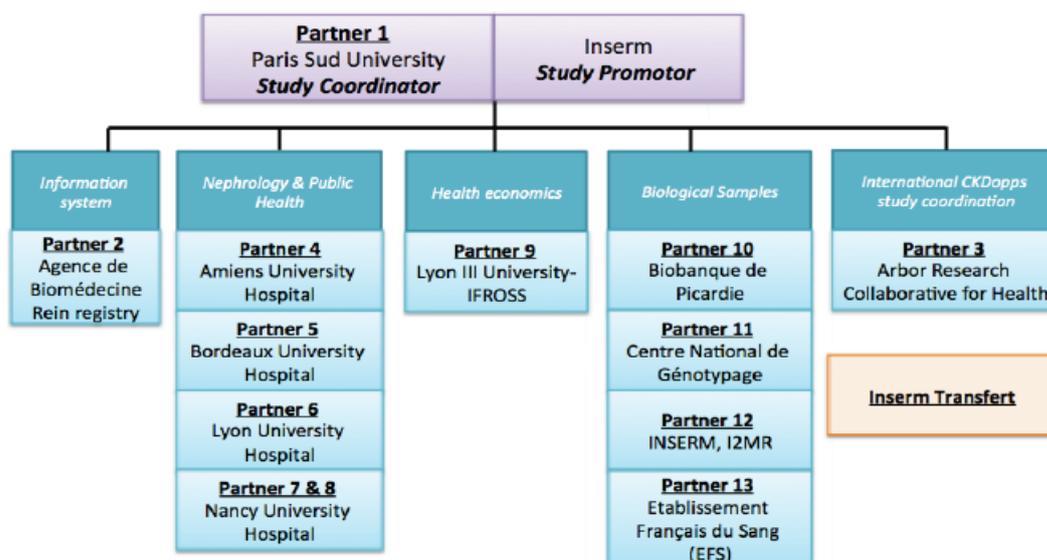
Les principaux résultats de CKD REIN montrent aussi que les épisodes d'insuffisance rénale aiguë peuvent être une cause ou une conséquence d'une hospitalisation et qu'ils peuvent être évités. Fréquents dans CKD REIN, ces épisodes étaient principalement pris en charge en dehors des services de néphrologie, souvent banalisés et sous-déclarés dans les résumés de sortie des hôpitaux, et donnant lieu à une surveillance et une gestion inadéquates des patients après l'événement.

Enfin, CKD REIN permet d'étudier les médicaments utilisés dans des conditions réelles, alors que les patients atteints de maladie rénale chronique sont souvent exclus des essais cliniques et particulièrement exposés aux prescriptions inappropriées de médicaments et à des événements indésirables. CKD REIN a mis en évidence les taux élevés de ces problèmes, ce qui a servi de base à des recommandations de l'Académie nationale de Pharmacie (2019).

4 Gouvernance du dispositif

L'établissement coordinateur de la cohorte est l'université Paris Sud et son promoteur scientifique l'Inserm. Elle donne lieu à un partenariat institutionnel entre 13 partenaires publics ou à but non lucratif, parmi lesquels l'Agence de la Biomédecine et Inserm Transfert, signataire de l'accord de consortium.

³¹Son objectif est d'étudier le rôle et les mécanismes de la toxicité liée au microbiote intestinal dans la progression de de la maladie rénale chronique par la production de toxines urémiques...



La gouvernance de CKD REIN comprend cinq comités ayant les missions suivantes :

- un conseil de partenariat comprenant des représentants des institutions partenaires et des partenaires industriels *Gold* (ceux ayant le niveau d'engagement le plus élevé) qui se réunissent chaque année pour évaluer l'état d'avancement et la situation financière du projet ;
- un comité de pilotage comprenant les 13 partenaires de la cohorte et Inserm Transfert qui se réunissent mensuellement et prennent toutes les décisions concernant l'étude ;
- un comité scientifique comprenant 15 experts internationaux qui se réunit chaque année, conseille le comité de pilotage sur les priorités de recherche et examine les projets soumis par des chercheurs universitaires ou industriels ;
- un comité industriel comprenant tous les partenaires industriels, Inserm-Transfert et trois membres du comité de pilotage qui se réunit au moins deux fois par an et apporte une contribution dans une perspective industrielle ;
- un comité d'éthique composé de 4 membres, dont un représentant des associations de patients, qui conseille le comité de pilotage sur tout aspect éthique pertinent à l'étude.

Un accord de consortium signé par les institutions partenaires et Inserm Transfert décrit la composition et les missions des cinq comités CKD REIN.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

Le budget initial estimé de la cohorte CKD REIN était de 11,4 millions d'euros.

Le PIA a financé 5,86 millions d'euros sur la période 2011-2019. Un financement supplémentaire du PIA (Biobanque) de 224 k€ a été obtenu en 2019 pour couvrir les besoins de la collecte des échantillons au suivi de 5 ans.

Pour atteindre les objectifs de l'étude, plusieurs cofinancements publics et privés ont été réalisés.

S'agissant des financements publics, le financement obtenu auprès du Programme hospitalier de recherche clinique (PHRC) pour mettre en place la partie française de l'étude internationale CKDopps a été utilisé en totalité pour développer la cohorte CKD REIN après la fusion des deux projets.

D'autres financements publics ont été obtenus pour mener des études ancillaires, dont 15% environ sont alloués à la cohorte (*cf.* tableau 2).

Tableau 1 : Sources et montants des financements publics (2010-2019)

Sources de financements	Montants en k€
PHRC (2010; PI Christian Combe, Univ Bordeaux)	540
Fondation de France (2017; PI Aurélie Untas, Univ Paris-Descartes)	60
ANR <i>Proteomark</i> (2018; PI Ziad Massy)	400
Agence national de Sécurité du Médicament (2018; PI Sophie Liabeuf, CHU Amiens)	145
International Society of Nephrology (2018; PI Sandra Wagner, CIC Nancy)	50

S'agissant des financements industriels, ceux-ci ont été obtenus *via* le partenariat public-privé géré par Inserm-Transfert (*cf.* tableau 3) ont été utilisés en totalité pour soutenir la cohorte (principalement pour l'emploi des attachés de recherche clinique et la biothèque). Au titre de la période 2012-2018, 2,86 M€ ont été effectivement perçus.

Tableau 2 : Sources et montants des financements industriels (2012-2021)

Partenaires industriels (période d'engagement)	Nombre de périodes d'engagement de 3 ans souscrites à la mi-2019	Montants en k€
MSD (2012-2018)	2	900
Baxter (2012-2018)	2	300
Fresenius Medical Care (2012-2021)	3	450
Amgen (2012-2021)	3	450
Lilly (2013-2019)	2	325
Otsuka (2015-2020)	2	375
GSK (2012-2015, 2017-2020)	2	350
Sanofi (2012-2015)	1	150
Vifor Pharma (2018-2021)	1	275
Astra Zeneca (2018-2021)	1	325
Total	-	3 900

5.2 Perspectives budgétaires pour la période 2020-2024

Le budget total demandé par les responsables de la cohorte dans le cadre de l'évaluation finale pour atteindre les objectifs de l'étude sur la période 2021-2024 est de 2,2 M€.

Ce budget comprend les dépenses de personnel, les coûts des prestations externes et les frais généraux et de gestion administrative nécessaires pour 5 des 13 partenaires sur la période. Les salaires du personnel employé en 2020 sont déjà couverts par le PIA, à l'exception de celui du nouveau data manager qui sera recruté en 2020 pour les données SNDS.

Comme suite à l'évaluation finale, le comité de pilotage chargé d'instruire cette demande a proposé en novembre 2019 un complément de financement du projet de 1,8 M€ jusqu'au 30 juin 2025, soit environ 80 % de la dotation demandée.

5.3 Les partenariats industriels

Un partenariat public-privé (PPP) a été conçu et mis en place en 2012 avec Inserm Transfert. Son objectif est de réunir des partenaires industriels pour mieux exploiter les ressources de la cohorte à travers des projets de recherche scientifique.

Ce partenariat pluriannuel, d'une durée de 3 ans renouvelable, permet à plusieurs partenaires privés d'accéder aux ressources de la cohorte dans des conditions partagées, dans un cadre qui se veut transparent et garantissant le plein respect de l'indépendance scientifique et des règles éthiques. En termes contractuels, il consiste en un accord-cadre, identique pour tous les partenaires privés, signé avec chacun d'eux, ainsi qu'avec l'Université Paris Sud et Inserm Transfert.

Le PPP ouvre les droits et accès suivants :

- un tableau de bord annuel des données, à l'inclusion et puis chaque année pendant le suivi. Le tableau de bord présente les résultats agrégés (moyenne, médiane, pourcentage) pour un certain nombre de variables sélectionnées par les partenaires privés.
- le droit de soumettre deux projets de recherche spécifiques (PRS) dans le cadre de l'engagement de 3 ans dans le PPP. Chaque PRS est soumis au Comité de pilotage pour examen selon des modalités spécifiques. S'il est approuvé, le partenaire privé a accès aux résultats des analyses de données dans un format agrégé, comme spécifié dans le protocole. Chaque PRS fait l'objet d'un accord *ad hoc* précisant à la fois les modalités et le calendrier associés à la réalisation du PRS, ainsi que ses coûts, qui sont pris en charge par le partenaire privé.
- un accès aux experts CKD REIN qui sont fortement impliqués dans la définition et le développement des PRS.
- un siège au comité industriel, lieu de partage et d'échange entre l'équipe de la cohorte et l'industrie pharmaceutique sur l'avancement du projet de cohorte, ses publications et communications, mais aussi pour les partenariats dits « *Gold* » un siège au conseil de partenariat

Le PPP bénéficie également de la relation de CKD REIN avec le consortium CKDopps. Ce lien offre de nouvelles opportunités à CKD REIN de mettre en place des PPP dans le cadre d'études internationales demandant un accès aux ressources de CKD REIN.

En 2012, six industriels ont manifesté leur intérêt pour la cohorte CKD REIN et sont devenus partenaires. Depuis 2012, trois d'entre eux ont renouvelé leur engagement et quatre autres industriels ont rejoint le PPP, ce qui fait près de six partenaires en moyenne par an.

Trois des six partenaires privés actuels de CKD REIN sont engagés jusqu'en 2021. Toutefois, il faut noter qu'à ce stade du projet, la possibilité de renouvellement de leur engagement au-delà de cette échéance est faible.

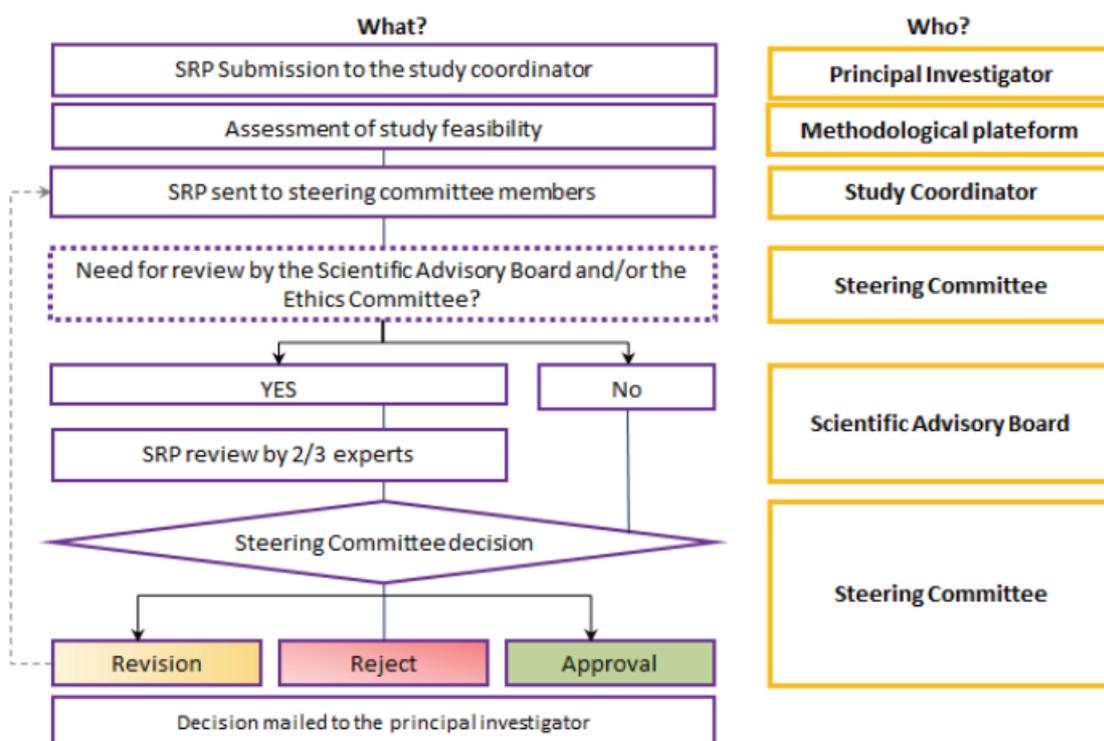
6 Accès aux données et éthique

CKD REIN est une étude épidémiologique non interventionnelle avec constitution d'une biothèque. Outre le CCTIRS, cette recherche est soumise à l'avis d'un Comité de Protection des Personnes (CPP) et à la CNIL.

L'information pour le consentement éclairé du malade porte sur :

- le recueil de données médicales à partir du dossier du service de néphrologie ;
- la conservation d'échantillons de sang et d'urine à des fins de recherches biologique et génétique ;
- la transmission par la Poste du suivi de l'adresse en cas de déménagement ;
- l'accès aux données des registres de santé tels que le registre REIN, au RNIPP et au SNIIRAM.

L'accès des chercheurs aux données et échantillons de la cohorte suit un processus spécifique de demande et d'approbation résumé par le graphique suivant :



7 Appréciations portées par la mission

La cohorte CKD REIN se consacre à l'étude d'une pathologie chronique, la MRC, importante, tant au regard de la santé publique que des comptes de l'assurance-maladie.

Et cela avec une méthodologie rigoureuse et ouverte, la conduisant ainsi à explorer de nombreuses dimensions de cette affection, médicales -comme les troubles métaboliques ou cardio-vasculaires associés, mais aussi sociales -qualité de vie, en particulier.

Sa focalisation sur la MRC avant le stade thérapeutique de la dialyse ou de la greffe garantit une complémentarité avec le registre REIN. C'est donc très logiquement que cette cohorte se voit soutenue et appuyée par l'ABM.

Le projet scientifique est cohérent et bien cadré, avec une fin des inclusions en 2016 et un suivi envisagé de façon utile sur 8 ans, avec une fin prévue en 2025

Il suscite un intérêt marqué de la part des industriels qui y ont contribué, initialement, de manière significative.

Il est à noter, à ce sujet, les faibles perspectives actuelles de prolongation de ces financements industriels au-delà de 2021.

Dans un autre registre, on peut regretter l'absence pour l'heure d'appariement avec le SNDS, dans l'attente du CASD puis de France Cohortes.

Enfin, la demande formulée d'un *data manager* à plein temps semble, aux yeux de la mission, devoir être réévaluer compte-tenu de la taille réduite de la cohorte.

LA COHORTE COBLANCE

1 Présentation générale du dispositif

1.1 Objectifs

Le cancer de la vessie constitue la 6^{ème} cause de cancer en Europe et la 5^{ème} en France avec une incidence élevée (environ 13 000 nouveaux cas par an) et des traitements qui ont peu évolué au cours des trente dernières années.

La cohorte Coblance (« cohorte prospective pour une étude intégrée des cancers de vessie ») est la première étude prospective menée en France sur cette maladie. C'est la plus grande étude au monde sur le sujet. Elle a pour objectifs :

- d'établir une classification des sous-types de cancer de la vessie selon des critères pathologiques et moléculaires et en intégrant des données génomiques ;
- d'identifier des biomarqueurs diagnostiques et pronostiques de la maladie ;
- d'évaluer les associations entre les variants génétiques, les facteurs environnementaux et liés au mode de vie, les sous-types moléculaires de cancer et l'évolution de la maladie ;
- de décrire les modalités de traitement, les consommations de soins et la qualité de vie des patients ainsi que les coûts attribuables au cancer de la vessie.

La cohorte est portée par un large consortium réunissant l'Institut national de la santé et de la recherche médicale (INSERM), établissement coordinateur, l'Institut Gustave Roussy, l'Institut Curie, la Fondation Jean Dausset, l'université Paris – Est Créteil, l'hôpital Foch, les CHRU de Lille et Clermont-Ferrand, le CHU de Nîmes et l'Assistance publique – Hôpitaux de Paris.

1.2 Principales caractéristiques de la cohorte

- Cohorte de patients volontaires couvrant des sujets d'âge adulte traités pour un cancer de la vessie confirmé histologiquement
- Inclusion et suivi réalisés dans 14 centres de soins sur le territoire métropolitain
- Création d'une large base de données épidémiologiques, cliniques, biologiques, économiques et de mode de vie, observationnelle et prospective

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants et modalités d'inclusion

Les inclusions dans la cohorte Coblance ont démarré en novembre 2012 et se sont achevées en juin 2018, avec un taux de refus de participation relativement faible (5,7%) : le nombre de participants s'établit au final à 1 805, soit 90 % de l'objectif fixé à l'origine du projet (2 000), dont 79 % de fumeurs

(54 %) ou ex-fumeurs (25 %) sachant que le tabagisme est identifié comme le principal facteur de risque du cancer de la vessie. Depuis juin 2018 la cohorte est fermée.

Au mois de juin 2019, le suivi a effectivement concerné 1 124 patients au bout d'un an (62 %) et 703 patients à deux ans (39 %), ce qui traduit une attrition relativement importante et rapide, qui a nécessité la mise en place de dispositifs correctifs de relance des patients pour s'efforcer de la réduire, conformément aux préconisations du jury dans le cadre de l'évaluation à moyen terme ; en outre 275 décès sont recensés au 30 juin 2019.

2.2 Contenu, collecte et stockage des différents types d'informations

Les informations recueillies à l'inclusion puis ultérieurement dans le cadre du suivi pour un certain nombre d'entre elles portent notamment sur :

- des données sociodémographiques, relatives au mode de vie et à l'environnement, aux comorbidités, à la présence de cancer dans la famille (sous forme de questionnaires) ;
- des données cliniques portant sur le management de la maladie, les éléments de diagnostic, la pathologie et les traitements, la progression de la maladie, les récurrences, la survenue du décès ;
- des données de qualité de vie, sous forme d'auto-questionnaires ;
- des données de consommation de soins (hospitalisations, imagerie, examens de ville, arrêts maladie à partir de 6 mois après la chirurgie initiale).

Les échantillons biologiques collectés (environ 8 000 globalement) consistent en échantillons sanguins (100 % des patients à l'inclusion), échantillons de tumeurs, soit congelés (1 395 patients à l'inclusion, soit 77,3 %), soit conservés dans des blocs de paraffine (1 588 patients, soit 88 %), échantillons d'urine (1 560 patients, soit 86,4 %) et d'ongle (1 093 patients, soit 60,7 %). Afin de permettre des analyses génétiques, de l'ADN/ARN a été extrait à partir des échantillons urinaires, sanguins et de tissus tumoraux.

Les données collectées par des personnels spécialisés (attachés de recherche clinique) dans les 14 centres investigateurs alimentent une base de données dédiée au suivi de la cohorte qui est hébergée à l'Institut Gustave Roussy.

Les échantillons sont traités et stockés dans trois centres de ressources biologiques (CRB) différents, qui sont partenaires du projet dans le cadre du consortium et de sa gouvernance :

- les échantillons sanguins au CRB du centre d'étude du polymorphisme humain (CEPH) de la Fondation Jean Dausset ;
- les échantillons de tissus tumoraux au CRB de l'Institut Curie ;
- les échantillons d'urine et d'ongles au CRB de l'hôpital Henri Mondor (APHP).

2.3 Articulation avec d'autres cohortes ou dispositifs d'information

Le projet d'appariement de type probabiliste avec les données du SNDS, indispensable pour enrichir le suivi et élargir les perspectives de recherche, est en cours d'examen par les différentes instances compétentes (l'INDS a été saisi en juillet 2019).

2.4 Horizon et perspectives envisagés pour le dispositif après 2020

Il est prévu un suivi actif d'une durée de six ans conformément aux recommandations de l'association européenne d'urologie (après l'inclusion à l'occasion de la première visite, sont programmées des visites de routine à trois mois, six mois, douze mois, puis annuellement). Compte tenu de la date de fin des inclusions, le suivi actif est prévu pour durer jusqu'en 2024. Le suivi sera ensuite passif.

La richesse des données de la cohorte, notamment en vie réelle, et leur intérêt en phase avec les nouveaux axes de développement des thérapies du cancer de la vessie ont déjà permis (Merck en 2016 – dans le cadre d'une extension d'indication du pembrolizumab pour le traitement du cancer de la vessie métastatique auprès de la Haute autorité de santé - HAS) et paraissent pouvoir permettre (Ferring, Janssen) la conclusion de partenariats significatifs avec les laboratoires pharmaceutiques.

3 Principaux résultats et retombées en termes de recherche et de santé publique

En septembre 2019, compte tenu du retard relativement important pris pour la mise en place de la cohorte du fait des difficultés rencontrées pour la mise à niveau des centres investigateurs et dans la fixation des mécanismes de conventionnement, l'étude COBLANCE n'a pas encore produit de résultat scientifique. Toutefois, afin d'accroître la visibilité de la cohorte, ainsi que le jury en a souligné la nécessité dans le cadre de l'évaluation intermédiaire, l'équipe COBLANCE a rendu public le protocole de l'étude, a récemment présenté des communications orales en congrès à son sujet et prépare pour publication un article décrivant la population à l'inclusion.

Par ailleurs, des perspectives sont ouvertes dans plusieurs directions, permettant de conforter l'exploitation des données et le financement de la cohorte.

L'équipe COBLANCE ainsi a répondu à plusieurs appels à projet lancés par des institutions françaises :

- en tant qu'équipe de recherche, elle a ainsi été retenue pour le projet SEnACLE, financé par l'institut national du cancer (Inca) à hauteur de 0,89 M€ sur la période 2019 – 2022 et consacré à l'étude de l'impact du tabagisme sur les caractéristiques moléculaires et le pronostic du cancer de la vessie ;
- elle a été également retenue pour le projet Comebac, financé par l'Inca et l'Iresp à hauteur de 0,19 M€ consacré notamment à l'étude des coûts de prise en charge des cancers de la vessie, identifiés par la littérature scientifique comme très élevés, en utilisant les données de COBLANCE chaînées aux données du SNDS ;
- en outre, elle s'inscrit dans les priorités d'étude définies dans le cadre du programme transversal de l'INSERM sur la variabilité génomique.

Afin d'accroître sa visibilité au niveau international, l'équipe COBLANCE s'est engagée dans des collaborations internationales avec des équipes étrangères portant également des cohortes de patients atteints d'un cancer de la vessie (études Epicuro et Isblac en Espagne, BCPP au Royaume Uni).

Le jury international a souligné l'excellence de l'étude, fondée notamment sur la robustesse de sa méthodologie et sur l'expertise et la dimension interdisciplinaire de l'équipe qui l'a mise en place, dont les développements restant à réaliser (génomique) et l'appariement prochain au SNDS devraient encore accroître l'intérêt scientifique.

4 Gouvernance du dispositif

L'étude COBLANCE est portée par l'équipe « Méthodologie et épidémiologie clinique pour l'oncologie moléculaire », une des neuf équipes du centre de recherche en épidémiologie et santé des populations (CESP), unité mixte de recherche et de service INSERM – Université Paris Sud – Université Versailles Saint-Quentin.

COBLANCE dispose d'une gouvernance à trois niveaux, relativement lourde, reposant sur :

- une instance décisionnelle : le comité exécutif, qui se réunit, soit en format restreint, comprenant la coordinatrice scientifique (Simone Benhamou – INSERM), la cheffe de projet et les représentants des principaux autres membres du consortium (IGR, Institut Curie, hôpital Foch, université Paris Est - Créteil), soit en format élargi à tous les membres du consortium, ce qui permet d'assurer une représentation des CRB et des centres investigateurs. Le CE en format restreint qui concentre les décisions stratégiques tout en prenant les principales décisions opérationnelles se réunit tous les deux mois, le CE en format élargi se réunit annuellement ;
- une instance consultative : le comité scientifique et éthique, qui compte vingt scientifiques internationaux, se réunit une fois par an ;
- s'y ajoute un « comité stratégique » dans lequel sont représentés l'ensemble des établissements parties à l'accord de consortium.

Le support et le soutien à l'activité de la cohorte (aspects réglementaires et financiers) sont fournis la direction régionale de l'Inserm compétente. Inserm Transfert, filiale de l'Inserm, prend en charge la définition et la mise en œuvre des partenariats avec les industriels.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

Selon les données communiquées par l'agence nationale de la recherche (ANR), les recettes du projet COBLANCE se sont globalement élevées sur la période 2012-2019 à 4,95 M€, dont 3,95 M€ au titre des financements du PIA (compte tenu du retard pris pour la mise en place de la cohorte, une part significative de ce financement reste disponible – 1,87 M€ au début de 2018 – et pourra être utilisé avec l'accord de l'ANR jusqu'en 2024 pour couvrir les charges liées à la mise en place complète de la cohorte).

Sur la période 2012-2019, 63 agents ont contribué au projet COBLANCE, dont 28 urologues et pathologistes et 17 attachés de recherche clinique (ARC) exerçant dans les centres investigateurs, 7 techniciens exerçant dans les trois CRB, le reste des agents (11) étant ou ayant été directement en charge de la mise en place de la cohorte au titre de l'équipe projet. Hormis le chef de projet COBLANCE, ces personnels n'ont consacré qu'une part, variable, de leur activité au projet ; en moyenne annuelle, 7,8 ETP ont été mobilisés pour le projet en période d'inclusion et 5,5 ETP en

période de suivi. Les 63 agents qui ont contribué au projet COBLANCE sont pour l'essentiel des personnels permanents (personnels statutaires de l'Inserm ou des hôpitaux, agents titulaires d'un CDI) ; 8 étaient ou sont en CDD

Les financements du PIA ont couverts les charges liées aux ARC, à une partie du coût des personnels des CRB et à une partie du coût de l'équipe projet.

Outre le financement au titre du PIA (3,95 M€), le projet COBLANCE a obtenu de 2012 à 2019 les principaux financements publics et privés suivants (en M€) :

Financements d'établissements publics (Inserm)	Ligue nationale contre le cancer	Laboratoires Merck
0,52	0,17	0,31

5.2 Perspectives budgétaires pour la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale de septembre 2019 portait une demande de financement complémentaire de 1,65 M€ destiné à achever les opérations de suivi des patients inclus et à consolider la cohorte. Le projet de budget prévisionnel global cumulé sur la période 2020-2024 associé à cette demande s'élevait à 3,51 M€.

Dans ce cadre des renforts étaient demandés notamment en statisticiens au titre de l'analyse des données résultant de l'appariement au SNDS, en ingénieurs et techniciens pour les extractions et analyses moléculaires.

Comme suite à l'évaluation finale, le comité de pilotage chargé d'instruire cette demande a proposé en novembre 2019 un complément de financement du projet à hauteur de 0,9 M€ jusqu'au 30 juin 2025.

6 Accès aux données

Les modalités d'accès des chercheurs aux données ou échantillons collectés dans le cadre de COBLANCE sont définies par une charte qui prévoit la publication annuelle d'appels à projet par l'équipe responsable de la cohorte et l'examen en trois temps des projets soumis : vérification par l'équipe de la complétude de la candidature, examen par le comité exécutif de sa faisabilité, appréciation par deux membres du comité scientifique de son intérêt en termes de recherche.

Les projets présentés par les industriels sont soumis au comité exécutif et au comité scientifique de COBLANCE. Les modalités de partenariat mises en place en lien avec Inserm Transfert prévoient que les financements versés doivent contribuer pour partie à financer les charges de constitution et de maintenance de la cohorte.

7 Appréciations portées par la mission

COBLANCE est une cohorte à la thématique circonscrite, bien reconnue niveau européen. Elle dispose de collections biologiques et génétiques très complètes et bénéficie d'importants cofinancements sur projets de l'Iresp et de l'Inca. Elle a en outre engagé une série de collaborations prometteuses pour l'étude des expositions environnementales et professionnelles.

Le rôle de structuration et d'appui joué par la Fondation Jean Dausset est à cet égard à souligner et permet un travail bien coordonné entre les centres.

Toutefois, la cohorte a rencontré une série d'obstacles réglementaires et informatiques difficiles à franchir et a connu un démarrage retardé impliquant une fin de suivi reportée à 2024. Son arrivée à maturité ne sera effective qu'à compter des prochaines années.

Sous réserve de sa consolidation effective sur la base de la poursuite de financements publics suffisants, les perspectives de partenariats industriels paraissent significatives.

LA COHORTE CRYOSTEM

1 Présentation générale du dispositif

1.1 Objectifs

L'allogreffe de cellules souches hématopoïétiques (CSH) est le seul traitement curatif de nombreuses maladies du sang malignes ou non (leucémie, lymphome, aplasie médullaire, par exemple). La complication majeure en est immunologique : la maladie du greffon contre l'hôte (GvHD pour *Graft-versus-Host Disease*, en anglais), dans laquelle le système immunitaire du donneur reconnaît le patient greffé comme étranger et le rejette (incidence : de 10% à 20%). Si la GvHD constitue la première cause de mortalité liée à la greffe, la physiopathologie de cette maladie et des complications à long terme sont encore mal connues. C'est dans ce contexte que la cohorte nationale CRYOSTEM a été initiée en 2010 (service Hématologie Greffe Hôpital Saint-Louis, Paris, et Centre de Thérapie Cellulaire, Institut Paoli-Calmettes, Marseille) avec le soutien de la Société Francophone de Greffe de Moelle et de Thérapie Cellulaire (SFGM-TC). D'abord centrée sur la seule GvHD, CRYOSTEM a élargi, depuis novembre 2016, son action à l'ensemble des complications de l'allogreffe de CSH. Les objectifs de CRYOSTEM sont :

- la constitution d'une collection d'échantillons biologiques homogènes et de qualité, issus de couples patients/donneurs pré- et post-allogreffe selon une cinétique bien définie ;
- la valorisation de cette collection dans des études scientifiques d'envergure sur les complications de l'allogreffe de CSH pour en comprendre les mécanismes biologiques et développer de nouveaux tests diagnostic et thérapies ;

Le but ultime de CRYOSTEM est l'amélioration de la prise en charge des patients allogreffés grâce à une meilleure compréhension des complications. En parallèle de ses objectifs initiaux, CRYOSTEM s'est fixé des objectifs secondaires pour assurer la constitution d'une collection performante et de qualité :

- fédérer la majorité des acteurs nationaux, ici unités de greffe et CRB, dans un réseau couvrant la quasi-totalité du territoire ;
- maintenir un niveau de conformité réglementaire optimal en adéquation avec l'évolution de la cohorte ;
- mettre en place une base de données sécurisée, unique et évolutive, utilisable à distance par tous les centres pour l'enregistrement et la centralisation des données et des échantillons.

1.2 Principales caractéristiques de la cohorte

Initialement, cette cohorte prospective et multicentrique associait 33 des 36 unités de greffe françaises et 23 Centres de Ressources Biologiques (CRB). La collection compte aujourd'hui près de 200 000 échantillons, issus de plus de 5 700 patients et 2 300 donneurs. Dans le cadre d'appels à projets, plus de 6 000 échantillons ont ainsi été valorisés dans 10 projets de recherche depuis 2015.

CRYOSTEM constitue l'unique cohorte européenne et internationale, financée par des fonds publics, dans le champ des complications de l'allogreffe de CSH. Du fait de son envergure et du taux élevé d'inclusion, la cohorte CRYOSTEM est représentative de la population de patients transplantés en termes d'âge, sexe, type de greffe, pathologie.

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

Lancée le 01/07/2012, la cohorte a été close le 31/12/2019. Tel que déposé à l'ANR en 2010, le projet initial prévoyait d'inclure 616 patients greffés par an. Au 31 juillet 2019, 5 749 patients adultes et pédiatriques, ainsi que 2 374 donneurs, ont été inclus dans le protocole CRYOSTEM, dépassant ainsi d'environ 33 % l'objectif initial.

Depuis son initiation, et jusqu'au 31 décembre 2017, la cohorte CRYOSTEM a connu une croissance constante en termes d'inclusions, avec un rythme moyen d'environ 1 100 inclusions/an entre 2013 et 2017. Depuis le 1er janvier 2018, l'arrêt des inclusions est rentré en vigueur suite au vote du Comité Plénier de novembre 2017 pour répondre aux contraintes budgétaires liées au financement des PIA et à la pérennisation de la collection. Au 31 juillet 2019, 94 % des consentements des receveurs et donneurs étaient présents dans la base de données. Grâce à sa structuration performante, CRYOSTEM a réussi à inclure 82 % des patients greffés dans les centres de son réseau (en référence aux chiffres d'allogreffe 2017 de l'Agence de la Biomédecine).

2.2 Modalités d'inclusion et d'échantillonnage

Une fois son consentement écrit obtenu (ou de celui de ses ayant-droits), le patient peut être inclus dans le protocole. De la même manière, les donneurs intrafamiliaux géno- ou haplo-identiques sont éligibles au protocole. Depuis novembre 2016, et en collaboration avec l'Agence de la Biomédecine, les donneurs volontaires non-apparentés du fichier français peuvent être également inclus dans la cohorte. Les unités de greffe partenaires du réseau CRYOSTEM sont en charge de l'identification des patients éligibles, de leur information et de la récupération des consentements. Le patient ayant donné son consentement par écrit est inclus dans le protocole. Dans le cas de patients mineurs, le(s) représentant(s) légal(aux) consent(ent) pour l'enfant. À tout moment du protocole, le patient (ou le donneur) peut retirer son consentement, impliquant la destruction de ses échantillons biologiques. Chaque patient inclus dispose d'un suivi biologique et clinique jusqu'à deux ans post-greffe. La cinétique de prélèvements inclut des prélèvements à la survenue de la GvHD, aiguë et/ou chronique et au cours de sa prise en charge, et prend en compte les événements médicaux qui peuvent survenir (rechute, non prise de greffe, décès...).

2.3 Contenu, collecte et stockage des différents types d'informations

La collection de ressources biologiques annotées s'articule autour du recueil d'échantillons biologiques et de données cliniques. Pour chaque patient inclus dans le protocole CRYOSTEM, ayant donné son consentement écrit, des prélèvements sanguins sont intégrés au parcours de soin du patient selon une cinétique de prélèvement bien précise. Avec l'élargissement de la thématique à toutes les complications de l'allogreffe de CSH en novembre 2016, les prélèvements sont systématiques à 100 jours et 1 an post-greffe et un nouveau prélèvement systématique à 2 ans post-greffe a été introduit.

Le réseau CRYOSTEM regroupe aujourd'hui la totalité des unités de greffe en activité sur le territoire français. Chaque unité de greffe collabore en binôme avec un CRB, afin de limiter le délai entre le prélèvement et le traitement de l'échantillon sanguin : en 2018, environ 65% des prélèvements sanguins ont été traités en moins de quatre heures, assurant ainsi une préservation optimale des protéines plasmatiques.

Les CRB procèdent au traitement pré-analytique des prélèvements sanguins selon les procédures et protocoles en vigueur, en accord avec la norme française NF S 96-900. De plus, la gouvernance de CRYOSTEM fournit des réactifs définis comme critiques aux CRB (RPMI, solution de lyse et tubes Ficoll) permettant de garantir l'homogénéité de traitement des échantillons. Trois types d'échantillons cryopréservés sont générés à partir d'un prélèvement sanguin : cellules viables en DMSO, plasma et culots leucocytaires.

Le stockage des échantillons dérivés cryopréservés est décentralisé : chaque CRB gère en local le stockage des échantillons générés dans le cadre du protocole, CRYOSTEM restant l'unique responsable de la collection. Les CRB sont responsables de la mise à disposition des échantillons pour les projets de recherche sélectionnés.

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

Une collaboration a été initiée en avril 2015 avec l'intergroupe français d'étude des leucémies aiguës myéloïdes responsable du protocole BIG-1. A ce jour, plus de 250 patients du protocole BIG-1 ont été inclus dans la cohorte CRYOSTEM. Depuis juillet 2018, CRYOSTEM collabore avec la cohorte prospective multicentrique L.E.A. (Leucémies de l'Enfant et de l'Adolescent, Hôpital de la Timone, Marseille).

Enfin, la cohorte CRYOSTEM a initié récemment un rapprochement avec le réseau F-CRIN (*French Clinical Research Infrastructure Network*) piloté par l'INSERM. Horizon et perspectives envisagés pour le dispositif après 2020.

3 Principaux résultats obtenus

CRYOSTEM a réussi à créer la seule collection de ressources biologiques en Europe dédiée aux complications de la greffe de moelle osseuse. CRYOSTEM a également développé un modèle de biobanque en réseau innovant et efficient. Depuis mi-2018, CRYOSTEM met son expertise en réseau au service de protocoles cliniques portés par la SFGM-TC pour la réalisation d'études ancillaires et aussi d'autres cohortes pour l'enrichissement de leurs propres données épidémiologiques (ex. cohorte L.E.A.). CRYOSTEM est la seule cohorte internationale disposant d'échantillons biologiques post- allogreffe de CSH jusqu'à deux ans post-greffe.

Plus largement, au regard de la politique de santé publique, l'allogreffe de CSH reste encore une problématique grave, compte tenu des taux de morbidité et de mortalité liés à la procédure, mais aussi de la prise en charge du patient pré- et post- allogreffe. C'est donc fort logiquement que le Plan Greffe 1 (2000-2003) de l'Agence de la Biomédecine a fait du soutien à la recherche dans le domaine de l'allogreffe de CSH une priorité, notamment avec les PHRC (Programme Hospitalier de Recherche Clinique) dédié à la recherche clinique.

4 Gouvernance du dispositif

CRYOSTEM s'est doté du statut d'association loi 1901 à but non lucratif le 18 juillet 2015 pour faciliter la gestion administrative et financière. Depuis le 1^{er} mai 2016, l'Association CRYOSTEM remplace la SFGM-TC au titre de partenaire coordinateur. La gouvernance de CRYOSTEM est constituée du comité de pilotage, du comité plénier et des comités stratégique et scientifique.

Le comité de pilotage est responsable de la coordination des opérations au sein du réseau et de la transmission des informations. Il fait le lien entre l'ANR, les partenaires et les différents comités constituant la gouvernance. Le Comité de Pilotage veille à la bonne exécution et la mise en œuvre des décisions validées par le Comité Plénier. Le Comité de Pilotage se réunit mensuellement.

Le comité plénier est en charge de la validation des orientations stratégiques proposées par les comités scientifique et stratégique et de toute éventuelle modification du projet, lors de la réunion annuelle (tenant également lieu de revue de direction pour la certification ISO) durant le congrès de la SFGM-TC. Environ une soixantaine de médecins et personnels de CRB assistent chaque année à la réunion du Comité Plénier.

Le comité scientifique est en charge de la valorisation de la collection à travers les appels à projets annuels et de la sélection des projets qui pourront accéder à la collection ainsi que de la gestion et de l'administration de projets de recherche scientifique d'envergure portés par CRYOSTEM. Il s'appuie sur des experts internationaux indépendants pour l'évaluation des candidatures soumises dans le cadre des appels à projets. Les réunions du comité scientifique ont lieu tous les deux mois environ.

Le comité stratégique gère les aspects stratégiques et propose des orientations relatives au fonctionnement du réseau. Le comité stratégique se réunit au moins une fois par an.

CRYOSTEM entend ouvrir sa gouvernance à des structures publiques telles que l'INSERM et des universités, pour étendre les collaborations académiques.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

CRYOSTEM est l'une des 10 cohortes sélectionnées par l'appel à projets de 2010 de l'ANR et financée depuis 2011 grâce au soutien du « Programme des Investissements d'Avenir » (PIA). Un soutien financier de 5,24 M€ a été apporté dans ce cadre, complété par 0,48 M€ de contributions diverses, selon les données communiquées par l'ANR, notamment de la part de l'INCA dans le cadre de projets de recherche.

5.2 Perspectives budgétaires pour la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale portait une demande de financement complémentaire de 1,32 M€ destiné à consolider la cohorte. Le budget prévisionnel global cumulé du projet sur la période 2020 – 2024 associé à cette demande s'élevait à 2,34 M€.

A l'issue de l'évaluation finale, le comité de pilotage chargé d'instruire cette demande a proposé un complément de financement de 0,9 M€ jusqu'au 30 juin 2025.

5.3 Les partenariats industriels

Les journées scientifiques annuelles sont organisées et financées dans leur quasi-globalité grâce au soutien d'industries pharmaceutiques pour des montants allant de 500 à 5.000 euros.

Depuis la fin de l'année 2017, CRYOSTEM a créé le fonds de dotation HTC Project avec pour objectif de lever des fonds. Le tableau ci-dessous présente les montants alloués et leurs affectations. Depuis la fin 2017, près de 580 000 euros ont été levés par le HTC Project.

Donateur	Montant alloué	Affectation
Association Gueriduncancer	20 000,00 €	Projet CRYO-LEA
Donateurs individuels via la plateforme Commeon	117,52 €	
MSD France	500 000,00 €	
Association Tribal Sport	3 800,00 €	
Association Aquitaine Espoir	20 000,00 €	Projet Cartographie du système immunitaire (projet CS 17-01 CRYOSTEM <i>"Deep phenotyping, transcriptomic and metabolomics mapping of donors and recipients after HSCT"</i>)
Association Vaincre la Leucémie / Association Laurette Fugain	40 000,00 €	Projet Exome (projet CS 17-05 CRYOSTEM <i>"Role of the exome sequencing in the GvHD prediction post-HSCT"</i>)
Donateur individuel	5 000,00 €	Projet AlloZithro (projet CS 18-01 CRYOSTEM <i>"Deciphering biological mechanisms involved in the regulation of anti-tumor response after HSCT by azithromycin intake"</i>)
Donateur individuel	6 010,00 €	Conception et impression de la BD dédiée au HTC Project <i>"97 jours en chambre stérile"</i>
Association Ensemble avec Benoît	1 000,00 €	Sans affectation particulière
Donateurs individuels via les plateformes LyfPay et HelloAsso	2 322,50 €	Sans affectation particulière

6 Accès aux données et éthique

Les consentements et les documents d'information utilisés dans le cadre des inclusions dans la cohorte CRYOSTEM ont été rédigés en collaboration avec différents comités de patients et validés par le CPP Sud-Méditerranée I en février 2012. Toutes les autorisations réglementaires (CPP, CNIL, CCTIRS, MESRI) ont été obtenues et ont permis la première inclusion en juillet 2012. Les patients sont informés de l'utilisation des ressources biologiques et des données pour des recherches sur les caractéristiques génétiques. Depuis fin 2018, la gouvernance de CRYOSTEM s'est engagée à l'intégration des exigences du RGPD entré en vigueur le 25 mai 2018. Un délégué à la protection des données pour CRYOSTEM auprès de la CNIL a été désigné, avec une date de prise d'effet le 24/05/2019.

En ce qui concerne l'accès aux données, depuis mi-2015, les membres de la communauté scientifique nationale et internationale, académique et industrielle, peuvent accéder à la collection de ressources biologiques CRYOSTEM via les appels à projets annuels. Les appels à projets sont régis par des conditions générales révisées et éditées annuellement en collaboration avec le cabinet d'avocats Fidal. Le comité scientifique de CRYOSTEM est en charge de l'évaluation et de la sélection des candidatures. Une fois sélectionné, des entretiens sont organisés avec les porteurs de projets pour déterminer les critères de sélection des ressources biologiques et programmer les mises à disposition.

Les ressources biologiques CRYOSTEM sont accessibles selon la tarification prévue par les conditions générales, dépendante du statut du porteur de projet.

7 Appréciations portées par la mission

CRYOSTEM offre l'exemple d'une initiative de cohorte et de biobanque circonscrite dans son objet, par conséquent limitée dans son effectif, isolée institutionnellement lors de son lancement et ayant néanmoins réussi à installer sa place dans le paysage français des cohortes.

Ce résultat, très positif, est dû à la méthodologie particulièrement rigoureuse de l'équipe et à son haut niveau scientifique, qui lui ont permis d'associer à sa démarche tous les centres de greffe français.

Les retombées scientifiques et médicales, individuelles comme collectives, sont d'ores et déjà très satisfaisantes et devraient l'être encore davantage dans les prochaines années (notamment le nombre des publications).

La structuration institutionnelle actuelle -association régie par la loi de 1901- est fragile et pourrait être réexaminée dans le cadre de la future gouvernance associant les Universités. De même, il existe une certaine fragilité des aspects communication et relations avec les industriels dans le cadre juridique actuel.

Une autre faiblesse potentielle réside dans l'absence de centralisation de la biobanque sur un site unique.

Il n'existe pas, pour l'heure, de chaînage avec les données du SNDS.

À noter, enfin, des recoupements mal maîtrisés du projet associé et complémentaire CRYO LEA avec la cohorte HOPE-EPI.

LA COHORTE HEPATHER

1 Présentation générale du dispositif

1.1 Objectifs

La cohorte a été lancée en août 2012 dans le contexte d'un changement de paradigme quant aux options thérapeutiques de l'infection chronique par les virus des hépatites B (VHB) et C (VHC), avec plus de 70 nouveaux médicaments alors au stade de l'étude clinique. Dès lors, son principal objectif était d'évaluer l'efficacité clinique et la sécurité de ces nouveaux traitements en « vie réelle », et de hiérarchiser leurs apports thérapeutiques respectifs. La cohorte comporte une biobanque centralisée et un chaînage avec le SNDS.

Son ambition est d'être une plateforme ouverte à la communauté scientifique française et internationale, qu'elle soit publique ou privée, et cela à travers l'établissement de projets de recherche spécifiques (PRS). Mais aussi d'intégrer la génétique, la pharmacogénomique, la santé clinique et environnementale, les données comportementales et les coûts afin de permettre des recherches trans-disciplinaires et translationnelles.

1.2 Principales caractéristiques de la cohorte

La cohorte ANRS CO22 HEPATHER a été initiée en collaboration avec l'Association française pour l'étude du foie (AFEF). C'est une cohorte unique, par son ampleur et ses objectifs. Elle permet d'étudier l'évolution de la maladie hépatique liée au virus B ou C (VHB ou VHC) chez plus de 21 000 patients au stade chronique de l'infection ou guéris. Cette cohorte, dont l'Inserm-ANRS est le promoteur, associe organismes de recherches, sociétés savantes, institutions de santé et industriels. En 2018, du fait du contexte scientifique et thérapeutique, le partenariat industriel, géré par Inserm Transfert, initialement composé de 6 laboratoires, s'est réduit à 3. En effet, les nouveaux médicaments contre le VHC permettant de guérir près de 95 % des patients, les industriels ont revu leur stratégie de développement et d'investissement.

La plateforme épidémiologique a inclus 21 269 patients au 31 décembre 2018 dans 38 centres experts en hépatologie, impliquant plus de 50 personnels de recherches. La biothèque, gérée et stockée par la société *Cell and Co* à Pont du Château en Auvergne, contient plus de 900 000 échantillons biologiques, et continue d'être exploitée pour les PRS. De nouveaux PRS hébergés au sein de la cohorte ont été initiés en 2018.

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

Les ouvertures des centres participants se sont échelonnées d'août 2012 à septembre 2014. 32 centres ont été ouverts et actifs. Après une suspension au 31/12/15, les inclusions ont repris le 05/07/16 afin d'inclure les patients des cohortes ANRS CO12 CIRVIR (cohorte de patients atteints du VHB ou du VHC cirrhotiques) et ANRS CO23 CUPILT (cohorte de patients VHC transplantés traités

par AAD) pour continuer leur suivi sur le long terme. 6 « nouveaux centres » ont été ouverts en 2016 et actifs. La période d'inclusion a pris fin au 30/06/18 en France et prendra fin au 30/04/19 dans le centre de Bruxelles en Belgique.

Au 31 décembre 2018, plus de 21 000 patients étaient inclus., dont 15 000 avaient été atteints par l'hépatite C et 6 000 par l'hépatite B.

2.2 Modalités d'inclusion et d'échantillonnage

Les inclusions des patients et le recueil de leurs données sont collectés dans un cahier d'observation électronique. De ce fait, cet outil est le principal indicateur d'avancement du projet. On peut y suivre entre autres le nombre d'inclusions, les visites de suivi réalisées. Des tableaux de bord ont été mis en place au début de l'étude, ils permettent de suivre les caractéristiques des patients inclus.

Afin de suivre le remplissage des données et d'augmenter l'efficacité de ce recueil, un bilan électronique de suivi des saisies en ligne a été mis en place en mars 2017 et accessible pour les centres. Ainsi en temps réel, les centres ont connaissances de l'effectif, du pourcentage et des codes patients pour lesquels soit 1. la visite d'inclusion est incomplète, 2. aucune donnée de suivi n'a été saisie ou 3. aucune nouvelle n'a été renseignée depuis plus de 15 mois. Par ailleurs, à la suite de chaque monitoring, un courrier est envoyé aux centres.

Afin d'améliorer le suivi des patients ne venant plus en consultation, le Comité de Protection de Personnes a donné son accord pour le recueil d'information auprès des patients par téléphone. Au cours du dernier trimestre 2018, un web designer a été recruté afin de construire le site internet de la cohorte. Celui-ci est accessible depuis le 27 Mars 2019.

2.3 Contenu, collecte et stockage des différents types d'informations

Un des points forts de cette cohorte est la constitution d'une biothèque. Celle-ci est réalisée uniquement à l'inclusion des patients, et représente 40 tubes par patient. Toutes les biothèques réalisées sont stockées à -80°C. Initialement, la biothèque était gérée par l'INSERM SC10 et stockée à l'Établissement Français du Sang (EFS) à Beynost. Ce flux important d'échantillons a saturé le système de stockage existant.

Suite à la publication d'un appel d'offre en août 2013, depuis le premier trimestre 2014, la coordination et le stockage de la biothèque HEPATHER sont désormais assurées par la société Cell&Co, à Pont du Château en Auvergne (utilisation de tubes 2D, code-barre, de collectes régulières dans les centres, ...). Depuis fin 2015, les biothèques des cohortes ANRS HEPATHER, CIRVIR et CUPILT sont stockées chez *Cell&Co*. La biothèque représente à ce jour plus de 900 000 échantillons qui sont à disposition pour la communauté scientifique.

Des biothèques de suivi spécifiques ont été créées pour certains projets ancillaires de la cohorte : l'Observatoire de la Résistance et le projet LICAVIR.

Les données informatiques de la cohorte sont stockées sur des serveurs sécurisés situés à l'Institut Pierre Louis d'épidémiologie et santé publique, UMR de l'INSERM et Sorbonne Université.

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

En août 2018, l'ANRS a obtenu l'autorisation de la CNIL pour l'accès aux données du Système Nationale des Données de Santé (SNDS) des patients inclus dans la cohorte depuis l'année de leur inclusion. Ce chaînage permettra de suivre les consommations de soins des patients y compris ceux qui ne se présentent plus en consultation.

Cet accès nécessite l'autorisation des patients : un consentement spécifique a été mis en place en septembre 2014. Au 31 décembre 2018, il a été proposé à 74,9 % des patients de participer à ce chaînage. Il y a eu un taux de refus de 4,1 %. Ainsi, 15 932 patients avaient exprimé leur accord.

L'appariement ne devait prendre place qu'à la fin 2019, compte tenu de la nécessité d'établir des circuits d'information avec des tiers de confiance (CNAV).

2.5 Horizon et perspectives envisagés pour le dispositif après 2020

Les enjeux scientifiques futurs pour 2020-2025 pour la cohorte HEPATHER sont les suivants :

- la réversibilité de la fibrose du foie et son évaluation en particulier dans l'hépatite C guérie sur le plan virologique ;
- la guérison de l'hépatite B chronique traduite par la négativation de l'antigénémie HBs (dite guérison « fonctionnelle ») : plus de 250 millions de personnes sont atteintes d'hépatite B chronique dans le monde, dont 280 000 en France (malgré la vaccination préventive) ; la guérison fonctionnelle est aujourd'hui identifiée comme l'objectif thérapeutique des cinq prochaines années, permettant d'envisager l'interruption des traitements antiviraux au long cours actuellement proposés à ces patients ;
- le carcinome hépatocellulaire (4ème cause de décès par cancer chez l'homme en France) et notamment son épidémiologie, sa physiopathologie et les nouvelles stratégies curatives et préventives, incluant les immunothérapies anti checkpoints immunitaires ;
- la stéatose hépatique et le syndrome métabolique, puisque, malgré la guérison, un tiers de patients ont une hypertension artérielle, un diabète et/ou une prise de poids qui risquent de limiter les capacités de réparation de leur hépatite virale guérie ou viro-supprimée ;
- les innovations thérapeutiques en développement (anti-fibrosants, antiviraux, anti stéato-hépatite, anticancéreux).

La poursuite de la cohorte HEPATHER sur la période 2020-2025, permettrait de mettre en place une collection biologique de suivi qui en augmenterait l'attractivité et le potentiel pour la réalisation de recherches cliniques et translationnelles (notamment autour des biomarqueurs) pour des partenaires académiques ou industriels (puisque nous disposerions alors de deux points de prélèvement séparés de 5 à 8 ans). Les responsables de la cohorte proposent, par ailleurs, l'inclusion d'une seconde vague de patients atteints de l'hépatite B afin d'augmenter la puissance des analyses à venir et surtout d'adapter les thématiques d'intérêt citées précédemment aux progrès diagnostiques et thérapeutiques.

A l'instar de ce qui a été réalisé pour les antiviraux dans l'hépatite C, l'ambition est que la cohorte HEPATHER devienne l'instrument d'un haut niveau de qualité scientifique qui guide les décisions publiques et l'évaluation coût-efficacité des traitements de la fibrose du foie, de la stéato-hépatite

métabolique, de l'hépatite B, du cancer du foie, avec des analyses sociales et économiques, souhaitées au départ mais non réalisées, à reprendre et développer.

3 Principaux résultats et retombées en termes de recherche et de santé publique

Les six premières années de la cohorte (2012-2018) ont permis de répondre aux principales questions scientifiques et de santé publique relatives aux AAD (efficacité clinique, sécurité des traitements à court et moyen terme). Les questions non résolues autour de l'hépatite C en 2018 concernent la tolérance à plus long terme de l'utilisation des AAD, l'organisation du suivi médical des patients guéris, du dépistage des complications (cancer essentiellement) et de la prise en charge des comorbidités (stéato-hépatite, consommation d'alcool) et leur évaluation coût-efficacité. Ces questions pourront être en partie étudiées grâce au chaînage des données de la cohorte avec le SNDS attendu en 2019.

Les données de la cohorte ont permis de publier 9 articles dans des revues scientifiques et de soumettre des résumés scientifiques au congrès nationaux et internationaux. La cohorte a notamment montré l'effet bénéfique des traitements antiviraux à action directe (AAD) contre le VHC sur la survie des patients et la survenue d'hépatocarcinome ; ces résultats ont été publiés dans la revue *Lancet* en février 2019.

Le projet est inséré dans des collaborations européenne et internationale entre des cohortes ayant pour objet de suivre les traitements respectivement de l'hépatite B et de l'hépatite C.

En termes de santé publique, les résultats de la cohorte sur l'évaluation « en vie réelle » (accessibilité, efficacité et sécurité) des traitements AAD ont été importants pour les autorités sanitaires (DGS, HAS et ANSM) et leur définition d'une politique efficace de lutte contre les hépatites infectieuses. Ce fut en particulier le cas pour la décision de recommander le traitement universel (janvier 2017).

Les études sur les disparités sociales observées en matière d'hépatites et leurs conséquences n'ont toutefois pas été une priorité de ces premières années de mise en œuvre et pourraient être relancées à l'avenir.

4 Gouvernance du dispositif

La cohorte HEPATHER est placée sous la responsabilité directe de l'ANRS, agence autonome de l'Inserm dédiée à la recherche sur la recherche et les hépatites virales. Elle associe l'INSERM, INSERM-Transfert et deux universités : Sorbonne Université and l'Université de Paris.

La gouvernance du projet, assez complexe, comprend :

- le comité exécutif : il s'assure du bon déroulement au quotidien de la cohorte, et notamment la mise en place des procédures (documents types de soumission des PRS, charte d'accès), le suivi de la réalisation des objectifs de la cohorte, et propositions d'actions. Il est composé des responsables scientifiques du projet, de la chef de projet au centre de méthodologie et de gestion, du promoteur et d'Inserm Transfert ; il se réunit une à deux fois par mois depuis le début de la cohorte ;

- le conseil scientifique : composé de chercheurs spécialistes des principales thématiques abordées et impliqués dans des projets de recherche au sein de la cohorte, son rôle est de conseiller l'équipe en charge d'HEPATHER concernant les aspects scientifiques et d'aider à l'établissement de liens avec la communauté scientifique française et internationale ; il se réunit 4 fois par an depuis le début de la cohorte ; ce comité associe, à titre de partenaires non votants, des représentants de chaque partenaire industriel ;
- le comité des partenaires : organe associant les partenaires publics et, en session plénière, privés a pour mission de veiller au bon fonctionnement de la cohorte ; il permet d'assurer la direction stratégique de la cohorte, en cohérence avec les autres organes de la gouvernance. Il s'est réuni une fois en 2018 ;
- le comité consultatif des institutions : il rassemble des représentants des institutions, notamment l'AFEF, l'ANSM, la HAS, la DGS ; ces institutions ont été consultées séparément par les responsables scientifiques et l'ANRS, afin d'établir un lien institutionnel dans le cadre de la cohorte ; il n'y a pas eu, à ce jour, de réunion physique de ce comité ; toutefois, l'ANRS entretenant une relation suivie avec chacun d'eux, l'état d'avancement de la cohorte a été adressé à chacun des membres de ce comité.
- le comité consultatif des partenaires industriels, qui rassemble ces partenaires ;
- le comité de déontologie : il a été constitué en 2017 et a pour mission la validation des procédures mise en place dans la cohorte, notamment la gestion des liens d'intérêt.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

En 2012, l'ANRS a assuré le lancement de la cohorte par son financement propre, en particulier pour prendre en charge les dépenses de personnel. La subvention de l'ANRS est prévue jusqu'en 2022 (à fin 2018, la contribution de l'agence atteint 6,03 M€). À cet égard, l'année 2017 a été marquée par un surcoût de 0,6M€ lié aux centres cliniques hospitaliers.

Le financement de l'ANR au titre du PIA s'est élevé à 10,01 M€; quant aux industriels de la pharmacie, ils ont financé la cohorte à hauteur de 12,50 M€.

5.2 Les partenariats industriels

Un partenariat public/privé a été mis en place afin de proposer un accès aux ressources de la cohorte à la communauté scientifique privée. Le partenariat industriel est encadré par une convention, avec engagement financier de ces partenaires, pour la participation à la constitution et la maintenance de la cohorte. Il a été conçu dans un cadre qui se veut transparent et désireux de respecter l'indépendance scientifique de la cohorte. Les partenaires industriels ont accès aux données et aux échantillons HEPATHER *via* des projets de recherche spécifiques (PRS) faisant l'objet de contrats et financements spécifiques, ainsi qu'à des tableaux de bord.

Deux niveaux financiers de partenariat ont été définis :

- *Platinum* : 500 000 €/an ; ce niveau d'engagement permet un accès gratuit aux échantillons et aux tableaux de bord sur leurs produits ;
- *Gold* : 150 000 €/an ; l'accès aux échantillons est payant.

Le partenariat a été engagé à partir de fin 2012 pour 5 ans pour 6 partenaires (MSD, Gilead, BMS, Roche, Janssen Cilag, Janssen Therapeutics). Abbvie est entré dans le partenariat en 2014 pour 5 ans. Fin 2017, 3 partenaires n'ont pas souhaité renouveler leur engagement : BMS, Janssen et Roche. En 2018, 3 PRS industriels ont été réalisés, et 3 PRS sont en cours. Au total des apports des partenaires se sont élevés à 12, 50 M€.

5.3 Perspectives budgétaires pour la période 2020-2024

Depuis 2018, le nombre des partenaires industriels a décliné : BMS, JANSSEN et ROCHE ne participant ainsi plus au financement de la cohorte. GILEAD et ABBVIE sont engagés par des contrats pluri-annuels jusqu'à fin 2019 en tant que partenaires « *platinum* », et MSD jusqu'à 2020 en tant que partenaire « *gold* ». À partir de 2020, les industriels se verront offrir la possibilité de renouveler leurs partenariats avec davantage de flexibilité.

Dans ce contexte, la cohorte vise aussi à attirer de nouveaux financements. Ainsi, le modèle économique a été adapté afin que les industriels puissent désormais réaliser de simples projets sans pour autant s'engager à devenir des partenaires.

Le dossier transmis dans le cadre de l'évaluation finale de septembre 2019 portait une demande de financement complémentaire de 2,9 M€. Le budget prévisionnel cumulé global sur la période 2020-2024 associé à cette demande s'élevait à 13,45 M€.

Comme suite à l'évaluation finale, le comité de pilotage chargé d'instruire cette demande a proposé en novembre 2019 un complément de financement du projet à hauteur de 1,50 M€ jusqu'au 30 juin 2025.

6 Accès aux données et éthique

La cohorte est conduite en accord avec les règles posées par la Déclaration d'Helsinki et par la législation nationale en matière de recherche biomédicale. Elle a été approuvée par le "CPP Ile de France 3", le Comité national d'éthique, l'ANSM et la CNIL. Le consentement écrit de chaque patient est recueilli préalablement à son inclusion.

L'accès aux données et/ou échantillons de la cohorte est défini par une charte qui a été validée par le comité des partenariats en 2015. Il est réalisé à travers la soumission d'un projet de recherche spécifique qui, en fonction de sa nature, est examiné au plan scientifique et méthodologique et évalué par le comité exécutif et/ou le comité scientifique de la cohorte.

7 Appréciations portées par la mission

La cohorte HEPATHER présente un grand intérêt scientifique, concernant une pathologie (l'infection par le VHC) grave et dont la prévalence est notable dans notre pays. L'équipe responsable est d'un haut niveau et reconnue comme telle, y compris au niveau international.

La méthodologie mise en œuvre est rigoureuse et de qualité. Les résultats obtenus sont satisfaisants et le potentiel ultérieur de recherche dans le domaine des hépatites reste substantiel.

La cohorte bénéficie en outre du fait d'être directement abritée par l'ANRS, qui lui fournit un appui juridique, informatique, et logistique important, et du fait d'être très reconnue de la part des autorités sanitaires, qu'il s'agisse de l'ANSM ou de la HAS, qui a demandé aux industriels de l'utiliser pour présenter des évaluations de leurs traitements en vie réelle.

Un certain nombre de faiblesses existent néanmoins. En particulier, le nombre limité de recherches dans le champ de l'hépatite B, même si le choix de centrer les recherches initiales sur le VHC et les nouveaux traitements était justifié par le contexte thérapeutique à l'initiative du projet. Depuis, un groupe de travail pour l'étude de l'hépatite B a été mis en place. Les études sur les facteurs socio-économiques associés à l'hépatite, prévues dans le projet initial mais non réalisées, méritent aussi d'être relancées.

Par ailleurs, la place importante occupée par les partenaires industriels dans les comités de pilotage et scientifique du projet peut interroger, alors même que le nombre important de ces partenaires industriels et le montant élevé de leurs contributions au budget de la cohorte est aujourd'hui devenu, une fois la question de l'évaluation du traitement de l'hépatite C levée, un facteur de dépendance financière et de fragilité.

LE PROJET HOPE-EPI

1 Présentation générale du dispositif

Selon les données de l'Institut Curie, 1 700 enfants de moins de 15 ans et 800 adolescents de 15 à 18 ans sont diagnostiqués d'un cancer chaque année en France. Il s'agit de la deuxième cause de mortalité des moins de 15 ans après les accidents. Bien qu'ils ne représentent que 1 à 2 % de l'ensemble des cancers recensés tous les ans, les cancers pédiatriques présentent des spécificités puisqu'ils affectent généralement des tissus embryonnaires ou en croissance. Les cancers de l'enfant ou de l'adolescent sont ainsi souvent assimilés à des maladies rares qui exigent des stratégies thérapeutiques adaptées.

Les leucémies, les tumeurs du système nerveux central et les lymphomes constituent les trois principales pathologies cancéreuses affectant les tissus embryonnaires ou en croissance des mineurs.

Les cancers de l'enfant et de l'adolescent constituent donc un enjeu de santé publique très important identifié comme une priorité par le troisième Plan cancer (2014-2019) dont deux actions appellent d'une part à structurer le suivi à long terme des enfants et adolescents et d'autre part à améliorer la prise en compte des séquelles physiques des traitements des cancers.

Le projet HOPE-EPI s'inscrit dans ce contexte.

1.1 Objectifs

- Construire une plateforme nationale dédiée à la recherche en oncologie pédiatrique, adossée au registre national des cancers de l'enfant - RNCE (résultant de la fusion des bases de données du registre national des hémopathies malignes de l'enfant et du registre national des tumeurs solides de l'enfant), en rassemblant, afin de les mutualiser, d'en améliorer la qualité, de les enrichir et d'en faciliter l'accès aux chercheurs, un vaste ensemble de données de cancérologie pédiatrique, en les adossant à des cohortes permettant un suivi sur longue période plus approfondi des enfants atteints.
- Alimenter à partir de cet ensemble de données la recherche en oncologie pédiatrique prioritairement sur les thématiques suivantes : les facteurs de risque génétiques et environnementaux des cancers de l'enfant, les effets iatrogéniques des traitements, les disparités dans le traitement des cancers de l'enfant et leurs déterminants, l'évolution à court, moyen et long terme de l'état de santé des personnes ayant souffert d'un cancer pendant l'enfance ou l'adolescence.

Le projet HOPE-EPI avait en 2010 pour ambition très large d'intégrer à la nouvelle plateforme les cohortes préexistantes FCCSS (*French childhood cancer survivor study* – cohorte d'adultes traités avant 2000 pour un cancer de l'enfant, de type lymphome ou tumeur solide, qui comptait en 2018 plus de 16 000 patients inclus) et LEA (*cohorte prospective multicentrique des leucémies de l'enfant et de l'adolescent* – cohorte constituée à partir de 2004 de patients atteints de leucémie aigüe lymphoblastique âgés de moins de 18 ans au moment du diagnostic et diagnostiqués à partir de janvier 1980, qui comptait en 2018 plus de 4 000 patients inclus) et de les fusionner avec une

nouvelle cohorte COHOPER (*COHort of the PEdiatric cancer Registries*) pour constituer une cohorte unique.

Toutefois l'évaluation à moyen terme (2017) du projet HOPE-EPI a constaté l'échec de l'intégration des deux cohortes préexistantes et de la nouvelle cohorte au sein du même projet, sur fond de divergences d'orientations scientifiques entre équipes, et demandé à l'équipe en charge du projet HOPE-EPI de redéfinir son ambition, le versement des financements du PIA étant suspendu dans l'attente de la définition de cette réorientation. Celle-ci a proposé de recentrer le projet HOPE-EPI sur un cœur de données et d'instruments (*cf.* point 1.2 ci-après) excluant les cohortes FCCSS et LEA, proposition qu'a validée en 2018 le secrétariat général pour l'investissement (SGPI), en décidant dans ce cadre le reversement à l'ANR des fonds versés au titre du PIA pour ces deux cohortes.

A la mi 2019, le projet HOPE-EPI est donc recentré sur un projet de moindre ampleur que prévu en 2010, rebaptisé «plateforme d'observation des cancers de l'enfant» (CCOP)³², dans un contexte qui reste cependant conflictuel entre les participants restant parties prenantes du projet et les porteurs des cohortes n'y figurant plus³³.

1.2 Principales caractéristiques du projet

La plateforme HOPE-EPI (*Recherche Epidémiologique en Hémato-Oncologie Pédiatrique*) intègre, à partir des données collectées dans tout ou partie des 32 centres d'oncopédiatrie, les composantes suivantes qui sont rendues interopérables avec les bases de données du RNCE :

- le programme GEOCAP, qui consiste à géocoder, en vue de recherches sur les facteurs de risques environnementaux et les disparités territoriales des cancers de l'enfant, le dossier des patients inclus dans la plateforme d'observation (avec identification de l'adresse au diagnostic et, pour une partie des dossiers, de l'adresse à la naissance) ; les dossiers géocodés font l'objet d'un échantillonnage annuel afin d'être représentatifs de l'ensemble des naissances en France ;
- un second programme consiste en la description de l'histoire détaillée des différents types de traitements reçus par les patients (dosages de médicaments et de chimiothérapies, interventions chirurgicales, protocoles et modalités des radiothérapies, greffes de moelle osseuse) ;
- le programme PEDIART est une base de données qui recense les doses et la fréquence des radiothérapies reçues dans le cadre de leur traitement par des patients diagnostiqués depuis 2013, ainsi que notamment les protocoles techniques suivis pour leur administration, les organes concernés par le traitement et les organes à risque exposés aux radiations ;
- la base de données BIOCAP est une « biobanque virtuelle » qui centralise les informations codées et cryptées relatives aux échantillons tumoraux et non tumoraux collectés et stockés depuis 2011 (2009 pour l'Ile de France) dans les biobanques des hôpitaux dans lesquels sont traités les patients inclus ;
- la cohorte COHOPER, qui concerne le suivi et la surveillance épidémiologique de tous les cas de cancer de l'enfant diagnostiqués depuis 2000 identifiés dans le RNCE (environ 30 000 cas, dont 25 000 survivants), a pour objectifs de suivre, surveiller et évaluer leurs risques de rechutes, de séquelles et de second cancer, sur la base d'un suivi systématique à partir des

³² Dans la suite de la présente fiche la dénomination « HOPE-EPI » est cependant maintenue

³³ Cette situation de conflit se traduit notamment par la suspension en 2018, toujours en vigueur à mi 2019, de l'alimentation par un ensemble de centres de soins en oncopédiatrie du RNCE et de certains modules de la plateforme

dossiers médicaux et des données médico-administratives, et plus généralement l'évolution de leur état de santé, sur la base de questionnaires.

2 Caractéristiques du projet

2.1 Avancement du projet

Les différentes composantes de la plateforme connaissent des états d'avancement variables à mi - 2019, les objectifs de centralisation de données portés par chacun des modules n'étant à cette date pas complètement atteints et étant fragilisés par le refus de certains centres de soins en oncopédiatrie de transmettre les données correspondantes (concernant PEDIART et BIOCAP notamment) :

- le programme GEOCAP a permis de recueillir l'adresse au diagnostic pour 97 % des dossiers inclus dans la plateforme et les caractéristiques géographiques détaillées associées (GIS), décrivant notamment la proximité avec diverses sources de risque (installations nucléaires, lignes à haute tension, axes à trafic routier très important, etc...), sont documentées pour 19 500 dossiers (60 000 dossiers miroirs) ;
- les données d'historique des traitements suivis pendant cinq années après le diagnostic sont complètes pour 12 700 des 20 900 cas de cancer primaire de l'enfant diagnostiqués entre 2000 et 2011 ;
- la base de données PEDIART, alimentée par les 19 centres d'oncopédiatrie dotés d'une infrastructure technique sécurisée pour l'échange de données avec la plateforme HOPE-EPI, comporte 1 700 dossiers de radiothérapie de patients diagnostiqués depuis 2013 ;
- la base BIOCAP centralise les données de près de 20 000 échantillons stockés dans les biobanques des centres d'oncologie pédiatrique correspondant à 4 944 cas de patients inclus dans la plateforme entre 2009 et 2016 ;
- le suivi des 25 000 patients de la cohorte COHOPER repose sur le recueil de données supplémentaires à celles recueillies en routine dans le cadre du RNCE : les causes de décès, à partir du Capi-DC (appariement réalisé), les cas de rechute et de second cancer avant l'âge de 18 ans, à partir d'un recueil spécifique d'information par enquête cinq ans après le diagnostic, les cas de rechute, second cancer ou complications, dans le cadre d'un appariement de la base du RNCE au SNDS (la demande d'appariement individuel est en cours : le dossier est déposé à la CNIL). S'y ajoutent l'état de santé, le mode de vie et la qualité de vie à l'âge adulte, à partir de questionnaires adressés aux patients volontaires dont le lancement n'a pas encore démarré (l'administration de ces questionnaires nécessite une autorisation qui est en cours d'instruction). En parallèle, une collaboration est développée avec l'Inca en matière d'exploitation et d'analyse des données issues de l'appariement au SNDS de la plateforme qui recense l'ensemble des personnes atteintes d'un cancer identifiées depuis 2010.

2.2 Articulation avec d'autres cohortes ou dispositifs d'information

Les différents modules de la plateforme HOPE-EPI sont interconnectés avec le registre national des cancers de l'enfant.

La plateforme HOPE-EPI collabore avec la cohorte Cancer de l'Inca.

La cohorte CONSTANCES doit être utilisée comme cohorte témoin pour l'analyse des réponses aux futurs questionnaires d'état de santé, mode de vie et qualité de vie (à cette fin, les questionnaires prévus dans le cadre de la cohorte COHOPER ont été harmonisés avec ceux de CONSTANCES).

2.3 Horizon et perspectives envisagés pour le dispositif après 2020

La période qui s'ouvre à compter de 2020 doit permettre la consolidation des différents modules de la cohorte, le suivi effectif des patients survivants inclus dans la cohorte COHOPER et l'exploitation des données qui seront produites notamment dans le cadre de son appariement au SNDS.

3 Principaux résultats et retombées en termes de recherche et de santé publique

La production scientifique reposant sur les données produites par les différentes composantes de la plateforme reste à mi-2019 relativement limitée : onze publications sont ainsi recensées.

La plupart des publications (neuf) s'appuient sur des données géocodées produites par GEOCAP et portent sur les liens entre expositions environnementales et risque de cancer de l'enfant : notamment les expositions au trafic routier et à la pollution de l'air (accroissement du risque de leucémie myeloïde aigüe), aux radiations UV (établissement d'un lien entre une exposition à des radiations de haute intensité et du risque de survenue d'une leucémie myeloïde aigüe), aux radiations émises par les lignes à haute tension (accroissement du risque de leucémie de l'enfant), aux installations nucléaires ; deux autres publications utilisent les données fournies par PEDIART.

Un premier projet de publication décrivant l'infrastructure a été soumis en 2019 à relecture ; la publication de cette présentation permettra d'en accroître la visibilité internationale.

4 Gouvernance du dispositif

Le projet HOPE-EPI fait l'objet d'un consortium (accord signé en 2016, soit cinq années après le lancement du projet) réunissant à l'origine l'université Paris-Descartes (Université de Paris), établissement coordinateur, l'Inserm, le centre Léon Bérard (Unicancer), l'Université Paris-Sud, l'Aix-Marseille Université et l'Assistance publique – Hôpitaux de Marseille. Toutefois suite au conflit entre équipes qui s'est traduit par le retrait du projet des porteurs des cohortes FCCSS et LEA, si les trois derniers établissements cités sont toujours formellement membres du consortium, les composantes d'HOPE-EPI dont leurs équipes avaient la responsabilité n'appartiennent plus au projet tel qu'il a été reconfiguré sur la base de l'évaluation intermédiaire de 2017-2018.

Il est porté par l'équipe « Epidémiologie des cancers de l'enfant et de l'adolescent », une des neuf équipes du centre de recherche en épidémiologie et statistiques (CRESS), unité mixte de recherche Inserm – Université de Paris.

HOPE-EPI dispose d'une gouvernance à trois niveaux, largement renouvelée à l'issue de l'évaluation intermédiaire qui avait pointé l'échec de l'organisation précédente puis du recentrage du projet, reposant sur :

- une instance opérationnelle : le comité de pilotage, qui réunit mensuellement les responsables des différentes composantes de la plateforme et règle l'ensemble des questions liées au fonctionnement et à l'avancement du projet et les relations de celui-ci avec le registre national ;
- une instance décisionnelle : le « comité des établissements », dans lequel sont représentés les établissements signataires de l'accord de consortium, se réunit une fois par an et statue sur les orientations stratégiques du projet, son budget, son rapport d'activité préparé par le comité de pilotage et sur toute décision importante relative à l'organisation du projet ;
- s'y ajoute un « comité directeur » (*steering comitee*) dans lequel siègent des scientifiques représentant les spécialités et disciplines nécessaires au pilotage scientifique du projet (épidémiologie, radiothérapie, oncologie pédiatrique, pathologie, biologie), dont trois sont désignés par la Société française d'oncologie pédiatrique, et auxquels doivent être ajoutés quatre représentants respectivement d'associations de parents d'enfants malades, de L'Institut de santé publique de l'INSERM, de L'INCA et de Santé publique France ; ce comité se réunit annuellement pour examiner et valider la stratégie du projet.

En l'état, ces instances ne comportent pas le conseil scientifique dont sont dotées les autres cohortes soutenues par le PIA.

Le support et le soutien à l'activité de l'équipe HOPE-EPI (aspects réglementaires et financiers) sont fournis la direction régionale de l'INSERM compétente.

5 Ressources, budgets et financements

5.1 Sur la période 2011-2019

Selon les données communiquées par l'agence nationale de la recherche (ANR), les recettes du projet HOPE-EPI se sont globalement élevées sur la période 2012-2019 à 12,16 M€ dont 6,27 M€ au titre du financement du PIA, étant souligné que le versement des financements au titre du PIA a été suspendu comme suite aux difficultés rencontrées par le projet et dans l'attente de la stabilisation de son nouveau format issu des décisions prises sur la base des constats de l'évaluation intermédiaire (cf. points 1-1 et 1-2 supra).

L'équipe HOPE-EPI compte dix-sept membres relevant de différents employeurs parties prenantes du projet (l'INSERM, l'université Paris-Descartes, le CHU de Nancy, l'AP-HP). Neuf bénéficient d'un emploi permanent (statutaire ou CDI) et huit sont en contrat à durée déterminée.

Outre le financement au titre du PIA (6,27 M€), le projet HOPE-EPI a perçu de 2012 à 2018 les financements suivants (en M€) :

Institut national du cancer	Fondations et associations caritatives ³⁴	Canceropôle Ile de France	Autres institutions publiques ³⁵	Entreprises privées	Financements européens	Total
2,57	1,09	0,74	0,88	0,19	0,36	5,88

5.2 Perspectives budgétaires pour la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale de septembre 2019 portait une demande de financement complémentaire de 3,10 M€, dont 2,29 M€ de dépenses de personnel (74 %) correspondant à 50 % du coût de l'équipe HOPE-EPI, afin de consolider les différentes composantes de la plateforme et mettre en œuvre l'appariement au SNDS. Le projet de budget prévisionnel cumulé sur la période 2020-2024 associé à cette demande s'élevait à 6,30 M€ correspondant pour 60 % de ce montant à des dépenses de personnel.

Comme suite à l'évaluation finale de septembre 2019, le comité de pilotage chargé d'instruire cette demande a proposé en novembre 2019 un complément de financement du projet de 0,9 M€ jusqu'au 30 juin 2025, couvrant ainsi 30 % de la demande.

5.3 Les partenariats industriels

Si la plateforme a reçu des financements importants de la part des fondations ou associations engagées dans la lutte contre le cancer (la Ligue nationale contre le cancer, la Fondation pour la recherche sur le cancer, Enfants et Santé), qui ont contribué au financement de plusieurs de ses composantes (notamment BIOCAP et la cohorte COHOPER), ses données n'ont à ce stade généré qu'un seul partenariat industriel ponctuel en 2018 avec la société Novartis (dans le cadre de la constitution et du dépôt à la HAS du dossier d'évaluation d'une nouvelle thérapie génique – Car-t Cells – destinée au traitement des leucémies aiguës lymphoblastiques à cellules précurseurs B réfractaires de l'enfant)

6 Accès aux données

Il est prévu la mise en place d'appels à projets de recherche scientifique par l'équipe HOPE-EPI utilisant les données de la plateforme, telles que définies dans le catalogue des données de l'infrastructure ; les projets seront soumis à une procédure d'examen par les instances de gouvernance d'HOPE-EPI, en deux phases faisant intervenir, pour la première, l'équipe projet et le comité directeur appelés à se prononcer sur la base d'une lettre d'intentions sur la faisabilité du projet de recherche et, pour la seconde, le comité directeur et un conseil scientifique, qui reste à créer, qui évalueront la pertinence scientifique du projet sur la base d'un dossier complet.

Toutefois, compte tenu du retard important pris pour la mise en place des différentes composantes de la plateforme, aucun appel à projets n'a à ce jour été publié.

³⁴ Ligue nationale contre le cancer, ARC, Enfants et Santé

³⁵ Anses, Invs, Iresp, ministère chargé de la santé

7 Appréciations de la mission

Les désaccords entre les partenaires du projet d'origine, qui ont conduit à l'abandon du projet d'intégration initialement prévu des trois cohortes en oncologie pédiatrique et à la suspension des financements du PIA aux deux cohortes préexistantes, perdurent à la mi -2019 et fragilisent le déploiement du projet y compris dans son format recentré sur une plateforme d'observation des cancers de l'enfant (CCOP).

Le non fonctionnement du registre national des cancers de l'enfant depuis 2018 lié au blocage de certains cliniciens en onco-pédiatrie refusant de l'alimenter nécessite que les divergences qui subsistent entre les parties prenantes fassent l'objet d'une action particulière des pouvoirs publics et que la réforme de la gouvernance du projet mise en place progressivement depuis 2018 continue à être fortement accompagnée, notamment par l'Inca. Le projet de cohorte COHOPER paraît encore peu avancé compte tenu des difficultés de mise en œuvre de l'infrastructure de base conçue à partir des registres, notamment pour la partie comportant des questionnaires auprès des patients survivants.

Le développement de la plateforme, qui paraît nécessiter un effort de relativement long terme compte tenu des retards pris, devrait pouvoir s'appuyer, compte tenu de la prolongation des financements du PIA proposée sur la période 2020 – 2025 à un niveau assez éloigné de la demande de financement permettant la consolidation du projet, sur les services mutualisés portés par l'Inserm dans le cadre du projet d'unité mixte de services France Cohortes.

L'avenir des deux autres cohortes (exploitation de FCSS et poursuite de LEA) dont les apports sont aussi intéressants et utiles est d'un autre côté mal assuré, et mériterait aussi d'être clarifié, tout en évitant que les différents projets (CCOP, nouvelles inclusions de LEA et CRYOSTEM) puissent doubler auprès des mêmes familles ayant eu à subir le cancer d'un de leurs enfants et en continuant à rechercher, même sans unification, les articulations indispensables.

LA COHORTE MEMENTO

1 Présentation générale du dispositif

La maladie d'Alzheimer et les maladies apparentées touchent, d'après la Fondation Alzheimer, environ 900 000 personnes en France. Chaque année, 200 000 nouveaux cas sont diagnostiqués et ce nombre pourrait à l'avenir augmenter fortement avec le vieillissement de la population. En dépit de tentatives répétées, aucune thérapie n'a pu être mise au point de façon efficace contre la maladie, sachant que, pathologie d'évolution lente avec des signes qui peuvent exister de façon précoce, elle présente des caractéristiques et des évolutions hétérogènes, dont la diversité et les déterminants n'ont pas encore été complètement appréhendés.

1.1 Objectifs

- Faire progresser, dans le cadre de la mise en œuvre du plan Alzheimer, la connaissance l'histoire naturelle de cette maladie en analysant les facteurs conduisant à son développement ou à celui de démences connexes chez les personnes souffrant de plaintes cognitives ou de déficits cognitifs légers
- Disposer d'une plateforme de recherche permettant des analyses multidimensionnelles de ces facteurs de risque au regard de la santé cérébrale des personnes, et incluant des évaluations sociodémographiques, neuropsychologiques, des co-morbidités associées, ainsi que des marqueurs de neuro-imagerie, génétiques et sanguins.

1.2 Principales caractéristiques de la cohorte

- Cohorte spécialisée suivant 2323 patients inclus entre 2011 et 2014 dans 26 centres « mémoire, ressources et recherche » (CMRR)
- Constitution d'une base de données réunissant résultats d'examens cliniques, prélèvements sanguins destinés à des analyses biologiques et génétiques, et données de neuro-imagerie

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

Après une phase pilote conduite dans cinq centres, l'inclusion des patients a été réalisée entre avril 2011 et juin 2014, 2 323 des 2449 participants ciblés ayant accepté de participer à l'étude. La cohorte est fermée depuis cette date.

Le suivi est réalisé tous les six mois, et avait été prévu initialement sur une période cinq ans suivant la dernière inclusion.

66% des patients non décédés ont été suivis à 5 ans, 350 d'entre eux ayant développé une démence à l'issue de cette période.

2.2 Modalités d'inclusion

Les patients ont été recrutés, sur la base du consentement, dans 26 des 28 centres « mémoire, ressources et recherche » (CMRR) institués dans le cadre des plans Alzheimer, centres qu'ils consultent soit de façon spontanée, soit sur le conseil de leur médecin traitant ou de centres de consultation mémoire.

Les critères d'inclusion retenus ont visé à identifier des patients sujets à des troubles de la mémoire à un stade précoce et éventuellement susceptibles de développer ensuite une maladie d'Alzheimer ou une démence connexe.

Ils étaient ciblés sur deux catégories de personnes, non démentes à l'entrée dans l'étude :

- celles émettant « une plainte cognitive subjective » et âgées de 60 ans ou plus ;
- celles ayant un déficit cognitif léger (au seuil de la moyenne moins un écart type) détecté depuis moins d'un an au moins dans un domaine cognitif : mémoire, langage, *praxis*, capacités visuo-spatiales, fonction exécutives.

En l'absence d'échantillonnage, les biais de recrutement ont été analysés en comparant les participants à la cohorte MEMENTO à la Banque nationale Alzheimer (BNA) d'une part, aux données de l'Insee issues du recensement d'autre part.

Le suivi des patients est réalisé tous les six mois jusqu'à décembre 2019, et a été poursuivi à 5 ans pour 66% des participants inclus non décédés.

Les responsables de la cohorte ont pris des dispositions pour limiter l'attrition (suivis intermédiaires par téléphone, questionnaires *ad hoc* via un informant ou un médecin traitant...), mais les taux d'attrition ainsi que les décès s'accroissent au fil du temps, ce qui peut poser problème.

2.3 Contenu, collecte et stockage des différents types d'informations

Les informations recueillies lors des examens cliniques concernent les caractéristiques sociodémographiques des participants, leur histoire personnelle et familiale, leurs modes de vie (tabac, alcool, activité physique...), leurs médicaments ou traitements, les résultats de leurs examens physiques et neurologiques ainsi que d'une batterie de tests neuropsychologiques, et leurs plaintes subjectives cognitives ou non cognitives. Ces données donnent lieu à une saisie électronique et sont centralisées dans une base de données administrée à partir du centre d'investigation clinique de Bordeaux, plateforme commune au CHU de Bordeaux, à l'Inserm et à l'Université de Bordeaux.

Des données d'imagerie sont recueillies par la voie d'IRM et de PET scans, sous l'égide du Centre d'acquisition et de traitement des images (CATI) pour la maladie d'Alzheimer basé à Saclay, et financé par la Fondation Alzheimer.

Les prélèvements biologiques permettent de recueillir du liquide céphalo-rachidien et des échantillons sanguins pour des mesures courantes (glycémie, lipidémie...) ou des analyses génétiques (ADN).

Ces prélèvements sont pour moitié conservés à l'Institut Pasteur de Lille et pour l'autre moitié, à titre de « biobanque miroir », par le centre de ressources biologiques (CRB) du CHU de Bordeaux.

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

La cohorte s'inscrit dans le cadre de l'animation scientifique des 28 CMRR implantés dans les centres hospitalo-universitaires suite aux plans Alzheimer.

Elle n'est en revanche pas directement liée à la banque nationale Alzheimer (BNA) créée lors du plan Alzheimer de 2008 et qui rassemble les données des centres de consultation mémoire (plus d'un million de sujets suivis à partir de l'activité de ces centres), sachant toutefois que celle-ci connaît des problèmes d'exhaustivité et de qualité.

Memento n'a pas encore non plus été appariée avec le SNDS, ni mise en perspective des utilisations potentielles de la cohorte CONSTANCES, qui vise à proposer un examen neurologique et cognitif de façon systématique après 45 ans et à suivre les différents aspects de la « fragilité » à partir de 50 ans.

Un appariement est toutefois envisagé à horizon rapide avec le SNDS, une demande en ce sens étant en cours pour un appariement de type probabiliste.

2.5 Horizon et perspectives envisagés pour le dispositif après 2020

Les inclusions ayant été achevées en 2014, le suivi initialement annoncé était appelé à s'achever à la fin 2019.

Seuls 350 patients ayant à ce stade développé une démence, les responsables de la cohorte souhaitent prolonger ce suivi pendant dix années supplémentaires jusqu'à 2030 sur la base d'un examen clinique annuel et d'un appariement avec le SNDS, pour tirer parti sur longue période des potentialités offertes par les données collectées.

Ils envisagent aussi une collecte, sur la base de dons, des cerveaux des personnes décédées en vue d'études anatomiques, génétiques et moléculaires.

Les taux d'attrition s'accroissant toutefois avec le temps, les responsables ont par ailleurs émis l'idée d'une réouverture de la cohorte à 600 nouveaux participants dans 10 CMRR entre 2020 et 2022, auxquels seraient proposés de nouveaux examens (neuro-imagerie, électroencéphalogrammes) non prévus dans le projet initial.

3 Principaux résultats et retombées en termes de recherche et de santé publique

La cohorte MEMENTO a pris le relais, pour cette pathologie spécifique et les démences connexes, des cohortes historiques Paquid (3 800 personnes âgées suivies de 1998 à 2018) et Trois Cités (9 300 personnes âgées de 65 ans ou plus villes suivies à partir de 1999), qui ont donné lieu à des suivis sur longue période.

Les responsables de la cohorte soulignent les résultats d'ores et déjà obtenus concernant la définition biologique de la maladie d'Alzheimer, l'association entre hypertension et démence *via* les anomalies cérébrovasculaires et la neurodégénération (communications au congrès mondial de juillet 2019), ainsi que la stratification des participants grâce à de nouvelles mesures obtenues *via* l'IRM.

Ces travaux confirment de leur point de vue l'intérêt de la cohorte MEMENTO à des fins à la fois d'inférence dans la population, de compréhension des mécanismes et de définition de cibles de prévention.

Plus d'une trentaine de projets de recherche ont été déposés et acceptés, à la fois concernant l'analyse statistique de la cohorte et le suivi des patients. La cohorte donne en outre lieu à une série de projets ancillaires, dont certains menés en partenariat avec des industriels (*cf. infra*). On peut notamment citer :

- le projet VASCOD, qui vise à étudier l'impact des facteurs de risques vasculaires et de l'inflammation sur l'évolution cognitive (400 participants dans 10 centres avec un financement PHRC) ;
- le projet AMYGING, qui vise à évaluer les liens entre la charge amyloïde et l'incidence de la démence (contrats partenariaux avec GE *Healthcare* et AMID) ;
- le projet Sommeil, cognition et troubles de la mémoire (SCOAL) conduit au sein du CHU de Bordeaux (avec un financement de l'ANR) ;
- le projet INSIGHT, étude observationnelle conduite au sein de l'IHU ICM à la Pitié-Salpêtrière qui complète les observations par des électroencéphalogrammes et vise à évaluer les associations entre la protéine bêta-amyloïde et divers marqueurs cognitifs ou issus de la neuro-imagerie (349 participants, dont 249 communs avec Memento) ;

Le nombre de publications à partir de la cohorte commence à monter en charge, avec 19 articles acceptés ou publiés et 15 en cours de soumission, mais les publications les plus importantes sont plutôt attendues pour l'avenir.

La cohorte est engagée dans des collaborations internationales, par exemple avec la cohorte Framingham concernant l'analyse de biomarqueurs sanguins, et participe à des consortia internationaux comme l'*European Alzheimer DNA Biobank* (EADB) ou les *Cohorts for Aging Research in Genomics* (CHARGE).

Elle n'a été examinée par le jury d'évaluation du PIA qu'en 2019, dans la mesure où son financement est jusqu'ici passé par une autre voie (Fondation Alzheimer, puis ministère de la recherche *cf. infra*).

À cet égard, les différents interlocuteurs rencontrés par la mission ont souligné l'intérêt de Memento de par la qualité du phénotypage mis en œuvre en termes cliniques, biologiques génétiques et d'imagerie. Ils regrettent néanmoins que cette cohorte n'ait pu être constituée à partir d'un socle d'observation général (registre ou très grande cohorte en population) susceptible de servir de base témoin, ce qui semble la voie d'avenir pour le suivi des maladies neuro-dégénératives pour lesquelles les pouvoirs publics souhaiteraient aussi développer des approches moins cloisonnées.

4 Gouvernance du dispositif

La cohorte MEMENTO repose sur un partenariat entre la Fondation Alzheimer, fondation de coopération scientifique créée en 2008 avec le concours de l'Inserm et du ministère de la recherche pour le financement du Plan Alzheimer, l'Université de Bordeaux et le CHU de Bordeaux. C'est ce dernier qui est le porteur administratif de la cohorte (à travers sa direction de la recherche clinique et de l'innovation).

La cohorte associe également, en tant que partenaires, la Fédération des centres mémoire, ressources et recherche) (FCMRR) et la plateforme d'imagerie CATI (*cf. supra*).

L'ensemble de ces partenaires siège actuellement dans le comité de pilotage institutionnel de la cohorte, dont les réunions se sont espacées et qui doit être prochainement recomposé autour des seuls partenaires financeurs (dont l'Inserm qui porte le poste de l'investigatrice principale) ; les partenaires intervenant dans la mise en œuvre de la cohorte continueront à participer à son équipe de management.

De même, le conseil de « stratégie scientifique » de la cohorte a vu son mandat s'achever en 2016 et n'a pas encore été reconstitué.

La mise en place d'un mécanisme d'évaluation est enfin envisagée à compter de 2020.

5 Ressources, budgets et financements

5.1 Sur la période 2011-2019

Memento a été financée sur fonds publics de 2011 à 2016 principalement par la Fondation Alzheimer, qui y a consacré 15,6 M€ (2,6 M€ pour la mise en place et la coordination de l'étude, 1,6 M€ pour la création de la biobanque, 8,3 M€ pour l'inclusion et le suivi des participants et 3 M€ pour la plateforme de neuro-imagerie).

Depuis 2016, la Fondation Alzheimer ne finance plus la coordination de la cohorte, qui a bénéficié d'une subvention du ministère chargé de la recherche à hauteur de 1,73 M€.

S'y ajoutent, à hauteur d'environ 1,3 M€, les contributions des partenaires institutionnels, à savoir le CHU de Bordeaux, l'Inserm et l'université de Bordeaux, ainsi que 4,3 M€ d'autres financements (PHRC et financements privés -*cf. infra*).

En termes de ressources humaines, l'équipe de projet en charge de la cohorte comportait, outre les deux co-investigatrices (relevant statutairement respectivement de l'Université de Bordeaux et de l'Inserm), une cheffe de projet, un biostatisticien, une *data manager*, trois attachées de recherche cliniques, une coordinatrice administrative et une coordinatrice d'études cliniques.

5.2 Perspectives budgétaires pour la période 2020-2024

La demande budgétaire présentée par la cohorte pour la période de 2020-2024 est de 4,3 M€, pour un budget total évalué à 7,5 M€, dont 0,7 M€ seraient par ailleurs supportés par les partenaires institutionnels et 2,5 M€ par des cofinancements, acquis ou à acquérir.

Ces demandes budgétaires comprennent d'une part la coordination de l'étude et le suivi des participants déjà inclus (à hauteur 3,1 M€) et d'autre part l'inclusion et le suivi de 600 nouveaux participants entre 2020 et 2022 (à hauteur 1,2 M€).

L'équipe de suivi resterait dans ce cadre composé de 6 à 7 personnes (hors investigatrice principale).

Comme suite à l'évaluation de la cohorte, le ministère chargé de l'enseignement supérieur et de la recherche a proposé de financer sur crédits budgétaires une enveloppe complémentaire pour assurer la poursuite du suivi des membres de la cohorte pour cinq années supplémentaires (0,6 M€ pour 2020).

5.3 Les partenariats industriels

La cohorte MEMENTO a bénéficié d'un ensemble de partenariats industriels ciblé sur des projets spécifiques :

- de la part de GE-*Healthcare* et AVID-Lilly, à hauteur de 1,23 M€, pour un projet *Amyging* visant à évaluer la charge amyloïde chez un sous-échantillon de patients ;
- de la part de Fujirebio pour l'analyse d'échantillons à l'aide de test *in vitro* ;
- de la part de Lilly pour un projet *Roadmap* de médecine personnalisée dont memento a rejoint le consortium ;
- de la part de Lundbeck en vue de la mise au point de tests neuropsychologique prédictifs de déclin cognitif ;
- de la part de Qynapse pour l'étude de l'apport de plusieurs marqueurs dans le domaine de la neuro-imagerie.

Ces partenariats ont représenté un apport financier d'environ 2,41 M€ sur la période 2011-2019., et 0,55 M€ sont déjà acquis à ce titre dur la période 2020-2024.

Ils sont négociés par la SATT (société d'accélération du transfert de technologie) régionale, en appui à l'Université de Bordeaux, et non, comme pour d'autres cohortes par Inserm Transfert.

6 Accès aux données

L'accès aux données se fait dans le cadre d'une charte qui fixe des règles relatives aux règles d'exclusivité et aux publications. Elle prévoit que les projets scientifiques soumis reçoivent l'approbation des investigateurs principaux et du comité de coordination scientifique de la cohorte, sur des critères comme la faisabilité, l'originalité scientifique, l'apport potentiel au réseau, mais aussi l'absence de concurrence avec les recherches en cours et de conflits avec les résultats et développements de la cohorte. Ces critères, formulés de façon limitative, pourraient être revus dans le cadre d'une gouvernance renouvelée.

La mise à disposition des données a été initialement envisagée *via* le centre d'accès sécurisé aux données (CASD), mais les tarifs pratiqués par celui-ci pour les chercheurs académiques ont conduit les responsables de la cohorte à choisir la plateforme *Dementia Platform UK* (DP UK), située au Royaume-Uni mais dont l'accès est gratuit, pour leur fournir depuis juin 2019 cet accès sécurisé.

7 Appréciations de la mission

Le démarrage de la cohorte a été envisagé à un moment où des traitements pour la maladie d'Alzheimer étaient espérés à un horizon rapide, alors que l'échec des essais cliniques conduit de nouveau à tenter de mieux caractériser l'histoire naturelle de la maladie et de définir une prévention ciblée par la recherche de biomarqueurs identifiant des « groupes à hauts risques ».

La cohorte MEMENTO, précisément centrée sur cette thématique, est reconnue au plan international et reçoit un soutien affirmé de l'Université et du CHU de Bordeaux, dans le cadre d'une politique de site affirmée en matière de recherche en santé.

Le fait que seule une partie des démences susceptibles d'intervenir au sein de l'échantillon des 2323 participants se soit déclarées à l'issue de cinq ans d'observation semble justifier la poursuite du suivi pendant cinq à dix ans supplémentaires, dans l'objectif de tirer parti des informations très riches et complètes (cliniques, biologiques, génétiques, d'imagerie) collectées par la cohorte.

Cela implique toutefois que, pour compenser une attrition devenant de plus en plus importante, un suivi « passif » soit parallèlement mis en place grâce à un appariement avec le SNDS, par exemple pour retrouver les personnes appelées à déménager en EHPAD

Cela importe d'autant plus que la cohorte, dont les inclusions ont été terminées en 2014, n'a pas forcément vocation à être ré-ouverte, dans la mesure où les responsables de la Fondation Alzheimer et du plan maladies neuro-dégénératives semblent plutôt préconiser pour l'avenir le développement de cohortes à partir de vastes bases de données ou cohortes généralistes en population.

La transparence dans la gouvernance de la cohorte et les conditions d'appropriation des données par des équipes de recherche extérieures au réseau de ses partenaires directs semblent par ailleurs devoir être améliorées, et les modalités, aujourd'hui tout à fait spécifiques, de mise à disposition de ces données (*via* la plateforme DPUK) pourraient être revues dans le cadre des projets France cohortes ou *Health data Hub*.

LA COHORTE OFSEP

1 Présentation générale du dispositif

La sclérose en plaques est une maladie auto-immune chronique du système nerveux central qui évolue par « poussée » inflammatoires imprévisibles, qui s'accompagnent fréquemment de troubles visuels ou de la sensibilité. Souvent invalidante, elle affecte plus de 110 000 personnes en France. Plus de 5000 cas sont diagnostiqués chaque année.

1.1 Objectifs

Trois objectifs principaux :

- enrichir la collecte de données, débutée en 2003 (30 000 patients), des données d'examen d'IRM et de prélèvements biologiques afin de pouvoir décrire de manière régulière la population des patients ;
- conduire des recherches spécifiques, notamment sur des sous-groupes de patients prioritaires : patients atteints de syndromes radiologiquement isolés, patients atteints de syndromes cliniquement isolés, patients atteints de formes progressives d'emblée et patients atteints de neuromyéélite optique de Devic ;
- depuis 2018, développer une cohorte « haute définition » (OFSEP-HD) visant à déterminer les facteurs prédictifs de l'évolution de la maladie.

1.2 Principales caractéristiques de la cohorte

Portée par un consortium réunissant l'université Claude Bernard de Lyon, les hospices civils de Lyon et la fondation Eugène Devic Edmus contre la sclérose en plaques, l'observatoire français de la sclérose en plaques OFSEP regroupe les données sociodémographiques et cliniques standardisées, enrichies, des examens d'IRM et d'échantillons biologiques de 62 700 patients suivis dans les 41 centres Edmus participants au projet.

La fondation Edmus est propriétaire du logiciel éponyme (pour *European Database for Multiple Sclerosis*) conçu en 1990 par le professeur Confavreux reposant sur un langage commun reconnu et partagé par la communauté des neurologues pour décrire la maladie.

OFSEP est à l'origine davantage un registre longitudinal qu'une cohorte de recherche, dont les liens avec la haute autorité de santé (HAS) sont étroits : le comité de transparence de la haute autorité de santé (HAS) recommande ainsi aux laboratoires pharmaceutiques de s'appuyer sur OFSEP pour évaluer le service rendu par un nouveau médicament.

C'est pourquoi le jury d'évaluation de 2017 a recommandé aux responsables du dispositif de développer en son sein une cohorte de petite taille recentrée sur un nombre limité de participants et complétée par un ensemble d'informations beaucoup plus riche, incluant des questionnaires patients. Cette réorientation du projet a été opérée en 2018, par la voie d'une cohorte dite « haute définition » limitée à 1600 participants.

2 Caractéristiques du projet

2.1 Date de démarrage, nombre de participants

- Base de données OFSEP : collecte depuis 2003 - 62 700 patients concernés ;
- Cohorte OFESP-HD : début de la collecte en juillet 2018 – 1 600 patients suivis annuellement depuis 2 ans inclus.

2.2 Contenu, collecte et stockage des différents types d'informations

Les données collectées sont :

- des données descriptives : caractéristiques sociodémographiques, poussées³⁶, handicaps, traitements suivis, événements indésirables, reportées sur une fiche type dite Fiche minimale OFSEP (FMO) et saisies sur la base Edmus de chaque centre ;
- des échantillons biologiques (sérum, urine, LCR, selles) stockés sur une base de données TK® ;
- des IRM anonymisées transférées et stockées sur la base Shanoir (Sharing NeuroImaging Ressources).

Les données issues de Edmus, TK® et Shanoir sont regroupées deux fois par an dans une base de données unique OFSEP. Cette base de données est utilisée pour suivre des cohortes des quatre sous-groupes de patients jugés prioritaires, auquel s'est ajoutée, depuis juin 2018, une nouvelle cohorte, dite OFSEP-HD (pour haute définition), centrée sur 1 600 patients dont l'objet consiste à identifier les facteurs pronostiques de l'évolution de la maladie, et à analyser ses stratégies de prise en charge.

2.3 Articulation avec d'autres cohortes ou dispositifs d'information

Cette base de données n'est pas appariée avec le SNDS : les responsables d'OFSEP envisagent de s'appuyer sur une équipe de recherche de l'EHESP (équipe Repère, EA 7449) pour la méthodologie et sur l'institut de recherche technologique B<COM> pour le stockage des données.

Les données du SNDS ont été par ailleurs mobilisées pour décrire les parcours de soin dans le cadre du projet SEPTYCO portée par l'équipe de l'EHESP.

2.4 Horizon et perspectives envisagés pour le dispositif après 2020

Les porteurs du projet avancent trois axes d'évolutions pour les prochaines années :

- l'extension du suivi à de nouveaux patients, pendant quatre années supplémentaires, dans un contexte de changement rapide des pratiques de soins ;
- le développement d'analyses génétiques pour identifier les déterminants génétiques en comparant des profils de sous-groupes de patients ;

³⁶ Apparition des symptômes neurologiques,

- l'amélioration de l'accès à des IRM de qualité, notamment dans les centres hospitaliers non universitaires.

3 Principaux résultats et retombées en termes de recherche et de santé publique

Parmi les résultats mis en avant par les responsables de la cohorte, on relève principalement la mise en évidence des effets de différents traitements utilisés pour soigner la sclérose en plaques :

- description de la diminution de l'incidence de la leucoencéphalopathie multifocale progressive (affection démyélinisante du système neuronal central) associée à la prise de natalizumab (un anticorps) ;
- description de l'efficacité à long terme des traitements modificateurs de l'évolution de la maladie sur l'accumulation de l'invalidité ;
- évaluation de la surmortalité liée à la sclérose en plaques ;
- évaluation de la survie associée à l'exposition aux interférons bêtas (des glycoprotéines).

79 projets de recherche ont été conduits dans le cadre d'OFSEP, 59 projets (74 %) au sein du réseau et 15 (26 %) par des partenaires extérieurs.

OFSEP est en outre membre du consortium international BigMSdata, qui regroupe notamment les registres nationaux danois, italiens et suédois.

4 Gouvernance du dispositif

Remaniée à la suite de l'évaluation à mi-parcours du projet, la gouvernance d'OFSEP distingue instances décisionnelles, consultatives et opérationnelles.

Les instances décisionnelles reposent sur un comité de pilotage mensuel de 15 membres dont les réunions sont préparées par un bureau de 3 membres assistant le coordinateur scientifique qui a la responsabilité générale du projet.

Les instances consultatives reposent principalement sur un comité consultatif international (en charge de la stratégie scientifique) et un conseil scientifique (en charge de l'examen des projets de recherche).

Le centre de coordination national constitue, sous l'autorité du coordinateur scientifique, la principale instance opérationnelle : il assure le bon fonctionnement de la cohorte et la gestion des projets de recherche, académiques ou industriels.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

Selon les données communiquées par l'agence nationale de la recherche (ANR), les recettes du projet OFSEP se sont globalement élevées sur la période 2012-2019 à 14,08 M€, dont 9,76 M€ au titre des financements du PIA (compte tenu des réorientations du projet décidées dans le cadre du plan d'actions qui a suivi l'évaluation intermédiaire, ce financement pourra être utilisé avec l'accord de l'ANR jusqu'au 30 juin 2021).

En 2019, 17 agents contribuent directement au projet OFSEP au niveau du centre de coordination nationale (équipe projet), représentant 14,1 ETP (5,5 au titre de la cohorte mère et 8,6 au titre de la nouvelle cohorte HD) ; le financement de 9,1 ETP est assuré par le PIA.

Les financements du PIA ont couvert les charges liées aux attachés de recherche clinique (ARC), à une partie du coût des personnels des centres de ressources biologiques (CRB) et à une partie du coût de l'équipe projet.

8 agents sont des personnels permanents en contrat à durée déterminée (pour une part employés par la Fondation Edmus) ; 9 sont en CDD.

Outre le financement au titre du PIA (9,76 M€), le projet OFSEP a obtenu de 2011 à 2019 les principaux financements publics et privés suivants (en M€), la part provenant des partenariats avec plusieurs laboratoires pharmaceutiques étant particulièrement significative.

Financements européens	Fondation pour l'aide à la recherche sur la sclérose en plaques (ARSEP)	Laboratoires pharmaceutiques
0,09	0,58	3,65

5.2 Perspectives budgétaires pour la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale porte une demande de financement complémentaire de 5,45 M€ destinés à permettre la poursuite du suivi, au développement de la nouvelle cohorte HD (inclusions) et à consolider la cohorte mère. Le budget prévisionnel global cumulé du projet sur la période 2020-2024 s'élève à 14,33 M€.

Toutefois, comme suite à l'évaluation finale le comité de pilotage chargé d'instruire la demande de financement complémentaire a proposé de ne pas prolonger le financement du projet au titre du PIA, ce qui peut poser question au regard des appréciations portées par la mission (*cf. infra*).

6 Accès aux données

Les données sont mises à disposition gratuitement des chercheurs académiques et des partenaires industriels, français ou étrangers, sur la base d'un dépôt de projets scientifiques examiné par le conseil scientifique. Une charte « publication » et une charte « accès aux données et propriété intellectuelle » encadrent la mise à disposition des données.

- Les équipes relevant des centres participants bénéficient d'un accès privilégié. Une brochure (en anglais) décrivant la procédure est en outre disponible sur le site web d'OFSEP.
- Le CCN intervient dans la phase amont de conception du projet afin d'en valider la faisabilité et apprécier le besoin de données.
- Le CCN peut effectuer les analyses statistiques préalables afin d'éviter autant que faire se peut les sorties de données.
- Les analyses statistiques internes sont facturés aux porteurs de projet.
- Les projets retenus font l'objet d'un contrat bilatéral fixant notamment les règles d'utilisation des données et les aspects financiers.

7 Appréciations portées par la mission

OFSEP repose sur un logiciel de données standardisé développé par les cliniciens et fortement approprié par eux dans le cadre de la fondation Edmus.

Le dispositif peut s'appuyer sur un fort soutien des institutions membres du consortium, avec un partage de tâches de *back office* clair (la gestion des crédits ANR par l'université Lyon 1, la gestion réglementaire et les questions de propriété intellectuelle par la fondation Edmus),

Il bénéficie de partenariats industriels importants (matérialisés par un accord pilote avec quatre laboratoires pharmaceutiques en 2017), soutenus par une reconnaissance affirmée de la part de la HAS.

L'appariement avec le SNDS est toutefois encore en projet, même s'il pourra s'appuyer sur une équipe de l'EHESP désireuse de conduire des analyses épidémiologiques à partir de la base appariée.

Le recentrage du projet initial sur la cohorte OFSEP-HD, réalisé suite à l'évaluation du projet en 2017 a permis d'enrichir le dispositif d'un ensemble complet de données cliniques, biologiques et radiologiques, ainsi que de questionnaires remplis par les patients (*Patient Reported Outcome Measure-PROM*). Il soulève néanmoins des questions liées à la taille de la cohorte suivie (moins de 2000 patients, ce qui peut être insuffisant pour développer les analyses prévues) et à la prise en compte des biais de sélection (liés au recrutement dans les seuls centres Edmus).

LA COHORTE PSY-COH

1 Présentation générale du dispositif

Les maladies mentales représentent un enjeu de santé publique majeure et une source de coûts très importants pour l'assurance maladie. Certains de ces troubles se manifestent chez les jeunes adultes, et peuvent être assortis de comportements suicidaires ou de co-morbidités somatiques, à l'origine d'une mortalité plus élevée. Parmi eux, la schizophrénie touche 0,7 à 1% de la population et les troubles bipolaires 1 à 2,5%.

1.1 Objectifs

- Assurer le suivi de 1 200 patients adultes ayant un trouble bipolaire ou une schizophrénie pendant au moins 5 ans.
- Identifier des marqueurs cliniques et biologiques de stratification, stades et trajectoires de ces pathologies, ainsi que les déterminants des rechutes éventuelles
- Analyser les co-morbidités somatiques et les coûts médico-économiques associés à ces maladies

1.2 Principales caractéristiques de la cohorte

- Cohorte spécialisée composée de deux sous-cohortes centrées sur des groupes de patients diagnostiqués respectivement pour troubles bipolaires (*Bipolar Disorders* -BP) et pour schizophrénie (SZ)
- Recrutement et suivi dans 21 centres experts hébergés au sein de services de psychiatrie universitaires
- Constitution d'une base de données biologiques comprenant plasma, ADN, ARN et cellules mononucléaires périphériques sanguines

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

L'inclusion des patients atteints de troubles bipolaires (BD) depuis moins de 5 ans a démarré fin 2014, et, après des retards, a dépassé en 2019 l'objectif recherché de 800 patients. Les examens de suivi à un an et à deux ans ont concerné respectivement 77% et 66% de ces patients.

L'inclusion des patients atteints de schizophrénie(SZ) et traités depuis moins de dix ans a débuté en 2015 et concerne 452 patients en juillet 2019. Les examens de suivi à un an et à deux ans ont concerné respectivement 70% et 47% de ces patients.

	Number of inclusions (realized/expected)	Patients seen at least once after inclusion / patients with enough follow-up	Patients seen at least twice after inclusion / patients with enough follow-up
PSY-COHorte BD	844/800	532/689 (77%)	339/512 (70%)
PSY-COHorte SZ	452/400	228/324 (66%)	127/269 (47%)

Ces patients seront suivis pendant une durée de 5 ans, la poursuite de la cohorte se faisant à partir de janvier 2025 sous la forme d'un suivi passif.

2.2 Modalités d'inclusion et d'échantillonnage

Les patients sont recrutés, sur la base du volontariat, dans des centres experts où ils sont généralement adressés par leur médecin traitant, à savoir :

- pour la cohorte BD, dans 11 centres experts (Paris, Créteil, Versailles, Colombes, Nancy, Grenoble, Marseille, Montpellier, Bordeaux, Besançon et Clermont-Ferrand), avec l'appui de 7 centres d'investigation clinique (CIC) ;
- pour la cohorte SZ, dans 10 centres experts (Colombes, Créteil, Versailles, Strasbourg, Grenoble, Marseille, Clermont-Ferrand, Bordeaux, Montpellier, Lyon) avec l'appui de 8 centres d'investigation clinique.

Le suivi des patients est prévu une fois par an durant cinq années consécutives, l'inclusion comme le suivi comprenant à la fois une évaluation clinique et le recueil de données biologiques.

Compte tenu des difficultés à mobiliser les patients atteints de troubles mentaux pour ce type de suivi, la cohorte s'est heurtée à d'une part à des difficultés de recrutement pour parvenir aux objectifs de recrutement souhaités, d'autre part à des taux d'attrition élevés concernant le suivi à un ou deux ans.

Ainsi, en mars 2019 :

- pour la cohorte BD, 532 patients avaient été suivis une fois et 339 deux fois après l'inclusion, et l'on disposait d'un ensemble de prélèvements biologiques pour 680 patients et de deux pour 201 d'entre eux ;
- pour la cohorte SZ, 228 patients avaient été suivis une fois et 127 deux fois, et l'on disposait d'un ensemble de prélèvements biologiques pour 377 patients, et de deux pour 112 d'entre eux.

Les promoteurs de la cohorte ont mis en œuvre, suite à l'évaluation de 2017, une série de mesures pour réduire cette attrition : simplification de l'évaluation annuelle et des questionnaires associés, réduction de la durée des visites, organisation de relances et d'outils de communication, envoi d'auto-questionnaires aux patients ne souhaitant pas revenir vers les centres.

Les objectifs de recrutement ont en revanche été conservés, et finalement atteints en 2019, dans la mesure où un objectif important des cohortes est d'être à même de décrire les différents stades de progression des pathologies considérées.

2.3 Contenu, collecte et stockage des différents types d'informations

Une part notable des informations collectées sont issues des examens cliniques et évaluations réalisées dans le cadre du soin et comprennent :

- des auto-questionnaires à l'inclusion sur des marqueurs de vulnérabilité à la maladie ;
- des tests neuropsychologiques réalisés à l'inclusion puis tous les deux ans ;
- des examens cliniques généraux réalisés à chaque visite de suivi (poids ; signaux vitaux, ECG...) ; des auto- et hétéro- questionnaires évaluant l'état symptomatique de la maladie, les troubles de l'humeur et le fonctionnement ;
- le recueil des traitements pris et en cours ;
- des modules spécifiques d'auto-questionnaires sur les facteurs environnementaux (migration, vie urbaine), économiques, et de styles de vie (exercice physique, alimentation, travail, événements de vie).

Les examens biologiques réalisés dans les centres comprennent quant à eux des bilans biochimique, lipidique, hépatique, thyroïdien et sanguin. Sont en outre effectués et centralisés tous les deux ans des prélèvements destinés à obtenir du sérum, du plasma, des cellules mononuclées du sang périphérique, de l'ADN et de l'ARN.

Ces prélèvements sont conservés, selon leur nature, à la plateforme de ressources biologiques de l'hôpital Mondor ou au CRB de la Pitié-Salpêtrière (ICM).

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

La cohorte s'inscrit dans le cadre d'un dispositif d'outils cliniques et d'évaluation partagés par les réseaux de 43 centres experts en troubles psychiatriques majeurs coordonnés par la fondation Fondamental, et qui rendent possible, pour chacune des quatre grandes pathologies suivies (troubles bipolaires, schizophrénie, troubles du spectre de l'autisme, dépression résistante), le recueil de données médicales et biologiques comparables dans le cadre d'un dossier patient informatisé ; les bases de données ainsi constituées, dites FACE (*Fondamental advanced center of expertise*) sont stockées sur un serveur situé à la faculté de Jussieu.

Des contacts ont été pris avec les responsables de la cohorte Constances pour comparer les données de PSY-COH avec celles recueillies en population générale, mais sans débouché concret à ce stade.

Un appariement est par ailleurs envisagé avec le SNDS à la fois pour mieux apprécier les caractéristiques des personnes suivies dans l'échantillon et pour pallier l'attrition de la cohorte.

2.5 Horizon et perspectives envisagés pour le dispositif après 2020

Il est envisagé que la cohorte poursuive le recrutement et le suivi des patients de façon à disposer d'au moins une série de prélèvements biologiques pour les 1200 patients prévus et de trois séries de prélèvements pour 400 d'entre eux.

Ces suivis cliniques et biologiques ont vocation à être conduits au maximum sur 5 ans, la cohorte ne donnant plus lieu à partir de 2025 qu'à un suivi « passif » à partir des données du SNDS.

L'enrichissement de l'exploitation des données de la cohorte est par ailleurs envisagé dans plusieurs directions, comme le génotypage des participants, les analyses épidémiologiques et médico-économiques permise par l'appariement projeté avec le SNDS ou le développement en lien avec le CEA d'outils d'analyse intégrée combinant les données biologiques, protéomiques et métabolomiques.

3 Principaux résultats et retombées en termes de recherche et de santé publique

Les responsables de la cohorte soulignent l'apport de cette dernière à des projets de recherche permettant d'éclairer une connaissance plus fine des différents stades de ces maladies, les syndromes métaboliques associés et l'intérêt de la prévention et du traitement des risques cardiovasculaires qui en découle. Des recherches ont aussi été conduites concernant l'amélioration de la sévérité des pathologies suite aux prises en charge, les facteurs jouant sur le risque de rechute et les coûts médico-économiques associés, en particulier pour la schizophrénie qui concerne des patients pour des patients d'âge jeune.

Le nombre de publications permis par la cohorte commence à se développer depuis 2017-2018 : il est évalué à plus d'une soixantaine, dont une vingtaine par an les dernières années.

Des journées d'information et de débats ont été organisées, en collaboration avec les associations de patients (ARGOS, UNAFAM...), d'une part en direction des participants à la cohorte (2017) et d'autre part, sous l'égide de la fondation Fondamental, en direction du public (une journée par an depuis 2015).

La cohorte est connue des milieux internationaux travaillant sur ces pathologies et donne lieu à des communications et des collaborations internationales (*Psychiatric Genetics Consortium*, ENIGMA, groupe international de chercheurs partageant des données d'imagerie cérébrale).

Les jurys d'évaluation, tout en reconnaissant l'intérêt de la cohorte pour éclairer des domaines de la psychiatrie rarement étudiés, avaient émis des préoccupations concernant la lenteur des recrutements et les taux d'attrition, auxquels les responsables de la cohorte ont tenté de répondre. La nécessité de mieux appréhender les biais de sélection a aussi été soulignée.

Compte tenu de sa petite taille et de sa spécificité, la cohorte a par contre peu de visibilité et d'écho parmi les responsables du ministère de la santé et les agences sanitaires.

4 Gouvernance du dispositif

La cohorte repose sur une coopération institutionnelle entre :

- la fondation Fondamental : cette fondation de coopération scientifique dédiée à la lutte contre les troubles psychiatriques majeurs et a structuré un réseau de 43 centres experts reconnus par le ministère de la santé, et en est l'animateur et le coordinateur ;
- l'équipe de psychiatrie translationnelle de l'université de Paris Est Créteil (UPEC) qui assume la responsabilité de la biobanque ;
- l'institut Pierre Louis d'épidémiologie et de santé publique de Sorbonne Université, pour les domaines ayant trait à l'épidémiologie et aux statistiques ;

- l'unité d'épidémiologie clinique et évaluation économique de l'Inserm, pour les études médico-économiques.

Ce partenariat sera prochainement élargi, dans le cadre d'un nouvel accord de consortium, au Commissariat à l'énergie atomique (CEA), pour les études faisant appel aux plateformes génétiques et métabolomiques et à d'intelligence artificielle) et à l'université de Paris- Saclay.

La gouvernance de la cohorte est assurée par le comité rassemblant ces partenaires institutionnels et un comité de pilotage opérationnel. Son comité scientifique ne se distingue pas de celui de la fondation Fondamental : il se réunit une fois par an et donne son avis sur les problèmes scientifiques et les projets de recherche se rapportant à la cohorte.

Deux éléments seront prochainement ajoutés à cette gouvernance : un comité des associations de patients et un comité d'évaluation.

5 Ressources, budgets et financements

5.1 Sur la période 2012-2019

Selon les données communiquées par l'agence nationale de la recherche (ANR), les financements reçus au titre du PIA se sont montés à 3,85 M€³⁷ pour 0,75 M€ obtenus dans le cadre de partenariats avec le privé (SANOFI).

Les frais de personnels ont représenté plus de 60 % des dépenses de la cohorte, essentiellement pour des contrats à durée indéterminée supportés par la fondation Fondamental.

5.2 Perspectives budgétaires pour la période 2020-2024

La demande budgétaire présentée pour la période de 2020-2024 en complément de financement de la part du PIA est de 1,5 M€ pour des dépenses totales évaluées à 6,84 M€ sur les cinq ans.

Les contributions des partenaires institutionnels sont supposées pouvoir y contribuer pour 1,1 M€, et celles issues d'autres partenariats ou de subventions à hauteur de 1,2 M€

L'objectif de ces suppléments budgétaires serait de permettre l'achèvement des opérations d'inclusions et de suivi des participants (examens cliniques et biobanques), dont la fin est prévue à la fin 2024.

Comme suite à l'évaluation finale le comité de pilotage chargé d'instruire la demande de financement complémentaire a proposé en novembre 2019 un complément de financement du projet à hauteur de 0,8 M€ jusqu'au 30 juin 2025.

³⁷ Le projet a par ailleurs bénéficié d'un financement dans le cadre du PIA au titre du Labex Bio - Psy

5.3 Les partenariats industriels

La construction du projet a bénéficié d'un mécénat de la part de Sanofi, fléché sur la cohorte BP, sans contrepartie en termes de suivi ou dépôt de projets de recherche.

Elle bénéficie aussi d'un mécénat de compétences de la part d'Altran, sous forme de mise à disposition de deux personnes, dont un chef de projet.

Ces partenariats ont été négociés par la fondation Fondamental, selon des règles et modalités qu'elle a définies.

6 Accès aux données

L'accès aux données se fait dans le cadre d'une charte qui fixe les règles relatives aux publications et prévoit que les projets scientifiques soumis recueillent à la fois l'avis du centre expert responsable, du comité Biobanque de la fondation Fondamental, de son comité scientifique et du comité de pilotage de la cohorte.

Ce processus a un délai estimé à deux mois, sachant que, compte tenu de la spécialisation de la cohorte, la quasi-totalité des demandes d'accès proviennent de chercheurs associés aux centres experts ou participants aux *consortia* internationaux dont elle est membre.

7 Appréciations de la mission

Les deux sous-cohortes constitutives de la cohorte PSY-COH très ciblées et reconnues dans leur champ, au plan national comme international. Elles s'inscrivent dans le prolongement d'un dispositif d'animation de centres experts, co-financé par le ministère de la santé et participent d'un réseau dense et fourni de relations avec les milieux de recherche, dont le Labex Bio-Psy et un ensemble de chercheurs en psychiatrie, épidémiologie statistique, économie de la santé, *machine learning* sur les données biologiques et – omiques...

Les préoccupations portées par ces cohortes s'agissant des liens entre pathologies mentales et pathologies somatiques, et de prise en compte des aspects socio-économiques de ces maladies ont une importance avérée en termes de santé publique. Leur horizon est en précisément défini, un suivi passif devant prendre le relai des examens cliniques et prélèvements à partir de 2025.

Les problèmes rencontrés pour atteindre le nombre de recrutement souhaités (en termes de prélèvements biologiques) et les taux d'attrition élevés observés imposent toutefois des efforts vigilants et continus pour être à même d'atteindre effectivement les objectifs affichés et apprécier la portée des résultats obtenus. L'effort doit à cet égard aussi viser l'appariement avec les données du SNDS, qui n'est pas encore réalisé, de même que le rapprochement envisagé avec la cohorte Constances en vue d'identifier les troubles bipolaires et la schizophrénie dans les bases médico-administratives.

Le management de la cohorte s'appuie sur une fondation de coopération scientifique qui permet des souplesses de gestion et offre la faculté de recruter du personnel à durée indéterminée. Ce choix se traduit toutefois par une intrication des instances de la cohorte et de celles de la fondation, et par des modalités spécifiques d'accès aux données et de relations avec les partenaires industriels, qui mériteraient d'être évaluées.

LA PLATEFORME DE COHORTES RADICO

1 Présentation générale du dispositif

Une maladie est dite rare lorsque moins d'une personne sur 2 000 en est atteinte. On estime ainsi³⁸ qu'il existe plus de 7 000 maladies rares dont souffriraient en France environ trois millions de patients, 75 % étant des enfants. Les maladies rares constituent plus du tiers des causes de décès d'enfants.

Il s'agit de pathologies complexes dans la plupart des cas d'origine génétique (plus de 3 700 gènes responsables sont identifiés en juin 2019, mais les gènes en cause restent dans de nombreux cas encore inconnus), en général chroniques, évolutives, dégénératives et invalidantes, pour lesquelles il n'existe pas de traitement curatif dans 95 % des cas.

Les maladies dites rares constituent donc un très important enjeu de santé publique, qui a justifié l'élaboration et la mise en place depuis 2005 de plans nationaux maladies rares (PNMR) réussis (en cours en décembre 2019 le PNMR3 2018-2022), comportant un important volet recherche et qui se sont traduits notamment par la création de la banque nationale de données maladies rares (BNDMR). Compte tenu en particulier du faible nombre par pays de patients atteints par une maladie donnée, qui nécessite la couverture géographique la plus large possible pour entreprendre et mener à bien des travaux de recherche, la stratégie nationale en matière de maladies rares est fortement articulée avec les initiatives européennes dans le domaine (cf. notamment la création des réseaux européens de référence – ERN – en matière de maladies rares en 2016 et la mise en place du programme européen conjoint (EJP) sur les maladies rares qui finance des projets de recherche).

Dans ce contexte a émergé le besoin de disposer dans un cadre coordonné et mutualisé de cohortes d'ampleur nationale voire européenne de patients atteints d'une maladie rare, produisant un ensemble de données permettant de répondre aux nombreuses questions posées par ces pathologies : description de leur histoire naturelle, compréhension de leurs mécanismes pathophysiologiques, analyse des corrélations phénotype / génotype, identification de nouvelles approches thérapeutiques et stratification des patients éligibles, estimation de leurs impacts sociaux et médico – économiques.

Le projet RADICO est conçu comme la réponse à ce besoin.

1.1 Objectifs

- Créer une plateforme nationale de services et d'expertises mutualisés permettant de centraliser la collecte, la structuration, la gestion, le contrôle qualité et l'analyse des données recueillies par des cohortes de patients atteints par une maladie rare à mettre en place.
- Construire ex nihilo au cœur de cette plateforme de services une infrastructure technique permettant la constitution, le stockage et la mise à disposition, de manière standardisée et conforme aux obligations de sécurité tant techniques que réglementaires, d'une base de données rassemblant toutes les données collectées dans le cadre des cohortes mises en place.

³⁸ Cf. le portail Orphanet créé en 1997 par l'INSERM dédié aux maladies rares

- Intégrer effectivement à la plateforme et au système d'information ainsi créés un premier ensemble de cohortes et en piloter les différentes étapes de montée en charge avec les porteurs de chacune d'elle, en animant les travaux de la communauté ainsi constituée.
- Intégrer cette plateforme française dans l'ensemble des réseaux et dispositifs européens qui contribuent à la recherche sur les maladies rares.

1.2 Principales caractéristiques du projet

L'infrastructure technique de stockage centralisé de la base de données issues des cohortes hébergées a été construite et installée sur le site de l'hôpital Trousseau (APHP) à Paris selon les modalités du *cloud computing* en intégrant les principes d'interopérabilité du système d'information et de *FAIR Data access* (données « faciles à trouver, accessibles, interopérables et réutilisables ») et en s'appuyant pour la construction des différentes composantes fonctionnelles de l'infrastructure sur des logiciels libres (notamment la suite *Research Electronic Data Capture – REDCAP* – de l'université Vanderbilt aux Etats-Unis).

En parallèle, le projet RADICO a sélectionné un premier ensemble de cohortes dans le cadre d'un appel à projet public à destination des centres experts hospitaliers maladies rares. Cet appel à projet (AAP) organisé en deux étapes de janvier à juin 2014 a permis la sélection d'une première vague de cohortes. 63 lettres d'intention avaient été reçues et examinées par le conseil scientifique de RADICO, 32 projets pré-sélectionnés et leurs coordinateurs invités à déposer un dossier complet. Chacun des dossiers complets reçus a été évalué par un collège d'experts étrangers. Les seize premiers classés par le conseil scientifique de RADICO ont été retenus, soit au final 25 % des projets déposés. En juin 2019 toutefois, compte tenu des difficultés techniques ou de pilotage rencontrées par trois des cohortes correspondantes qui se sont traduites par la décision des comités exécutif et scientifique du projet d'en arrêter la constitution, seules treize sont en cours de constitution au second semestre 2019 (la liste de ces cohortes avec une brève description de chaque maladie concernée et des principaux objectifs de la cohorte associée est fournie en annexe à la présente fiche). Ces treize cohortes couvrent 67 maladies rares.

2 Caractéristiques du projet

2.1 Dates de démarrage, nombre de participants

Compte tenu de la longueur du délai de préparation des travaux nécessaires à la mise en place des cohortes sélectionnées et de la procédure d'obtention des autorisations nécessaires de la part du comité consultatif pour le traitement de l'information en recherche en santé – CCTIRS – puis de la CNIL, dans un contexte d'évolution de la législation applicable aux recherches impliquant la personne humaine (loi n° 2012-300 du 5 mars 2012 dite « loi Jardé »), les inclusions dans les treize cohortes hébergées n'ont débuté qu'entre novembre 2016 pour la première ayant démarré et juin 2019 pour la dernière.

Les objectifs en termes de nombre de patients inclus varient considérablement d'une cohorte à l'autre : de 60 pour la plus petite (cohorte PP) à 2 550 pour la cohorte la plus importante (COLPAC) ; au total sur le périmètre des treize cohortes hébergées, l'objectif est de 13 188 patients inclus.

Les périodes prévisionnelles d'inclusion varient également fortement d'une cohorte à l'autre : de une année et demi pour la cohorte PP à cinq années pour la plupart des cohortes et jusqu'à dix années pour l'une d'entre elles (AC-ŒIL) ; de ce fait, la fin des périodes d'inclusion s'étale de décembre 2019 à mi - 2027.

Au 30 juin 2019 le nombre de patients inclus (y compris les patients pour lesquels un certain nombre de données collectées restent à saisir dans les bases de données) est de 5 595 ce qui est globalement en ligne avec l'objectif fixé à cette date dans le cadre du projet, avec cependant d'importantes variations cohorte par cohorte, certaines ayant dépassé leurs objectifs d'autre étant en deçà.

2.2 Modalités d'inclusion

Les données recueillies à l'inclusion et durant le suivi des patients peuvent varier d'une cohorte à l'autre, selon la ou les questions scientifiques posées : formulaires en ligne alimentés par les cliniciens, questionnaires accessibles par internet ou applications smartphone pour les patients, extraction depuis des bases de données préexistantes dans les hôpitaux (données prévalentes) parfois déjà partagées entre plusieurs centres spécialisés. Elles incluent généralement des données issues de l'examen de santé (données cliniques) et du diagnostic (mutations génétiques), de l'information sociodémographique, des échantillons biologiques et des données de qualité de vie recueillies auprès des patients à partir d'auto-questionnaires.

Les inclusions sont effectuées dans les circuits de prise en charge des patients structurés dans le cadre des plans nationaux maladies rares (PNMR) : en juin 2019, 271 centres participent à la collecte sur le territoire national auxquels s'ajoutent quatre centres de collecte implantés dans d'autres Etats européens (Belgique, Espagne, Italie et Allemagne), compte tenu de la dimension européenne de certaines cohortes (ECYSCO et Eurbio-Alcort).

2.3 Contenu, collecte et stockage des différents types d'informations

Les données collectées (diagnostics, traitements, données d'hospitalisation, questionnaires) par le circuit de prise en charge des patients, dont une partie alimente la BNDMR, sont in fine stockées de manière centralisée dans le système d'information créé dans le cadre du projet RADICO.

Six cohortes disposent d'une biobanque pour stocker les échantillons biologiques collectés (ADN, ARN, plasma, échantillons sanguins, urine ou selles) ou ont le projet de constituer une biobanque ; ces biobanques sont hébergées dans divers CRB sur le territoire national souvent en lien avec l'établissement d'exercice du coordonnateur scientifique de la cohorte (l'hôpital Necker pour les cohortes Eurbio-Alport et Ecysco, l'hôpital Trousseau pour la cohorte DCP, le CHU de Toulouse pour la cohorte AC-Œil, etc...).

2.4 Articulation avec d'autres cohortes ou dispositifs d'information

La plateforme RADICO et les cohortes dont elle héberge les données s'inscrivent dans le cadre de l'animation scientifique des 23 « filières de santé maladies rares » mises en place dans le cadre du PNMR2, qui regroupent sur le territoire national 109 centres de référence maladies rares », 1 757 « centres de compétence maladies rares » et 54 laboratoires hospitalo-universitaires de diagnostic moléculaire. Ces filières de santé jouent un rôle clef dans la pré-sélection pour RADICO de sujets de recherche à fort impact.

La plateforme est directement liée à la banque nationale de données nationale maladies rares (BNDMR) qui collecte un socle minimal de données pour chaque patient diagnostiqué, elle-même en cours de construction et qui constitue le registre national maladies rares que les cohortes contribuent à enrichir.

Compte tenu du fait que le processus d'inclusion des patients des différentes cohortes est toujours en cours à la mi -2019, l'appariement avec le SNDS n'a pas été réalisé ; les porteurs de la plateforme envisagent de l'étudier sous l'égide de l'INSERM dans le cadre de la mise en place de France Cohortes.

2.5 Horizon et perspectives envisagés pour le dispositif après 2020

La période 2020-2025 voire 2030 correspond pour la plupart des cohortes hébergées par la plateforme RADICO à la période de suivi actif des patients inclus, de durée elle-même variable selon les cohortes (deux, trois, cinq et jusqu'à dix années).

Le nombre de centres en milieu hospitalier effectuant les inclusions doit monter en charge pour atteindre rapidement une cible de plus de 600 sur le territoire national et 15 en Europe pour les cohortes à recrutement international, avec de nouveaux pays contributeurs (Hongrie, Luxembourg, Pays-Bas, Royaume Uni).

Concernant l'infrastructure de stockage des données elle-même, l'INSERM a fait réaliser au second semestre 2019 un audit de la solution technique (cabinet Ernst and Young) en vue de son intégration projetée au système d'information dont la création est envisagée dans le cadre du projet France Cohortes.

L'arrivée progressive à maturité des cohortes hébergées doit permettre d'intensifier et d'approfondir les collaborations dans le cadre des initiatives mises en place au niveau européen.

3 Principaux résultats et retombées en termes de recherche et de santé publique

Les premières inclusions ayant débuté en novembre 2016 pour la première cohorte (GeniDA) et en juin 2019 pour la dernière, avec une fin de période d'inclusion s'étalant selon les cohortes et en fonction de la durée de celle-ci du second semestre 2019 à 2027, les premières publications scientifiques s'appuyant sur le projet RADICO, au nombre de vingt-six à fin 2019, sont relativement récentes (la première remonte à 2015).

Ces premières publications concernent une première description de l'infrastructure RADICO elle-même, sa méthodologie et de ses enjeux (cinq publications entre 2015 et 2018) ; vingt et une autres

publications concernent depuis 2017 huit des treize cohortes hébergées et portent notamment sur la découverte de nouveaux gènes responsables (cohortes DCP – Dyskinésie ciliaire primaire – et ACCEil – Défaut congénital de l’œil), sur l’analyse des relations entre phénotype et génotype et l’amélioration à partir de la pathophysiologie de la maladie de l’établissement du diagnostic (cohorte PID – pneumonie interstitielle idiopathique).

Compte tenu de l’arrivée progressive à maturité des différentes cohortes hébergées le nombre de publications est en rapide croissance (quatre de 2015 à 2017, neuf en 2018, treize en 2019), traduisant ainsi que le jury d’évaluation finale l’a souligné l’important potentiel de l’infrastructure en matière de recherche.

4 Gouvernance du dispositif

Le projet RADICO dispose d’une gouvernance à trois niveaux reposant sur un comité exécutif comprenant les trois coordinateurs scientifiques et le directeur opérationnel du projet, qui se réunit hebdomadairement et a en charge le pilotage du projet, un comité consultatif institutionnel, qui constitue l’instance décisionnelle dans laquelle sont représentés les membres de l’accord de consortium, et un comité scientifique de dix-huit membres dont les représentants des trois instituts thématiques de l’INSERM concernés par le projet, qui se réunit une à deux fois par an et examine les orientations stratégiques du programme (il a été chargé dans ce cadre de définir les critères d’examen des candidatures et les réponses à l’appel à projet qui a permis à l’équipe RADICO d’arrêter en 2014 la liste des cohortes qui ont intégré l’infrastructure).

L’accord de consortium associe à l’INSERM, établissement coordinateur, l’Université Paris Descartes, l’Université Paris Diderot, l’Université Paris Est Créteil, Aix Marseille Université, l’Université de Montpellier, Les entreprises du médicament (LEEM) et l’Alliance pour la recherche et l’innovation des industries de santé (Ariis).

Le support et le soutien du projet (aspects financiers) sont fournis par la direction régionale de l’Inserm compétente. Inserm Transfert, filiale de l’Inserm prend en charge pour chaque cohorte concernée la définition et la mise en œuvre des partenariats avec les industriels.

Par ailleurs, chacune des cohortes ayant intégré l’infrastructure est elle-même dotée d’un conseil scientifique et d’un comité de pilotage auquel participent des membres de l’équipe RADICO.

5 Ressources, budgets et financements

5.1 Sur la période 2011-2019

Selon les données communiquées par l’agence nationale de la recherche (ANR), les recettes du projet RADICO se sont globalement élevées sur la période 2012-2019 à 12,90 M€ dont 10,51 M€ au titre du financement du PIA.

Outre les trois coordinateurs scientifiques du projet pris en charge par leur établissement d’affectation, l’équipe RADICO compte au 1^{er} octobre 2019 vingt membres, qui hormis le chef de projet (personnel statutaire de l’Inserm) sont tous en contrat à durée déterminée (CDD) de l’Inserm. L’équipe comprend trois pôles fonctionnels : un pôle Système d’information et e-Santé (quatre agents), un pôle Biométrie (trois agents) et un pôle Recherche clinique (onze agents).

Outre le financement au titre du PIA (10,51 M€), le projet RADICO a perçu de 2012 à 2018 les financements privés suivants (en M€) :

Boehringer ³⁹	Shire-Takeda ⁴⁰	Total
1,38	1,00	2,38

5.2 Perspectives budgétaires pour la période 2020-2024

Le dossier transmis dans le cadre de l'évaluation finale de septembre 2019 portait une demande de financement complémentaire de 3,5 M€ destiné à achever les opérations de suivi des patients inclus et à consolider les différentes cohortes hébergées par l'infrastructure. Le projet de budget prévisionnel cumulé sur la période 2020-2024 associé à cette demande s'élevait à 6,90 M€ correspondant pour 65 % de ce montant à des dépenses de personnel.

Comme suite à l'évaluation finale, le comité de pilotage chargé d'instruire cette demande a proposé en novembre 2019 un complément de financement du projet de 1,8 M€ jusqu'au 30 juin 2025.

5.3 Les partenariats industriels

Trois cohortes hébergées par l'infrastructure RADICO ont fait l'objet d'importants partenariats avec des entreprises pharmaceutiques : la cohorte GenIDA – Déficience intellectuelle et maladies du spectre autistique – avec Roche, la cohorte PID avec Boehringer et la cohorte MPS – mucopolysaccharidose – avec Shire – Takeda.

Ces partenariats, construits selon des règles définies par INSERM transfert, qui portaient sur des études post inscription ou de réinscription de médicaments, demandées par la haute autorité de santé (HAS), ont très largement contribué au financement des cohortes en cause compte tenu du montant total des apports (1,45 M€ de la part de Boehringer et 2,02 M€ de la part de Shire – Takeda) et de leurs modalités de fixation (ils comportent en effet une partie fixe forfaitaire destinée à la constitution et à la maintenance de la cohorte, qui ouvre accès à la base de données de la cohorte, et une partie modulable en fonction du projet de recherche, qui donne droit aux analyses effectuées à partir de la base de données).

6 Accès aux données

L'infrastructure RADICO alimentée par les données progressivement collectées par chaque cohorte, qui dispose d'un univers dédié au sein du système d'information qui comporte lui-même un ensemble de logiciels de traitement et d'analyse de données, est ouverte selon des modalités sécurisées aux équipes en charge de chaque cohorte.

Les modalités d'accès aux données de chaque cohorte, dans le cadre d'un projet de recherche, sont définies par une charte élaborée sous l'égide de l'Inserm : elle fixe notamment le principe d'une contribution financière pour les porteurs de projets n'appartenant pas à l'accord de consortium et prévoit que le projet de recherche soit examiné et validé par les instances de RADICO.

³⁹ Au titre de la cohorte PID pour un financement pluriannuel total prévu de 1,45 M€

⁴⁰ Au titre de la cohorte MPS pour un financement pluriannuel total prévu de 2,02 M€

7 Appréciations de la mission

Plusieurs cohortes ont une dimension internationale affirmée et l'infrastructure contribue étroitement aux initiatives européennes en matière de recherche sur les maladies rares.

Deux cohortes connaissent des difficultés ou des retards et font l'objet d'une vigilance particulière de la part du comité exécutif de la plateforme.

Les durées d'inclusion et de suivi prévues sont très différentes selon les cohortes (de trois à vingt ans au total) ce qui pourrait poser un problème de viabilité à terme de l'infrastructure, si ses contours demeurent identiques.

Au-delà de chacune des cohortes qu'elle abrite, RADICO a été pensée comme une plateforme à vocation en partie permanente. Son intégration au projet France Cohortes, dont RADICO constitue les prémisses par plusieurs aspects, et qui est souhaitée par ses promoteurs, doit être pensée spécifiquement afin d'en préserver les acquis.

La question de l'extension du périmètre de l'infrastructure RADICO à de nouvelles cohortes maladies rares voire à des cohortes portant sur des maladies communes, dans le cadre de nouveaux appels à projets, se posera lorsque les cohortes actuelles seront suffisamment consolidées. Ce questionnement dépendra du degré et des modalités d'intégration de RADICO à France Cohortes.

LES 13 COHORTES RADICO

RaDiCo-GenIDA
<p>Full title : Genetic of intellectual deficiency and autism spectrum disorders Principal Investigator: Pr. Jean-Louis MANDEL</p>
<p>Background: (ID) affects 1-2% of the population and more than half of the patients with a diagnosis of autism (or autism spectrum disorder, ASD). For most of patients, ID has a genetic origin. Rationale: this cohort will enable:</p> <ul style="list-style-type: none"> – to develop an alternative model of database for genetic ID, organised as a social network in which most of the clinical information would be entered by the patient / his/her family, on the basis of questionnaires established by professionals but understandable by families, – to create specific networks devoted to exchanges between patients or families sharing the same mutated gene, – to acquire relevant information for a personalised medical care, on risks of specific pathologic complications and on potential adverse effects of symptomatic treatments <p>Primary objective: Compile data of groups of people with ID and / or ASD sharing the same genetic mutation to improve knowledge of the natural history of the disease and the associated comorbidities.</p>
RaDiCo-SEDVasc
<p>Full title: French National Cohort on the vascular Ehlers-Danlos Syndrome (vEDS) Principal Investigator: Pr. Xavier JEUNEMAÎTRE</p>
<p>Background: Very rare genetic condition: prevalence of 1/150 000, autosomal dominant trait, related to > 250 mutations in the COL3A1 gene, encoding the pro-α chain of type III procollagen Rationale:</p> <ul style="list-style-type: none"> – <u>Heterogenous</u> clinical presentations, from <u>acrogeric</u> to <u>ecchymotic</u> forms. Altered assembly, stability and thus secretion and resistance to tensile stress of this <u>fibrillar</u> collagen -> early spontaneous arterial, digestive and obstetrical accidents. – Average age at first major complication: 24 years old; Average life expectancy: 50 years old – No cure for vEDS. <u>Celiprolol</u> (beta-blocker with vasodilating properties) limiting the occurrence of major fatal and non-fatal events, but still a high frequency of arterial complications – Natural history poorly described, impact of <u>Celiprolol</u> on clinical outcome in the long run poorly known – No study on health cost / social impact <p>Primary objective: To describe the natural course of vEDS, notably the chronology of complications (arterial, digestive, pulmonary and uterine)</p>
RaDiCo-AC-ŒIL
<p>Full title: French National Cohort on Congenital Defects of the Eye: natural history, genetic determinisms and improved ocular and extra-ocular outcome prediction for better patient management Principal Investigator: Dr Nicolas CHASSAING</p>
<p>Background: The genesis of ocular structures (induction, differentiation) can be altered leading to the presence of congenital abnormalities such as <u>anophthalmia</u>, <u>microphthalmia</u>, <u>aniridia</u>, or other anterior segment dysgenesis (<u>Axenfeld</u> anomaly -<u>Rieger</u>, anomaly of Peters). Rare genetic conditions: prevalence of 1 to 2/10 000 and frequently associated with extra-ocular malformations. Rationale:</p> <ul style="list-style-type: none"> – The genetic determinisms of these ocular malformations and their implications in prognostic terms remains to be assessed. – Limited published cohorts with very heterogeneous patients, making their use in clinical practice complex. – Predicting the visual and neurological prognosis of a child in whom the diagnosis of congenital anomalies of the eye has been made in the neonatal or antenatal period and to offer to these children a well-defined protocol of care is currently very difficult – A developmental disorder may be present, secondary to sensory impairment, or directly linked to developmental brain damage and thus leading to intellectual impairment <p>Primary objective: To describe the long-term outcomes of the patients with ocular developmental defects, focusing on visual and neuro-developmental issues.</p>

RaDiCo-ECYSCO
<p>Full title: European <u>Cystinosis</u> Cohort Principal Investigator: Dr Aude SERVAIS</p>
<p>Background: Lysosomal storage disease characterized by the abnormal accumulation of the amino acid cysteine and caused by the mutations in the gene CTNS, located on chromosome 17, coding for <u>cystinosin</u>. <u>Cystine</u> crystal accumulation in organs causing different symptoms: infantile, juvenile or ocular clinical presentation. Rare autosomal recessive genetic disorder: Incidence 1/180 000 live births - Estimated 140 cases in France and 500-600 in Western Europe.</p> <p>Rationale:</p> <ul style="list-style-type: none"> - Significant limitations in the knowledge of natural history and long-term manifestations - In adulthood, care for the <u>cystinosis</u> patients is fragmented with major geographical variability and long-term evolution remains unknown - An active and sustained academic cohort is necessary to avoid independent “drug-oriented” registries, company driven, which would thus lead to a fragmentation of the data <p>Primary objective: To understand the natural history and major long-term manifestations and outcomes of <u>cystinosis</u> in paediatric and adult cases.</p>
RaDiCo-DCP
<p>Full title: French National Cohort on Primary Ciliary Dyskinesia: Identification of severity criteria from deep phenotyping and phenotype-genotype correlation study Principal Investigator: Dr Estelle ESCUDIER</p>
<p>Background: Rare autosomal recessive respiratory disease due to an impaired <u>mucociliary</u> clearance: abnormal ciliary structure and/or function. Estimated incidence of ~1/10,000 live births. Prevalence difficult to determine. Great variety of ciliary defects reflecting the high genetic heterogeneity: >35 genes so far implicated in PCD explain the pathology in ~50% of the patients. Characterized by chronic upper and lower airway infections (i.e. <u>sinopulmonary</u> syndrome with rhinosinusitis and bronchitis, leading to nasal polyposis and bronchiectasis, respectively) starting in early childhood.</p> <p>Rationale:</p> <ul style="list-style-type: none"> - In half of cases, associated to laterality defects with situs <u>inversus</u> (Kartagener syndrome) - Infertility in most male patients (immotile spermatozoa flagella) and in some affected women (miscarriage and/or ectopic pregnancy) - 1/3 of PCD patients developing a severe form of the disease with respiratory failure. PCD-specific factors that could predict the severity of the disease are not known at present but would be of prime importance for improving patient care <p>Primary objective: Identification of predictive factors of PCD severity</p>
RaDiCo-PID
<p>Full title: French National Cohort on Idiopathic Interstitial Pneumonia: From infancy to elderly Principal Investigator: Prof <u>Annick</u> CLEMENT</p>
<p>Background: Idiopathic Interstitial Pneumonia (IIP)s encompass a group of diffuse infiltrative lung disorders of unknown origin that affect the lung architecture. Characterized by a progressive and often irreversible remodelling of the <u>alveolo-capillary</u> barrier. IIPs are rare diseases that can be observed at all ages, from infancy to elderly.</p> <p>Rationale: Patients with IIP represent a significant healthcare burden due to the chronic nature of the diseases and the lack of effective medical therapies, with associated high morbidity and premature mortality. The nature of inciting injury and subsequent alveolar epithelium dysfunction includes genetic and epigenetic factors as well as environmental and host co-morbidity components. All together these contributors, which remain poorly understood, affect disease expression and progression. Although IIP is most frequently diagnosed in adult patients, new information supports the view of an on-going process of lung dysfunction since early in <u>life</u>. To significantly progress in IIP knowledge and care, rigorously <u>phenotyped</u> IIP patient cohorts combining both paediatric and adult populations are required, then covering the all disease spectrum from infancy to elderly.</p> <p>Primary objective: Describe the phenotypic features of the paediatric and adult patients with IIP/PID, at diagnosis and during the follow-up.</p>
RaDiCo-EURBIO-Alport
<p>Full title: Study of the natural history of <u>Alport</u> Syndrome by implementing an international database Principal Investigator: Dr Laurence HEIDET / Co-Principal Investigator: Pr. Bertrand KNEBELMANN</p>
<p>Background: <u>Alport</u> Syndrome (AS) is a rare genetic disease, characterized by the association of a glomerular nephropathy progressing toward end-stage renal failure, with deafness, and ocular defect</p> <p>Rationale:</p> <ul style="list-style-type: none"> - Significant progress made in understanding the molecular mechanisms responsible for the disease, but little in the comprehension of the progression of renal failure and in the field of therapeutics - Studies and treatment indications would be greatly facilitated by the discovery of biomarkers allowing to predict, earlier than the date of onset of proteinuria, the progression to renal failure - The project will allow to expand to Europe the database on AS <p>Primary objective: To study the natural history of AS</p>

RaDiCo-PP
<p>Full title: Utility and validation study of a smartphone application for periodic paralysis Principal Investigator: Dr <u>Savine VICART</u></p>
<p>Background: Primary periodic paralysis (PP) is a group of rare genetic autosomal dominant diseases (estimated prevalence of 1/100.000 in Europe) due to mutations in genes encoding muscle ion channels. The symptoms usually starting in childhood or adolescence characterized by acute and reversible muscle weakness attacks. PP are classified as hyperkaliaemic (<u>hyperPP</u>) or hypokaliaemic (<u>hypoPP</u>) based on changes in blood potassium levels during the access. These changes are often subtle and transient, making the distinction between the 2 forms difficult. Three genes (CACN1A5, SCN4A and KCNJ2) have been identified in PP. The combination of EMG and molecular diagnosis allows an effective diagnosis of periodic paralysis.</p> <p>Rationale: Currently, there is no validated scale or tool to assess paralytic attacks in a precise and standardized way. Thus, it appears crucial to succeed in assessing more accurately both the clinical and medico-economic impacts of these pathologies to improve, in the future, healthcare and healthcare costs.</p> <p>Primary objective: To evaluate the benefit on the quality of collected information of real-time capture, via a smartphone application, of data on characteristics of paralysis attacks (number, duration, intensity, triggering factors, location, treatment), analysed at 6 months and verified at 12 and 18 months, as compared to standard retrospective collection via an interview during routine patient care visits.</p>
RaDiCo-MPS
<p>Full title: <u>Mucopolysaccharidosis</u> patients in France in the era of specific therapeutics Principal Investigator: Pr. Thierry BILLETTE / Co-Principal Investigator: Dr <u>Bénédicte HÉRON</u></p>
<p>Background: Rare diseases with an estimated global incidence of 1/25-30,000 births. Lysosomal storage disorders caused by accumulation of sulphated carbohydrate polymers deriving from sulphated monosaccharides or <u>glycosaminoglycans</u> (GAG) in the lysosomes leading to <u>multisystemic</u> disease manifestations. Autosomal recessive inheritance for all but MPS-II, Hunter disease, an X-linked disorder. Genes responsible for the 11 enzyme deficiencies corresponding to 7 clinical subtypes (MPS I-IV, VI et IX). Chronic, progressive <u>multivisceral</u> diseases. Age at first symptoms may vary according to the severity of the disease, in early infancy or childhood in severe cases (even antenatal).</p> <p>Rationale: Disease-specific treatments available for some of the disorders: enzyme replacement therapy (ERT) and haematopoietic stem cell transplantation (HSCT) offer substantial benefit but no cure. Several therapeutic options under development; gene therapy and substrate inhibitors.</p> <p>Primary objective: To characterize the epidemiology and natural history of MPS diseases.</p>
RaDiCo-COLPAC
<p>Full title: French National Cohort on Low Phospholipid-Associated <u>Cholelithiasis</u> (LPAC) syndrome: new knowledge on the epidemiology, clinical and genetic heterogeneity of the disease Principal Investigator: Dr Christophe CORPECHOT</p>
<p>Background: Rare, symptomatic form of cholesterol intrahepatic <u>cholelithiasis</u> of genetic origin (mutations in ABCB4 gene, no other predisposition genes have been so far identified) affecting young subjects (< 40 years old). Mostly affects women between 20 and 40 and frequently associated with intrahepatic cholestasis of pregnancy. Both clinically and genetically heterogeneous, with unpredictable progression and prognosis and yet unknown prevalence.</p> <p>Rationale: this cohort will enable:</p> <ul style="list-style-type: none"> – to define the epidemiological basis of the disease and to characterize the clinical and prognostic spectrum of the disease and its response to UDCA therapy, – to establish a genomic DNA bank that will permit to characterize the genetic heterogeneity of the disease and discover new predisposition genes <p>It should allow to improve the diagnostic and therapeutic management of patients and identify new mechanisms involved in the pathogenesis of <u>cholelithiasis</u>.</p> <p>Primary objective: To describe the various clinical, biological and radiological manifestations of the LPAC syndrome and their evolution.</p>

RaDiCo-FARD

Full title: French National Cohort for the Evaluation of the Burden of Rare Skin Diseases
Principal Investigator: Pr. Christine BODEMER

Background: Rare skin diseases, namely inherited epidermolysis bullosa, ichthyosis, ectodermal dysplasia, incontinentia pigmenti, neurofibromatosis type 1, albinism, pemphigus, mucous membrane pemphigoid, and palmoplantar keratoderma, share several characteristics to which the patients and their families are subjected: a high impact in terms of quality of life, psychiatric morbidity and costs. The treatment can be heavy, costly and complex (dressings and topical treatments not reimbursed by the health insurance fund). The importance of assessment of the individual burden of each disease to assess disability (health-related Quality of Life), social integration, home-life, and use of medical resources

Rationale: This project aims at evaluating the individual burden of these rare skin diseases in the aim of reducing it, using medical and non-medical strategies (adaptation of the environment, reassessment of life style, efficient therapeutic education programmes etc.)

Primary objective: Prospective assessment of the individual burden of the 9 selected rare skin diseases to assess disability in the broadest sense of the term: psychological, social, economic and physical, for patients and/or families.

RaDiCo-IDMet

Full title: French National Cohort on Imprinting Disorders and their metabolic consequences
Principal Investigator: Pr Agnès LINGLART / **Co-Principal Investigator:** Pr Irène NETCHINE

Background: Group of rare human congenital diseases affecting approximately 7,500 - 10,000 patients in Europe. Caused by changes in both gene sequences (“genetic mutations” in imprinted genes) and gene regulation (“epigenetic mutation”) of imprinted genes. Imprinted genes influence glucose metabolism, nutritional behavior, obesity, pubertal development, bone health and skeletal growth, brain function and behaviour. Nine clinically recognized disorders: Silver Russell Syndrome; Temple Syndrome; Prader-Willi Syndrome (PWS); Transient Neonatal Diabetes type 1; Angelman Syndrome; Inactivating PTH/PTHrP Signaling Disorder or Maternal uniparental disomy of chromosome 20; Precocious puberty syndrome; Beckwith Wiedemann Syndrome (BWS). Historical studies focused on single IDs characterization rather than exploring common underlying issues and outcomes

Rationale: This project aims at bringing together scattered knowledge from these 9 IDs to unveil suspected common features with a focus on metabolism

Primary objective: To describe the natural history of imprinting disorders (IDs) according to their metabolic profile.

RaDiCo-ACOSstill

Full title: French National Cohort of Children and Adults with Still's Disease
Principal Investigator: Dr Sophie GEORGIN-LAVIALLE

Background: Adult-Onset Still Disease (AOSD) and Systemic-Onset Juvenile Idiopathic Arthritis (SoJIA, Still's disease beginning in childhood), are two rare multifactorial diseases associated with systemic inflammation. Incidence: 0.1/100 000 in adults and 0.6/100 000 in paediatric population. Prevalence: 1 to 3/100 000 in adult and 3/100 000 in paediatric populations (currently 500 to 1500 adult patients and 450 paediatric patients known in France).

Rationale: Comparison of paediatric and adult forms, better understanding of the pathogenic mechanisms of the disease, identification of early diagnostic / prognostic markers, possibility of encouraging the evaluation of new therapies would support a better management of Still's disease patients.

Primary objective: To describe the natural history of the disease in adult and paediatric populations.

ANNEXE 5 : Quelques cohortes ou dispositifs de suivis longitudinaux soutenus par d'autres canaux que le PIA

À côté des 14 cohortes soutenues par le PIA, un ensemble varié de dispositifs d'observation et de recherche à dimension longitudinale sont utilisés pour l'expertise et la recherche en santé, dont l'annexe 1 présente une brève typologie.

Ces dispositifs sont extrêmement divers (autres cohortes, enquêtes ou panels appariés aux bases médico-administratives, utilisations spécifiques de ces dernières...), portés par des acteurs variés et souvent dispersés. Ils s'articulent différemment selon les champs d'observation et les thématiques d'intérêt, et il ne peut être question d'en tracer ici un panorama complet.

L'objet de la présente annexe se limite à porter un regard rapide sur quelques outils structurants, mis en place à l'initiative des organismes publics ou fortement soutenus par eux, et dont l'articulation avec des cohortes épidémiologiques comme celles du PIA mérite d'être envisagée à l'avenir.

Des champs comme le développement et la santé de l'enfant ou le suivi des médicaments en vie réelle ont par ailleurs donné lieu, dans des rapports récents^{41,42}, à des analyses détaillées des données d'observation et de suivi disponibles, et ne sont pas directement révoqués ici.

1 Les enquêtes ou bases d'études en population générale

1.1 Les opérations du système statistique : l'enquête EHIS et le projet d'EDP Santé

1.1.1 L'enquête EHIS

L'enquête *European Health Interview Survey (EHIS)*, pilotée en 2019 par la Drees et l'Irdes, est l'une des grandes enquêtes de la statistique européenne, rendue obligatoire en application d'un règlement-cadre de 2008. Elle est effectuée tous les six ans dans 28 pays.

L'enquête française de 2019 est la troisième de vagues d'enquêtes réalisées dans des conditions différentes en 2008 (*via* l'enquête Handicap-santé), puis en 2014 (où elle prenait appui sur l'enquête ESPS de l'Irdes). En France, il s'agit désormais, à l'exception des Baromètres santé de Santé publique France (SpF)⁴³, de la seule grande enquête santé généraliste « non thématique » représentative de l'ensemble de la population vivant en ménages ordinaires, répétée à un intervalle régulier et selon un protocole stable ; elle est également réalisée en 2019 dans chaque département et région d'outre-

⁴¹ Conseil de l'enfance et de l'adolescence, *Des données et des études publiques mieux centrées sur les enfants*, juin 2019 et Haut Conseil de la santé publique, *Le dispositif statistique de surveillance et de recherche sur la santé de l'enfant*, octobre 2019.

⁴² Bernard Bégau, Dominique Polton et Franck von Lennep, *Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé- L'exemple du médicament*, Rapport réalisé à la demande de Madame la ministre de la santé, mai 2017.

⁴³ Ces Baromètres Santé sont des enquêtes transversales, répétées à intervalle régulier, et centrées sur les opinions, attitudes et comportements de la population résidant en France en matière de santé.

mer (Drom). L'échantillon visé est au total de 20 000 personnes (12 000 pour la métropole et 8 000 pour les Drom).

Conduite par voie de questionnaire, l'enquête EHIS aborde trois thèmes principaux : l'état de santé de la population (santé perçue, maladie chronique, santé mentale), le recours aux soins (consultation des professionnels de santé, dépistage, vaccination, renoncement aux soins) et les déterminants de la santé (caractéristiques sociodémographiques, activité physique, alimentation, consommations de tabac et d'alcool).

Elle n'a pas en tant que telle de dimension longitudinale, mais elle sera appariée avec le SNDS de façon à disposer d'une vision des consommations de soins sur 10 ans (5 ans en rétrospectif et 5 ans en prospectif), un appariement avec les données socio-fiscales de l'Insee étant également prévu.

1.1.2 Le projet d'EDP santé

L'échantillon démographique permanent (EDP) de l'Insee rapproche les informations de différentes sources administratives et enquêtes pour des individus nés certains jours de l'année (4 jours par an jusqu'au début des années 2000, puis 16 jours par an, soit 4,4 % de la population). Il résulte de l'appariement de cinq types de données :

- les bulletins d'état civil de naissance, de mariage, de décès depuis 1968 ;
- les données issues des cinq recensements exhaustifs (1968, 1975, 1982, 1990, 1999) et des enquêtes annuelles de recensement (EAR) depuis 2004 ;
- le fichier électoral donnant les inscriptions électorales actuelles et passées depuis 1990 ;
- les informations issues du panel d'actifs salariés depuis 1967 : salaires, durée de paie, catégorie socioprofessionnelle, secteur d'activité... ;
- les données socio-fiscales issues de la base Fideli (Fichier Démographique des Logements et des Individus) et de la base FiLoSoFi (Fichier Localisé Social et Fiscal) depuis 2011.

Le projet « d'EDP Santé » vise à associer à ces données celles du SNDS (données issues du SNIIRAM, du PMSI et du CépiDC, cf. annexe 6) pour construire un instrument d'étude longitudinal, réactualisé tous les ans. Il est considéré par la Drees comme complémentaire de l'enquête EHIS, réalisée tous les six ans sur un échantillon relativement réduit.

L'objectif est d'obtenir une source permettant de croiser des données socio-économiques détaillées et des données médico-administratives sur la santé et le recours aux soins, et d'enrichir ainsi les possibilités d'étude des inégalités sociales et territoriales de santé et d'accès aux soins. L'échantillon couvert sera suffisamment large pour des études sur des catégories sociales ou des territoires à un niveau fin, grâce aux informations contenues dans l'EDP sur la situation démographique, sociale et économique, ainsi que les trajectoires spatiales et professionnelles des personnes suivies dans l'échantillon.

La réception des données est fin 2019 et leur ouverture aux chercheurs sera organisée en 2020, selon des modalités à déterminer.

1.2 Une cohorte consacrée à la nutrition: l'étude Nutrinet

Le lancement de cohortes sur le sujet de la nutrition a été facilité à partir des années 1990 par l'automatisation possible de la collecte des données. La cohorte SUVIMAX qui a précédé Nutrinet a d'abord utilisé le Minitel, pour le suivi de 13 000 sujets pendant 8 ans, 6000 ensuite.

L'étude Nutrinet, lancée en 2009 auprès des personnes de plus de 18 ans, vise quant à elle une population beaucoup plus large en s'appuyant exclusivement sur l'outil Internet. Les participants renseignent à l'inclusion cinq types de questionnaires portant sur leur alimentation (3 enregistrements alimentaires d'une journée), leur mode de vie, leur activité physique, leur santé ainsi que leurs caractéristiques anthropométriques. Ils répondent ensuite à des questionnaires de suivi annuels, ainsi qu'à des questionnaires mensuels optionnels permettant de disposer d'informations complémentaires sur leur état nutritionnel, leur santé ou leur qualité de vie.

En 2019, 160 000 participants ont été inclus dans l'étude, mais celle-ci est ouverte de façon permanente : l'inclusion se poursuit pour constituer des échantillons les plus larges possibles, un nouvel appel à participation venant d'être lancé en juin 2019.

Les principales thématiques abordées sont la compréhension des comportements alimentaires, l'étude des relations entre nutrition et maladies chroniques et l'identification des facteurs de risque ou des protections qui peuvent exister en la matière. L'étude repose sur évaluation détaillée de l'exposition nutritionnelle et des comportements alimentaires, complétée par un appariement, déjà réalisé, avec les données du SNIIRAM et par le recours à une biobanque concernant 20 000 sujets. Elle est délibérément tournée vers l'éclairage des politiques de santé publique, qu'il s'agisse de la mise en évidence des risques liés aux aliments ultra-transformés, ou, de la protection induite par la consommation d'aliments Bio, ou, enfin, de la mise en place d'un système d'étiquetage nutritionnel simplifié (Nutri-Score).

L'étude repose sur l'équipe de recherche en épidémiologie nutritionnelle (EREN), UMR constituée entre l'Inserm, l'Inra, le Cnam et l'université Paris 13. Une collaboration « légère » (confrontation des résultats) existe avec la cohorte Constances, mais sans articulation explicite des deux projets, ni projets de recherche en commun.

L'étude Nutrinet n'a pas été retenue au titre des cohortes ou infrastructures financées par le PIA, mais bénéficie de soutiens de Santé publique France (SpF), du ministère de la santé, de Fondation sur la recherche médicale, ainsi que de financements sur projets (ANR, INCA et ANSES). À la différence de l'expérience précédente de Suvimax, les financements industriels n'ont pas été sollicités, car considérés comme potentiellement incompatibles avec l'indépendance scientifique de l'étude.

À l'avenir, les sujets du financement de la biobanque, de la banque de selles et de la « génobanque » à associer à la cohorte resteront posés, de même que son ouverture plus large aux chercheurs, qui n'est réalisée aujourd'hui qu'au travers de collaborations spécifiques.

2 Les études sur les liens entre santé et travail

Priorité reconnue dans le cadre des Plans santé travail, cette thématique donne lieu à plusieurs cohortes ou dispositifs d'étude longitudinale soutenus par les pouvoirs publics, sachant que cette dimension est aussi l'un des sujets d'intérêt principal de la cohorte Constances.

2.1 Les enquêtes statistiques sur les Conditions de travail

Le principal instrument d'observation promu par le système statistique public (Dares) est l'enquête Conditions de travail, qui a désormais une dimension à la fois instantanée et longitudinale. Réalisée tous les trois ans (2013, 2016 et 2019) sous forme de panel, réalimenté pour être représentatif, elle interroge tous les actifs occupés, soit à trois reprises, soit en les conservant au moins une fois dans l'échantillon lorsqu'ils ne sont plus en emploi : on dispose donc de trois interrogations pour toutes les personnes en emploi (quel que soit leur statut) et de deux ou trois, selon les cas, pour ceux qui sont entrés ou qui ont pris leur retraite pendant cette période de six ans.

Les effectifs enquêtés sont de l'ordre de 33 000, dont 20 000 répondants à une troisième interrogation, avec des taux d'attrition de l'ordre d'un quart.

Les questionnaires de l'enquête sont très détaillés sur les facteurs organisationnels et les expositions professionnelles (physiques et psychosociales), mais ils apportent des informations limitées sur les pathologies et les autres variables d'environnement ou de comportement qui caractérisent la situation des personnes enquêtées.

L'enquête est toutefois appariée avec le SNDS, ce qui permet le suivi des dépenses de santé pendant toute la période, y compris pour les retraités sortis de l'échantillon. En revanche, l'enquête n'est à ce stade appariée ni avec les sources fiscales et sociales, ni avec les DADS/DSN. La carrière et les expositions antérieures sont à cet égard retracées par voie déclarative, sur la base d'un mini-module retraçant les événements de carrière passés, sans mobilisation des données administratives.

2.2 Les cohortes COSET sur les actifs indépendants et agricoles : un complément direct à la cohorte Constances

Ces cohortes sont portées et financées sur son propre budget par l'agence Santé publique France, dans le cadre de ses missions de surveillance, avec le soutien des ministères chargés du travail et de l'agriculture.

Les deux cohortes COSET ont été envisagées dès le milieu des années 2000 comme de nature d'apporter un complément indispensable à la cohorte Constances, centrée sur les salariés du régime général, alors que les effets des problèmes de santé au travail sont souvent différés et qu'il faut pouvoir les suivre d'un emploi à l'autre et après la sortie d'activité.

Il s'agit de cohortes constituées à partir d'assurés de la MSA et de l'ex-RSI âgés de 18 à 65 ans et donnant lieu à une collaboration avec les deux régimes. Les répondants à l'auto-questionnaire d'inclusion sont sollicités chaque année pour répondre à un questionnaire de suivi sur leur santé et leur situation socioprofessionnelle, des appariements étant par ailleurs prévus avec les bases médico-administratives du SNDS.

Le projet vise la représentativité des personnes invitées, mais pas forcément celles des participants volontaires, dont le nombre a atteint 30 000 pour Coset-MSA et 20 000 pour Coset-Indépendants. Des échantillons de non participants suivis de façon « passive » ont été constitués comme dans le cas de Constances pour apprécier cette représentativité, sachant que les données de la MSA sont par ailleurs très riches en termes de suivi et de prestations sociales.

Contrairement à Constances, les cohortes Coset ont *a priori* vocation à être ouvertes, c'est-à-dire à ré-inclure de nouveaux entrants, à la fois pour tenir compte des nouvelles générations et améliorer la représentativité de l'ensemble, importante pour les missions de surveillance de SpF.

2.3 D'autres dispositifs spécifiques dans le domaine de la surveillance : les études Espri et Epinano

Toujours du côté de la surveillance, et placées sous la responsabilité de Santé publique France, deux cohortes spécifiques peuvent par ailleurs être mentionnées :

- la cohorte Espri, qui concerne depuis le milieu des années 2000 10 000 artisans retraités ayant été exposés à l'amiante, sur la base d'un suivi « passif » (SNDS) ;
- la cohorte Epinano qui a fait l'objet d'une demande spécifique de la Direction générale du travail (DGT), avec l'objectif de repérer de façon prospective d'éventuels problèmes de santé chez les travailleurs exposés aux nano-matériaux. Il s'agit d'une exploration particulière portant sur d'éventuels impacts sanitaires non connus à l'avance, et qui connaît des difficultés à recruter des entreprises participantes compte tenu des problèmes de mesure des nanoparticules au niveau des postes de travail.

2.4 La cohorte Agrican sur les risques de cancers liés aux activités agricoles

La cohorte Agrican fait partie d'un consortium international de cohortes sur l'agriculture (Agricoh) coordonné par le Centre international de recherche sur le cancer (CIRC). Elle est portée par une UMR associant l'Inserm et l'Université de Caen, en partenariat avec la MSA et le réseau FRANCIM des registres du cancer.

Après une première étude initiée en 1995, la cohorte Agrican a été mise en place en 2005-2007 et étendue à l'ensemble des agriculteurs et salariés agricoles (actifs ou retraités) des onze départements métropolitains disposant de registres spécialisés du cancer. Environ 180 000 personnes ont été incluses dans l'enquête et 90 000 suivies de façon effective jusqu'à la période 2015-2018, sur la base de trois types d'informations : l'histoire des expositions professionnelles agricoles détaillées par type de culture et type de tâches, l'état de santé (hors pathologies cancéreuses) et les comportements ou modes de vie (alcool, tabac, habitudes alimentaires...).

L'étude analyse les différences de prévalence des cancers chez les travailleurs agricoles et dans la population générale. Elle a notamment pour objet d'identifier les effets spécifiques de l'exposition aux pesticides et reçoit à ce titre des financements du ministère de l'agriculture au titre du plan Ecophyto, lequel a soutenu la mise en place d'une biothèque associée à la cohorte (avec mesure des imprégnations biologiques à certains pesticides).

Une deuxième cohorte, Depare, en est issue et comprend environ un millier de femmes ayant eu une grossesse depuis leur inclusion pour étudier d'éventuels troubles de la reproduction ou du développement des enfants nés de ces grossesses.

3 Les études sur le vieillissement et la dépendance

À la suite des cohortes historiques Paquid (3 800 personnes âgées suivies de 1998 à 2018) et Trois Cités (9 300 personnes âgées de 65 ans ou plus villes suivies à partir de 1999), et du côté statistique, des enquêtes Handicap-Incapacités-Dépendance (1998) et Handicap-Santé (2008, sans dimension longitudinale), le dispositif de suivi du vieillissement fonctionnel et cérébral et de ses conséquences en termes de perte d'autonomie s'appuie, outre les cohortes Constances et Memento, sur plusieurs instruments à portée longitudinale.

3.1 L'enquête Share

Share est un dispositif constitué sous forme de panel longitudinal, portant sur les personnes âgées de plus de 50 ans. Créé en 2004, il comprenait initialement 2000 personnes tirées dans l'échantillon maître de l'INSEE, la vague 8 de 2019 en réunissant 3000. Ces personnes de 50 ans et plus sont suivies tous les deux ans dans la durée, mais le panel est « rafraîchi » à chaque vague de façon à disposer d'un échantillon représentatif de cette population. La poursuite du projet est prévue au moins jusqu'à 2024, permettant un suivi des actifs âgés dans 28 pays européens sur une période de 20 ans.

Sa principale caractéristique est d'être, depuis la vague 7, réalisé de façon homogène et entièrement décidée dans un cadre européen, à l'exception de l'échantillonnage, qui s'adapte aux bases de sondage nationales. Les questionnaires sont établis et discutés par des chercheurs dans une optique pluridisciplinaire, et comprennent plusieurs volets. Ils sont débattus dans le cadre de réunions internationales, ce qui limite la possibilité de poser des questions afférentes à chaque contexte national, par exemple sur la couverture des soins de santé.

L'objectif de ce suivi, qui peut aller jusqu'au décès, est d'éclairer le passage du travail à la retraite et les processus de vieillissement, notamment du point de vue de la santé et des conditions de vie. Les questions portent ainsi sur la santé physique et psychique, la perte d'autonomie, le bien-être, les comportements à risque, le recours aux soins. Ils comprennent des questions sur les revenus, la consommation, les dépenses de santé, les pensions perçues, ainsi sur le logement, la composition du ménage, les relations sociales et intergénérationnelles. Les vagues 3 et 7 du panel ont par ailleurs intégré des volets rétrospectifs sur les histoires et conditions de vie dans l'enfance, et des interrogations complémentaires ont été ajoutées aux différentes vagues de l'enquête. En matière de santé, en dehors des questions habituelles sur l'état de santé et les limitations fonctionnelles, des données biologiques ont été recueillies dans la vague 6 (gouttes de sang séché), la masse grasse a été mesurée en vague 9, et la force de préemption est testée régulièrement en tant qu'indicateur de « fragilité », de même que la dépression.

Préalablement au PIA, le projet Share avait été financé par l'appel d'offres Cohortes lancé dans le cadre de l'IRESP. Il a ensuite été proposé en 2010-2011 à l'AAP Cohortes du PIA 2, mais non retenu. Reproposé dans la perspective d'un appariement avec le SNIIRAM, il n'a pas non plus été retenu au titre des infrastructures innovantes du PIA3.

L'Insee a été l'opérateur des trois premières vagues du panel, avant que l'Irdes récupère cette responsabilité au moment de la vague 4, puis la cède en 2012 à l'équipe LEDA de l'Université Dauphine, constituée en UMR avec le CNRS. L'administration des questionnaires, est sous-traitée à une entreprise de sondages, le projet d'appariement avec le SNDS ayant à ce stade été abandonné. Les cofinancements publics principaux sont ceux de la CNSA, auxquels s'ajoutent des participations financières de l'infrastructure PROGEDO, de la CNAV et du Conseil d'orientation des retraites (COR).

La mise à disposition des données s'effectue directement au niveau européen, dans un cadre d'un dispositif ouvert *d'open data*. L'ampleur du projet pose la question de sa gestion dans la durée par une structure de type universitaire dotée de moyens et d'effectifs permanents limités, alors qu'il avait été porté initialement par le système statistique public, puis par l'Irdes.

3.2 Le dispositif statistique public de suivi du vieillissement et de la dépendance

Ce dispositif repose sur des données administratives longitudinales d'origine plurielle, appariées entre elles, et complétées par des panels ou, plus souvent, par des enquêtes statistiques transversales.

Dans le champ du vieillissement, la Drees met ainsi en place une base de données, dite Échantillon National d'Études sur l'Autonomie et l'hébergement des Seniors (ENEAS), issue des remontées de l'allocation personnalisée d'autonomie (APA) et de l'aide sociale à l'hébergement (ASH), appariées avec les informations issues de l'état civil (mortalité), les consommations de soins collectées par le SNDS, les allocations logement, et les données socio-fiscales.

Cette approche a vocation à s'étendre à l'avenir à l'ensemble des aides sociales dans le cadre d'un « échantillon transversal des aides et de la protection sociale » (ETAPS), qui comportera des dimensions longitudinales à la fois rétrospectives et prospectives. Cela devrait permettre des rapprochements avec les échantillons longitudinaux de retraités et de cotisants aux régimes de retraites (EIR et EIC), et donc, au-delà des trajectoires des personnes âgées dépendantes, une meilleure connaissance de l'entrée dans la perte d'autonomie.

Les enquêtes statistiques effectuées ou en projet s'inscrivent dans une perspective complémentaire à ces projets. S'agissant des enquêtes périodiques en population générale, l'enquête CARE menée en 2015 sur le champ des plus de 60 ans a ainsi interrogé 15 000 seniors vivant à domicile et 3000 en institution ; elle est enrichie par des appariements avec toute une série de données administratives, concernant les ressources fiscales et sociales (RFS) des ménages, les prestations sociales, familiales et vieillesse qu'ils perçoivent, l'APA, leurs consommations de soins (à partir du SNIIRAM), et les équipements de leur zone de résidence.

L'enquête « Autonomie » prévue en 2020 renouvellera l'opération en revenant, comme en 2008, à une approche globale du handicap et de la dépendance visant à décrire l'état de santé et les conditions de vie de l'ensemble des personnes ayant des problèmes d'autonomie.

4 Les dispositifs de suivi longitudinal des patients

Ces dispositifs sont extrêmement nombreux et, indépendamment même du suivi des traitements en vie réelle conduits par les laboratoires pharmaceutiques ou placés sous l'égide des autorités sanitaires, ils peuvent être portés par des agences comme l'ANRS ou l'INCA, par des opérateurs comme la Cnam, par des équipes de recherche hospitalo-universitaires (qui sont souvent mixtes avec des unités Inserm) ou par les établissements de santé eux-mêmes et leurs regroupements (Unicancer).

La diversité et la dispersion de ces dispositifs d'observation à dimension longitudinale rendent difficiles d'en avoir une vision globale, y compris par pour chaque type de pathologie, mais, au-delà des cohortes financées par le PIA, quelques-uns de ceux promus par les grands opérateurs publics méritent d'être signalés, permettant d'illustrer la variété et la complexité du panorama.

4.1 Des cohortes de patients vivant avec le VIH ou une hépatite virale sous l'égide directe de l'ANRS

L'ANRS a investi depuis sa création dans la mise en place de cohortes de personnes vivant avec le VIH et/ou les hépatites virales, dont elle porte directement la gestion et le financement en son sein. Outre la cohorte Hepather, soutenue par le PIA, plus d'une dizaine de cohortes ont ainsi été mises en place, et évaluées en 2018.

4.1.1 Des cohortes d'adultes vivant avec le VIH

- La plus ancienne est la cohorte ANRS CO3 Aquitaine. Initiée en 1987, elle a inclus plus de 10 000 patients infectés par le VIH-1 et suivis en milieu hospitalier dans le sud-ouest de la France. Son objectif principal est d'étudier les tendances temporelles de l'infection par le VIH en région Aquitaine, en s'appuyant notamment sur une biothèque.
- Une importante base de données hospitalière a été développée sur l'infection par le VIH à travers la cohorte ANRS CO4 FHDH ; initiée en 1989, elle incluait en 2018 plus de 183 000 participants à partir des données de 125 centres hospitaliers. Ses objectifs principaux portent sur l'étude des stratégies thérapeutiques, de la morbi-mortalité ainsi que sur l'impact de l'infection et/ou des traitements.
- La cohorte ANRS CO5 VIH-2 a pour objet d'étudier l'infection par le VIH-2, notamment la progression de cette infection et ses caractéristiques, ainsi que la réponse au traitement antirétroviral ; lancée en 1994, elle a inclus plus de 1 100 patients.
- La cohorte CO6 PRIMO concerne quant à elle les primo-infectés : initiée en 1996, elle a 2 200 patients vivant avec le VIH-1 depuis moins de trois mois ; son objectif principal est d'améliorer les connaissances sur la primo-infection et d'apporter des informations sur l'impact des traitements précoces.
- Deux autres cohortes concernent de patients porteurs du VIH à profils particuliers, soit qu'ils soient asymptomatiques depuis au moins 5 ans (cohorte ANRS CO21 Codex), soit qu'ils aient en même temps un cancer traité par anticorps (ANRS CO24 OncoVIHAC).

4.1.2 Des cohortes pédiatriques

Elles visent notamment à :

- décrire l'évolution au cours du temps du taux de transmission mère-enfant du VIH et de sa prévention et identifier les facteurs de risque de transmission (cohortes ANRS CO1);
- suivre les enfants infectés avant 15 ans, ainsi que la transition vers l'âge adulte et le devenir à long terme de ces jeunes infectés pendant l'enfance (cohorte ANRS CO25).

4.1.3 Une cohorte de patients co-infectés par le VIH et le VHC

Initiée en 2005, la cohorte ANRS CO13 HEPAVIH a inclus près de 1900 patients co-infectés par le VIH et le virus de l'hépatite C (VHC). Elle vise à préciser l'histoire naturelle de la co-infection ainsi qu'à mieux comprendre, à partir d'un suivi tous les six ou douze mois, les interactions entre les deux virus et leurs traitements.

4.2 Des cohortes ou projets de plateformes sur les maladies chroniques

4.2.1 L'étude ENTRED sur les patients diabétiques réalisée par Santé publique France

L'étude Entred mise en œuvre par SpF est réalisée en partenariat avec l'Assurance maladie, la HAS et l'ANSM, à partir de fonds exclusivement publics.

Elle porte sur les personnes âgées de 18 ans ou plus et traitées pharmacologiquement pour un diabète. L'édition 2019 est la troisième d'une série d'enquêtes menées en 2001 et en 2007 : elles recueillent des informations d'ordres épidémiologique, médical, économique et social ainsi que sur la qualité des soins, visant à améliorer les actions de prévention concernant les pratiques médicales, l'état de santé et la qualité de vie des diabétiques. L'enquête 2019 explore aussi des thématiques comme la littératie en santé, le recours aux soins ou l'adhésion aux traitements ; elle inclut des diabétiques de type 1 et des résidents des départements d'outre-mer.

Réalisée sur un échantillon aléatoire de 9 000 personnes diabétiques en métropole et 4 000 dans les Drom, cette enquête est de nature transversale (auto-questionnaires auprès des patients et des médecins-soignants), mais elle est complétée par la constitution d'une cohorte passive par extraction des données médico-administratives (consommation médicale, données d'hospitalisation, statut vital et causes médicales de décès), et ce sur une période de 20 ans (10 ans avant le tirage de l'échantillon et 10 ans après).

4.2.2 Le projet d'infrastructure de suivi de e-cohortes de patients atteints d'une maladie chronique porté par l'AP-HP

Le projet ComPaRe (Communauté des patients pour la recherche) est une infrastructure de recherche reposant sur l'AP-HP (centre d'épidémiologie clinique de l'Hôtel Dieu) et destiné à recruter et suivre une e-cohorte de patients adultes atteints d'une ou plusieurs maladie(s) chronique(s). Il est financièrement soutenu par l'AP-HP, l'Université de Paris-Descartes et l'Université Sorbonne Paris cité et bénéficie d'un ensemble assez large de partenariats associatifs.

Les patients seront recrutés sur quatre ans et suivis pendant dix ans, l'objectif d'inclusion étant de 100 à 200 000 patients, dont 33 000 déjà recrutés en 2019. L'idée directrice est de disposer de données plus proches du quotidien réel des patients, en sollicitant leur participation et leur implication.

Ce projet est fondé sur une plateforme accessible par Internet, outil de travail des équipes de recherche impliquées dans l'étude, et visant notamment à mettre à disposition des auto-questionnaires pour le recueil de données. La plateforme se propose ainsi de simplifier les procédures de recrutement et la logistique des études réalisées, notamment par les équipes hospitalo-universitaires.

L'utilisation de la plateforme pourra être plurielle, des thématiques et questionnaires autour des maladies chroniques pouvant être ajoutés tout au long de l'étude par les porteurs du projet ou des équipes extérieures, des sous-cohortes par pathologie et des essais d'intervention nichés étant également envisagés.

4.3 Des dispositifs de suivi des patients traités pour un cancer

Alors que quatre des cohortes soutenues par le PIA suivent des patients pris en charge pour une forme de cancer, et qu'une réflexion est en cours sur les registres, les initiatives sont multiples en ce domaine, notamment en milieu hospitalo-universitaire⁴⁴, dont certaines veulent avoir une portée fédératrice.

⁴⁴ Par exemple la cohorte Melbase portée par l'AP-HP qui est une cohorte clinico-biologique prospective, multicentrique, incluant environ un millier de patients présentant un mélanome métastatique pour un suivi de 3 ans.

4.3.1 La plateforme de données longitudinales en cours de développement à l'Inca

L'Institut national du cancer (Inca) a entrepris la constitution d'une plateforme de données en cancérologie (PDC), qui s'appuie sur les données individuelles issues de bases comme le SNDS et les registres des cancers, mais qui vise aussi à regrouper les données à visée de soins collectées au niveau local dans les « dossiers communicants de cancérologie » (DCC), lesquels contiennent des renseignements sur le patient, le diagnostic anatomopathologique, génétique et immunitaire de la tumeur, les propositions thérapeutiques et le parcours de soins. Cette plateforme pourrait ensuite être enrichie d'autres sources (tumorothèques, biothèques...).

Le socle de la PDC repose sur une « Cohorte cancer », extraction du Système national des données de santé (SNDS), qui rassemble les données médico-administratives de toutes les personnes atteintes d'un cancer identifiées depuis 2010 (8 millions de personnes), et ce avec un horizon de 25 ans à compter de leur inclusion. Ces données sont en cours d'appariement avec celles de 26 registres des cancers, qui fournissent une information sur le type morphologique et histologique du cancer, et dans certains cas, sur son stade au diagnostic. Les données des centres de coordination des dépistages sont appelées à compléter ce socle par des informations sur la participation de la population aux dépistages des cancers du sein, du col de l'utérus et colorectaux.

L'objectif de l'Inca est de colliger également, sous une forme dématérialisée et interopérable, les données à visée de soins collectées au niveau local dans les DCC : résumés des réunions de concertation pluridisciplinaire (RCP), comptes rendus d'anatomo-cytopathologie et de génétique moléculaire, programmes personnalisés de soins (PPS) pendant la phase aiguë du cancer et programmes personnalisés de l'après cancer (PPAC) ensuite, ce qui nécessite un travail important d'harmonisation et de modernisation des systèmes d'information professionnels.

Ce projet de plateforme se veut potentiellement ouvert à l'ensemble des chercheurs de divers domaines disciplinaires en leur proposant une offre de services intégrée dans la perspective d'analyses longitudinales (cf. annexe 6).

4.3.2 Le programme ESMÉ développé par Unicancer pour l'analyse des données de prise en charge dans les CLCC

Unicancer (fédération hospitalière et groupement de coopération sanitaire regroupant les 18 centres de lutte contre le cancer) a construit à partir de 2013 un outil de suivi des prises en charge concernant les cancers du sein métastatiques, des ovaires et du poumon. Il s'agit d'une base de données appariant les données du PMSI et celles des établissements, qui fait l'objet d'un suivi rétrospectif « passif », sans questionnaires auprès des patients. Ces données sont à la fois massives et de qualité (contrôles), ouvrant des perspectives d'utilisation pour les études et recherches épidémiologiques.

Le programme ESMÉ (« Épidémio-Stratégie Médico-Economique ») vise ainsi à structurer dans un format commun les données existantes des CLCC et des hôpitaux participants, pour pouvoir les analyser dans une perspective rétrospective longitudinale. Trois plateformes de données ont été constituées pour le cancer du sein métastatique, le cancer de l'ovaire et le cancer du poumon. Pour chaque pathologie, la plateforme s'appuie sur trois sources d'information : une base « patient », centralisant les données recueillies dans les dossiers patient informatisés (DPI); une base « traitement », centralisant celles relatives aux dispensations pharmaceutiques ; une base « séjour », centralisant la description des séjours hospitaliers. ESMÉ « poumon » connaît un développement particulier, dans la mesure où il prend appui sur un réseau de 40 établissements incluant des CHU au-delà des CLCC, et ce dans un contexte d'apparition de nouvelles thérapies ciblées.

Le financement du programme ESMÉ repose sur un modèle de partenariats public-privé (soutiens de Roche, Pierre Fabre, Pfizer, Astra Zeneca, MSD...). La base de données se veut toutefois à la disposition de la communauté scientifique, dans le cadre d'une gouvernance s'affichant comme indépendante des financeurs.

4.4 Des cohortes de consommateurs de soins développées par la Cnam à partir du SNDS, en partenariat éventuel avec l'ANSM et la HAS

L'assurance maladie a enfin développé ses propres cohortes ou simili-cohortes à partir des données du SNDS relatives aux consommations de soins, à des fins d'étude interne. Ces cohortes sont souvent closes rapidement à la fin de l'étude, mais certaines ont été conservées afin de faciliter leur réutilisation, notamment pour apprécier le devenir à moyen terme de certains patients.

L'objet principal de ces « cohortes » est d'analyser les usages des médicaments et des produits de santé, à la fois dans une perspective de régulation et de pharmaco-vigilance. Dans ce dernier domaine, a été mis en place en 2018, suite au rapport Bégau-Polton-Von Lennep sur le suivi des médicaments en vie réelle⁴⁵, un GIS Expertise publique en épidémiologie des produits de santé (Epiphare) associant la Cnam et l'ANSM: il a pour mission de piloter des études épidémiologiques en vie réelle sur les produits de santé, indépendantes de l'industrie pharmaceutique et de réaliser des études d'utilisation, ciblant les mésusages, les populations fragiles, les nouveaux médicaments et l'impact des mesures d'encadrement ou d'amélioration des pratiques.

D'autres cohortes de l'assurance maladie suivent des groupes de patients avec plusieurs entrées :

- des entrées du type « évènements de santé », tels les opérations de chirurgie bariatrique, l'implantation de prothèses de la hanche ou du genou, la prise d'anti-dépresseurs, d'antibiotiques ou d'anti-hypertenseurs, la prise en charge des pathologies de la prostate... ;
- des entrées liées aux situations médico-administratives, s'agissant des patients en affection de longue durée ou bénéficiaires de la CMU-C.

L'analyse des « parcours des patients », en partenariat avec la Haute autorité de santé (HAS), est en outre une priorité pour les années à venir au regard de la stratégie « Ma santé 2022 », avec un programme portant sur des sujets comme la bronchopneumopathie chronique obstructive (BPCO), l'obésité, les maladies coronaires, la prise en charge post-accident vasculaire cérébral, l'épilepsie, la dégénérescence maculaire liée à l'âge (DMLA), le reflux gastro-œsophagien (RGO), le diabète ou les troubles bipolaires.

⁴⁵ Bernard Bégau, Dominique Polton et Franck von Lennep, *ibid.*

ANNEXE 6 : Les potentialités liées au système national des données de santé (SNDS) et aux plateformes de données de santé en constitution

Les bases de données médico-administratives ont des interactions potentielles importantes avec les cohortes, qu'il s'agisse d'en extraire directement des cohortes dites « passives », de servir de base d'échantillonnage ou témoin, de dispenser les cohortes existantes de certains recueils d'informations ou d'enrichir les informations qu'elles collectent (*cf.* annexes 1 et 5).

La principale de ces bases en France est le système national des données de santé (SNDS), institué par la loi du 26 janvier 2016 de modernisation du système de santé, qui constitue une richesse importante pour les études et la recherche en santé, mais dont l'appariement avec les cohortes épidémiologiques est resté, à l'exception notable de ELFE et de CONSTANCES, jusqu'ici assez peu répandu, en raison notamment de la longueur des procédures et d'un cadre juridique changeant, concernant notamment le stockage sécurisé des données.

Par ailleurs, ce système est en cours d'évolution vers une plateforme de données de santé (*Health data hub*), juridiquement constituée en novembre 2019, visant à englober un ensemble de plus en plus large d'informations d'origine diverse (dossiers médicaux, résultats d'analyses biologiques et d'imagerie...), et à proposer des services aux chercheurs en matière d'appariement, de stockage sécurisé et de mise à disposition des données.

Ce dispositif national se double, du côté de l'Inserm, de la volonté de mettre en place, à travers le projet France Cohortes, une plateforme de stockage et de traitement destinée à offrir aux équipes des cohortes des solutions d'appariement et de stockage sur un site informatique propre, tandis que des initiatives existent aussi, par exemple du côté de l'Inca, pour mettre en place des bases de données à portée plus spécifique et plus réduite.

La présente annexe vise à présenter un panorama rapide de ces divers projets, qui nécessiteront des clarifications sur leur rôle respectif pour que les responsables des cohortes sachent où trouver des interlocuteurs compétents et bénéficier d'un appui pour tirer le meilleur parti possible des données susceptibles d'être combinées ou articulées avec celles qu'ils recueillent.

1 Le cadre actuel offert par le SNDS : des possibilités d'exploitation et d'appariements, des contraintes en termes de stockage sécurisé des données

1.1 Présentation du système national des données de santé (SNDS)

1.1.1 Le développement du SNDS

La France a construit, depuis une quinzaine d'années un entrepôt de données médico-administratives qui constitue un patrimoine potentiel important pour les politiques de régulation, l'expertise et la recherche.

Partant d'un côté du système national inter-régimes de l'assurance maladie (SNIIRAM) mis en place en 1998, et de l'autre du programme de médicalisation des systèmes d'information (PMSI), cet entrepôt a été constitué dès 2003, avec un chaînage⁴⁶ entre soins ambulatoires et soins hospitaliers opéré depuis 2006 sur données pseudonymisées.

Les usages de cet entrepôt s'étaient d'ores et déjà développés avant 2016, à la fois par l'assurance maladie, les agences sanitaires (ANSM pour la pharmaco-surveillance, INVS pour la surveillance épidémiologique) et, dans une moindre mesure, par la recherche.

Suite à des débats sur la complexité de la gouvernance du système et ses difficultés d'accès⁴⁷, la loi de modernisation du système de santé (LMSS) du 26 janvier 2016 a modifié et élargi ce dispositif en prévoyant que « *les données de santé à caractère personnel recueillies à titre obligatoire et destinées aux services ou aux établissements publics de l'État ou des collectivités territoriales ou aux organismes de sécurité sociale peuvent faire l'objet de traitements à des fins de recherche, d'étude ou d'évaluation présentant un caractère d'intérêt public, dans le respect de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés* ». Les finalités de ces traitements recouvrent notamment la recherche, les études, l'évaluation et l'innovation dans les domaines de la santé et de la prise en charge médico-sociale.

Les traitements associés au SNDS ainsi constitué, dont le ministère chargé de la santé pilote les orientations stratégiques, sont placés sous la responsabilité de la Cnam (en collaboration avec l'ATIH pour les données hospitalières et l'Inserm pour les causes de décès), tandis qu'un Institut national des données de santé (INDS), constitué sous forme de GIP où sont représentés tous les producteurs et les familles d'utilisateurs, est chargé d'accompagner les utilisateurs et de faciliter l'accès aux données.

La loi tente à cet égard d'avancer dans le sens d'un périmètre de données étendu (*cf. infra*), d'accès élargis (tant permanents qu'en termes de finalités) et d'une simplification du circuit des demandes d'accès au travers d'un « guichet unique » ; elle vise aussi à faciliter les procédures devant la Commission nationale de l'informatique et des libertés (CNIL).

1.1.2 Le contenu du SNDS

Le SNDS rassemble dans un entrepôt national trois blocs d'informations individuelles de santé d'origine administrative pseudonymisées (*cf. graphique 1*) :

- les données du SNIIRAM ;
- les données des hôpitaux et autres établissements de santé ;
- les données relatives aux causes de décès.

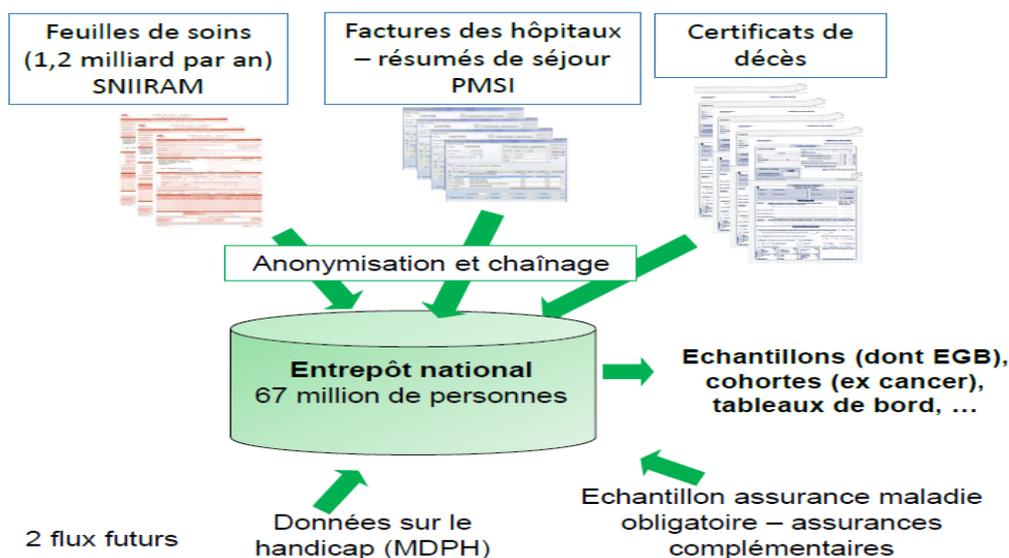
Ces données sont conservées pendant une période de 20 ans.

Le système prévoit aussi d'intégrer à l'avenir les données médico-sociales des maisons départementales des personnes handicapées (MDPH) et un échantillon représentatif des données de remboursement des organismes d'assurance maladie complémentaire

⁴⁶ Le chaînage consiste à relier entre elles des données issues de sources différentes pour constituer un ensemble d'observations se rapportant à un même individu.

⁴⁷ Pierre-Louis Bras, Rapport sur la gouvernance et l'utilisation des données de santé, septembre 2013.

Graphique 1 : Contenu du SNDS



Source : Dominique Polton, présidente de l'INDS

Les données ainsi constituées comprennent notamment des informations :

- sur les patients : âge, sexe, lieu de résidence accès à la CMU-C, reconnaissance d'une affection de longue durée (ALD) ; date et cause éventuelles de décès ;
- sur les soins de ville : détail des prestations remboursées intégrant le codage des actes (médicaux et de biologie), les médicaments et les dispositifs médicaux ;
- sur les soins hospitaliers caractéristiques des séjours dans la nomenclature du PMSI (motifs médicaux et actes), durées de séjour et modes de sortie, consultations externes.

Ces données administratives sont, selon les cas, plus ou moins à même de repérer les pathologies traitées : des informations existent sur celles-ci dans le cas des ALD et des soins hospitaliers, et peuvent être retrouvées de façon indirecte lorsqu'existent des médicaments ou des actes « traceurs » caractérisant spécifiquement certaines maladies.

Des travaux sont à cet égard réalisés, dans le cadre d'un réseau dit ReDSiam (réseau pour mieux utiliser les données du SNDS)⁴⁸, dont la coordination est assurée par l'UMS Inserm-UVSQ « Cohortes épidémiologiques en population » pour progresser sur le repérage, la validité et la fiabilité de l'identification des pathologies dans l'utilisation des données médico-administratives.

Toutefois, le système ne comprenait pas jusqu'ici de données cliniques ou paracliniques permettant de caractériser finement l'état médical des patients traités (compte rendus d'examen cliniques, résultats de prélèvements biologiques ou données d'imagerie). Il ne comprend pas non plus

⁴⁸ Le cœur de l'activité de ReDSiam est constitué par le développement, la discussion critique et la mise à disposition d'algorithmes spécifiques de pathologies basés sur les données du SNIIRAM par des combinaisons de diagnostics du PMSI et/ou d'ALD, de données sur les actes médicotechniques, les consommations de médicaments ou d'autres produits de santé plus ou moins spécifiques de la pathologie concernée, les consultations de professionnels,.... Ses activités reposent essentiellement sur le fonctionnement de groupes de travail thématiques par grands domaines pathologiques (appareil cardiovasculaire, appareil locomoteur, appareil respiratoire, endocrinologie, nutrition et métabolisme, handicap, maladies infectieuses, maladies neurodégénératives, maternité-périnatalité, polypathologie et multimorbidité, troubles mentaux et du comportement, tumeur), ainsi que de groupes de travail transversaux (méthodologie, qualité des soins).

d'informations sur les antécédents ou les facteurs de risques propres aux personnes concernées, non plus que (mis à part l'accès à la CMU-C) sur leur situation sociale.

L'intérêt d'appariements avec des données de cohortes comportant un phénotypage fin des sujets suivis aux différents stades de leur(s) pathologie(s), ainsi qu'avec des données sociales issues de différentes sources et avec des appréciations recueillies auprès des patients eux-mêmes sur leurs difficultés et leurs conditions de prise en charge apparaît donc clairement pour les études et la recherche en santé.

1.2 L'accès des cohortes au SNDS et le stockage sécurisé des données appariées

1.2.1 Les procédures d'accès au SNDS

Les principes posés par la loi de 2016 pour l'accès aux données du SNDS sont que toute personne ou structure, publique ou privée, à but lucratif ou non lucratif, peut accéder aux données du SNDS sur autorisation de la CNIL, en vue de réaliser une étude, une recherche ou une évaluation présentant un strictement encadré, avec un contrôle concernant l'absence de finalités interdites par la loi (promotion commerciale des produits de santé, sélection des risques ou tarification au risque dans les contrats d'assurance).

Les demandes sont à déposer auprès de l'Institut National des Données de Santé (INDS). Ce dernier est le point d'entrée unique des demandes d'accès au SNDS, avec pour mission d'assurer un traitement des demandes conformément au cadre réglementaire, dans les délais définis par la loi, et d'accompagner les utilisateurs du SNDS dans la prise en main des données.

L'INDS est en lien direct avec un Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé (CEREES), mis en place en mai 2017, et chargé d'examiner les demandes du point de vue méthodologique, afin de fournir un avis à la CNIL sur la cohérence entre la finalité de l'étude proposée, la méthodologie présentée et le périmètre des données auxquelles il est demandé accès. Cet avis vise à favoriser et raccourcir l'examen du dossier par la CNIL, de qui continue à dépendre l'autorisation effective du traitement considéré.

Pour certaines organisations chargées d'une mission de service public, une procédure spécifique d'accès permanent au SNDS est prévue. C'est par exemple le cas pour l'agence Santé publique France, l'Agence nationale de sécurité du médicament et des produits de santé, la Haute Autorité de santé, les chercheurs des CHU, des Centres de lutte contre le cancer, l'Agence de la biomédecine et l'Inserm.

Les projets demandant des appariements impliquant l'utilisation du numéro d'inscription au répertoire national d'identification (NIR) donnent lieu à une procédure particulière : nécessitant initialement, aux termes de la loi de 1978 relative à l'informatique, aux fichiers et aux libertés, un décret en Conseil d'État, ils font, depuis la loi du 7 octobre 2016 pour une République numérique, l'objet d'une procédure d'autorisation spécifique : s'agissant des projets de recherche scientifique publique, les appariements se font désormais à partir d'une « clé d'appariement non signifiante », code obtenu par une opération cryptographique spécifique à chaque projet de recherche⁴⁹.

⁴⁹ Antoine Bozio et Pierre-Yves Geoffard, *L'accès des chercheurs aux données administratives, État des lieux et propositions d'action*, Rapport du groupe de travail du CNIS au Secrétaire d'État chargé de l'industrie, du numérique et de l'innovation, mars 2017.

Le circuit standard pour les chercheurs est celui d'un dépôt de la demande en ligne à l'INDS (7 jours), suivi d'un avis du CEREES dans un délai d'un mois, suivi ensuite d'une autorisation de la CNIL, elle-même en principe dans un délai de deux mois.

Même si ces délais sont encadrés, les changements répétés de la législation entre 2016 et 2019 (intégrant en 2018-2019 la mise en conformité de la législation française « Informatique et Libertés », au cadre européen fixé par Règlement général sur la protection des données – RGPD -, qui retient une définition large de la notion de données de santé⁵⁰) ont conduit nombre de responsables de cohortes à retarder le dépôt des dossiers d'appariement projetés ou à devoir reprendre plusieurs fois la procédure de soumission de leurs projets.

1.2.2 Les implications et les pratiques en termes de stockage sécurisé des données

Ces changements réglementaires ont aussi des implications liées au fait que, d'après la loi de janvier 2016 « *l'accès aux données s'effectue dans des conditions assurant la confidentialité et l'intégrité des données et la traçabilité des accès et des autres traitements, conformément à un référentiel de sécurité défini par arrêté* ».

Ce référentiel de sécurité a été défini par un arrêté du 22 mars 2017. Celui-ci exige, pour tous les systèmes mettant à disposition des données potentiellement « ré-identifiantes » constituées à partir du SNDS, un cadre homologué nécessitant une étude préalable (analyse de risques, recettes et de tests assurant la bonne couverture des risques, homologation de sécurité sur le périmètre considéré, suivi opérationnel de la sécurité du système d'information). Les données non anonymisées doivent en particulier être stockées dans un environnement maîtrisé, dont l'accès par des utilisateurs doit se faire à partir d'un poste respectant des exigences particulières en matière de sécurité des systèmes d'information.

Aux termes de cet arrêté, les nouveaux systèmes d'information devaient être en conformité avec le référentiel de sécurité dès leur création, et les systèmes existants devaient procéder à cette mise en conformité dans un délai de deux ans, à savoir au plus tard en mars 2019.

Confrontées à cet impératif, les cohortes du PIA ont eu des pratiques diverses : attente des solutions susceptibles d'être proposées dans le cadre des projets France Cohortes ou du *Health data Hub* (*cf. infra*), utilisation potentielle de serveurs homologués au sein des établissements de santé ou des agences sanitaires, recours au Centre d'accès sécurisé aux données (CASD) pour héberger et mettre à disposition des chercheurs les données appariées dans des conditions sécurisées (*cf. encadré*).

⁵⁰Les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne *cf.* site de la CNIL, <https://www.cnil.fr/fr/quest-ce-que-une-donnee-de-sante>

Le Centre d'accès sécurisé aux données (CASD)

Projet initié par l'Insee dès 2010, et financé dans le cadre du premier appel à projets Equipex du PIA1 à hauteur de 4 M€, le CASD est un groupement d'intérêt public rassemblant l'État, représenté par l'Insee, le GENES, le CNRS, l'École polytechnique et HEC Paris, constitué par arrêté interministériel en décembre 2018.

Ce GIP a pour objet principal d'organiser et de mettre en œuvre au bénéfice des chercheurs publics des services d'accès sécurisé pour les données confidentielles, à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation, Il a également pour mission de valoriser la technologie développée pour sécuriser l'accès aux données dans le secteur privé.

Le CASD est un Equipex, dont le financement par l'ANR vient à échéance fin 2019, et qui participe à la très grande infrastructure de recherche (TGIR) PROduction et GEstion des DONnées – PROGEDO, chargée d'impulser et structurer une politique publique des données pour la recherche en sciences sociales.

Ses fonctions sont accomplies par le biais d'une infrastructure informatique centrale garantissant la confidentialité des données et d'un boîtier qui permet d'accéder à distance, sur la base d'une identification biométrique et d'une carte à puce, à une bulle sécurisée. Des moyens de calcul (logiciels et puissance de calcul) y sont aussi disponibles, permettant aux chercheurs de travailler sur les données et d'établir des résultats intermédiaires.

Le CASD a par ailleurs la possibilité de réaliser certains appariements lui-même en fonction des demandes qui lui sont adressées, en jouant le rôle de tiers de confiance sur des données comme le NIR. Il se donne aussi pour objectif d'assister les chercheurs dans les procédures d'habilitation vis-à-vis de la CNIL, d'accompagner la définition et la construction de leurs projets d'appariement et d'appuyer la documentation des bases de données constituées.

L'accès aux données via le CASD concerne à 95% à la recherche publique. L'infrastructure est financée, outre les apports des partenaires institutionnels, au moyen de redevances payées par les chercheurs, à partir des financements obtenus pour leurs projets.

Le CASD met aujourd'hui à disposition des données individuelles en provenance de l'INSEE, des ministères chargés de la justice, de l'éducation nationale, de l'agriculture ainsi que de l'économie et des finances s'agissant des données fiscales, sachant que pour ces dernières, l'accès doit obligatoirement s'effectuer par le biais de cet organisme.

Suite à la loi de 2016 exigeant un référentiel de sécurité pour l'accès aux données de santé, le CASD a été parmi les premières plateformes homologuées. L'accès aux données PMSI de l'ATIH est possible par son entremise, et il accueille également depuis 2019 les données de la cohorte CONSTANCES, dans le cadre d'une convention conclue avec l'Inserm (*cf. infra*). Il a aussi en perspective l'hébergement de bases de données issues de la Drees et de la Dares en vue de leur mise à disposition aux chercheurs.

Ces accès *via* le CASD donnent lieu à une sécurisation et à une traçabilité très rigoureuses, adaptées en particulier aux données fiscales, mais qui conduisent certains chercheurs à en déplorer la complexité ainsi que le coût qui leur est directement imputé.

Alors qu'il avait été à un moment envisagé d'intégrer l'Inserm dans le consortium gouvernant le CASD, celui-ci a finalement préféré, compte tenu de ses propres projets (*cf. infra*) et de l'échéance réglementaire de mars 2019 pour l'application du référentiel de sécurité afférent aux données médico-administratives, conclure avec le CASD une convention permettant une solution rapide, et le cas échéant réversible, en la matière. Cette convention, signée en juillet 2019, est principalement utilisée pour le stockage sécurisé et la mise à disposition des données issues de la cohorte CONSTANCES, appariée avec le SNDS.

2 Les perspectives et questions liées aux autres plateformes en cours de constitution

2.1 La Plateforme des données de santé, dite Health Data Hub

2.1.1 Le nouveau cadre mis en place en 2019

Au-delà des informations figurant dans le SNDS, dont les limites ont été précédemment mentionnées, un ensemble très riche de données de santé, beaucoup plus détaillées en termes de contenu médical (examens cliniques, résultats des prélèvements biologiques, données d'imagerie...), figurent dans les dossiers médicaux numérisés, constitués par les différents acteurs du système de santé. Les enjeux d'une utilisation secondaire de ces données concernent à la fois la prise en charge individuelle des patients, les politiques de régulation, la surveillance sanitaire et la recherche en santé.

Après que le Président de la République ait annoncé, en mars 2018, la création d'un « *Hub* » pilotant l'enrichissement et la valorisation du SNDS en y incluant les données cliniques des hôpitaux, de la médecine de ville et des autres producteurs de soins, une feuille de route a été proposée en 2018 par une mission de préfiguration pilotée par trois experts, dont la Drees a été rapporteur⁵¹.

La mise en place de ce nouveau cadre a été prévue par la loi relative à l'organisation et à la transformation du système de santé du 24 juillet 2019, dont l'article 41 crée une Plateforme des données de santé, qui se substitue à l'Institut national des données de santé, en élargissant notablement ses missions. Cette Plateforme vise à réunir, organiser et mettre à disposition les informations issues d'un SNDS progressivement enrichi à l'ensemble des données cliniques collectées lors des actes pris en charge par l'assurance maladie (ainsi qu'à celles relatives à la perte d'autonomie et celles recueillies par les services de protection maternelle et infantile, lors des visites médicales scolaires et lors des visites de prévention des salariés), les dispositions de la loi « informatique et libertés et le référentiel de sécurité du SNDS continuant par ailleurs à s'appliquer.

À la suite du vote de la loi, la plateforme a été constituée en novembre 2019 sous forme d'un groupement d'intérêt public dénommé « Plateforme des Données de Santé » (*Health Data Hub - HDH*). Ce groupement est constitué à la fois des membres de l'INDS et de nouveaux membres pouvant y apporter une contribution en tant que producteurs ou utilisateurs de données de santé, sachant que les données recueillies et les services rendus par la Plateforme peuvent aller au-delà de ses membres.

Le HDH demeure comme l'INDS le secrétariat unique par lequel transitent les demandes d'accès des porteurs de projet aux bases de données de santé, le CEREES devenant le comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé. Le HDH continue de la même façon à assurer l'accompagnement des utilisateurs et contribue à la mise en place de procédures simplifiées en accord avec la CNIL. Les possibilités d'accès aux données du SNDS sont en outre élargies à l'ensemble des traitements d'intérêt public de données concernant la santé afin, notamment, de permettre la constitution d'entrepôts de données.

Le HDH se voit par ailleurs conférer la mission d'effectuer en propre la mise à disposition des données du SNDS élargi, ainsi que la promotion de l'innovation et la réalisation de traitements de données pour le compte de tiers. Il doit aussi contribuer à la diffusion des normes de standardisation

⁵¹ *Health Data Hub*, Rapport de la mission de préfiguration pilotée par Marc Cuggia, Dominique Polton et Gilles Wainrib, octobre 2018.

pour l'échange et l'exploitation des données de santé et peut accompagner, notamment financièrement, les projets sélectionnés dans le cadre d'appels lancés à son initiative et les producteurs de données qui s'y associent.

2.1.2 L'avancement du projet

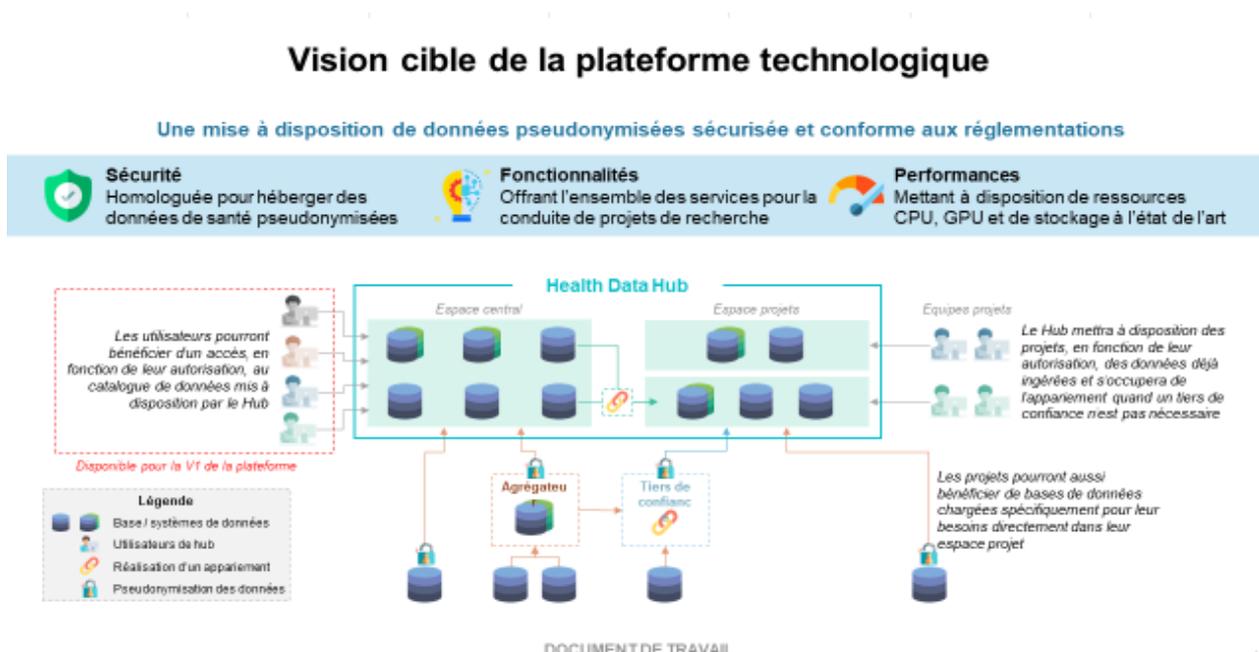
La mise en place du *Health Data Hub* a donné lieu à plusieurs chantiers, démarrés au premier trimestre 2019 dans le cadre d'une lettre de mission confiée à la Drees, puis pris en charge par la nouvelle structure constituée en fin d'année.

Celle-ci bénéficiera d'un budget de 19,6 M€ en 2020, et évalué à 76,8 M€ sur l'ensemble de la période 2019-2022. Ses effectifs prévisionnels devraient atteindre 28 personnes au début 2020, dont 15 nouveaux recrutements de profils variés, 49 à la fin de l'année et au total 75 personnes à la fin 2022.

Les principaux travaux d'ores et déjà menés concernent :

- le recensement des bases de données potentiellement concernées par une intégration dans le futur système, et l'engagement de discussions avec les différents promoteurs de cohortes ou de plateformes (dont l'Inca et l'Inserm -cf. infra) ; une étape importante évoquée depuis plusieurs années est en outre l'enrichissement du SNDS par le contenu des examens de biologie réalisés en ville, qui est en principe envisagé à l'horizon de deux ou trois ans, mais qui nécessite des investissements importants au sein des laboratoires s'agissant de l'homogénéisation du codage des données et de la métrologie fine des équipements ;
- la mise en place d'une plateforme technologique sécurisée qui a vocation à héberger de façon sécurisée des données individuelles de santé pseudonymisées, à offrir des espaces de travail, des services et un appui aux projets de recherche et, le cas échéant, à réaliser des appariements pour leur compte (cf. graphique2).

Graphique 2 : Vision cible de la plateforme du *Health data Hub*



Alors que le choix du CASD avait été parmi les solutions envisagées pour l'installation de cette plateforme, le choix finalement retenu, pour des raisons de « moindre complexité », a été celui du service d'hébergement de données *Cloud* de Microsoft, qui fait l'objet d'interrogations de la part de certains acteurs.

La plateforme homologuée devrait être en fonctionnement au début 2020, les prochaines étapes consistant à déterminer son « modèle économique », à savoir le système de tarification des accès et services proposés aux utilisateurs publics et privés, ainsi que ses relations respectives avec les projets de plateformes portés par l'Inca et l'Inserm (*cf. infra* points 2.2 et 2.3).

- le lancement de deux appels à projets permettant de sélectionner vingt projets « pilotes » visant à la fois à soutenir des usages innovants des données de santé et à participer à la construction de l'offre de services et du catalogue de données de la plateforme. Parmi ces projets, plusieurs concernent des appariements entre le SNDS et des registres ou des cohortes de patients (mais aucune dans le périmètre du PIA), comme l'évaluation des parcours de soins après un infarctus du myocarde, l'appréhension des trajectoires des patients Parkinsoniens (cohorte de 20 000 patients suivis dans les centres experts Parkinson) ou l'étude des liens entre syndrome métabolique et pathologies prostatiques (cohortes Observapur et Progene).

Plus généralement, la perspective du *Health Data Hub* soulève, à court-moyen comme à long terme, trois types d'enjeux et questionnements pour les cohortes :

- l'optimisation des lieux et des structures susceptibles d'effectuer les appariements souhaités entre leurs données et celles du SNDS, et le stockage sécurisé des bases qui en résultent ;
- l'alimentation de la Plateforme à partir de leurs propres données et les conséquences qui en découlent en termes d'ouverture et de mise à disposition de ces données aux chercheurs ;
- la mise en question à terme, au fur et à mesure de l'enrichissement du SNDS, de l'intérêt de certains recueils d'informations « actifs » (examens cliniques, prélèvements biologiques...) pour la réalisation de suivis longitudinaux, ou la possibilité de limiter fortement leur étendue.

2.2 La plateforme développée par l'Inca

Parallèlement au *Health data Hub*, l'Institut national du cancer (Inca) a entrepris la constitution d'une plateforme de données en cancérologie (PDC), décrite en annexe 5 à propos du suivi longitudinal des pathologies cancéreuses.

Partant d'une extraction du SNDS concernant les 8 millions de personnes repérées comme atteintes d'un cancer depuis 2010, la PDC vise à les appairer avec celles de 26 registres des cancers ainsi qu'avec celles des centres de coordination des dépistages.

L'objectif affiché est d'y adjoindre à terme, sous une forme dématérialisée et interopérable, et après un processus d'harmonisation des systèmes d'information professionnels, les données à visée de soins collectées au niveau local, notamment celles issues des réunions de concertation pluridisciplinaire et des programmes personnalisés de prise en charge.

Ce projet de plateforme veut comme le *Health data Hub* offrir un accès et proposer une offre de services aux chercheurs de différentes disciplines.

2.3 La plateforme du projet France Cohortes porté par l'Inserm

Dans le cadre de son projet d'UMS « France Cohortes », l'Inserm entend développer une plateforme informatique commune aux cohortes auxquelles l'organisme participe dans le cadre de son *data center* hébergé au sein du Centre Informatique national de l'enseignement supérieur de Montpellier (CINES), établissement public administratif créé en 1999 et qui héberge déjà les données de l'Inserm. Le projet, qui s'inspire largement du modèle du système d'information de la plateforme RADICO, a pour ambition de fournir un système d'information mutualisé qui :

- soit conforme au référentiel de sécurité du SNDS, aux normes exigées des hébergeurs de données de santé (HDS) et au règlement général pour la protection des données (RGPD),
- décharge les chercheurs de la gestion informatique des cohortes (gestion des matériels, des logiciels, des réseaux, des installations, de la maintenance),
- soit ouvert et interopérable avec d'autres systèmes d'information,
- soit doté de bulles sécurisées permettant aux chercheurs de traiter les données en toute sécurité.

La solution-cible (cf. graphique 3) envisage ainsi une architecture technico-fonctionnelle reposant sur huit briques⁵². Les quatre premières portent sur les questions de stockage de données :

« 1. Brique applicative où sont stockées les applications de collecte, d'alimentation, d'extraction et de consultation des données des cohortes (hors données SNDS).

2. Brique « Bases de données de production » où sont enregistrées les données des cohortes et leurs données SNDS associées.

3. Brique « Data Warehouse » (entrepôt de données) où sont historisées les données pour exploitation et aide à la prise de décision.

4. Brique « Exploitation » qui regroupe tous les outils d'exploitation, monitoring, maintenance, surveillance du système, consommation de ressources, traçage des opérations réalisées, sauvegardes, restauration, etc. »

Les quatre briques suivantes portent sur l'appariement et l'exploitation des données des cohortes dans un cadre sécurisé :

« 5. Bulles sécurisées Inserm qui fournissent des services sécurisés d'appariement et d'exploration statistique pour des besoins internes de traitement des données SNDS.

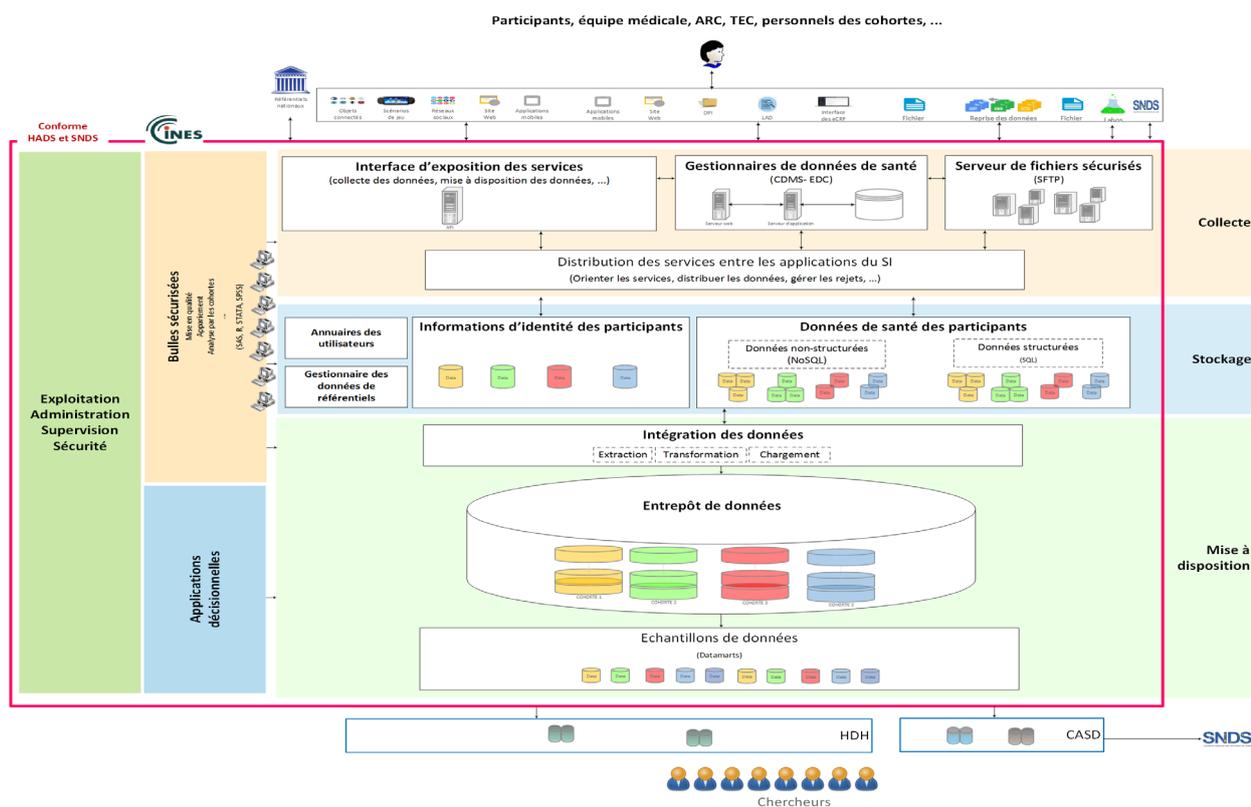
6. Bulles sécurisées du CASD qui fournissent des services sécurisés d'appariement et de calculs statistiques pour les chercheurs qui souhaitent travailler sur un échantillon de données des cohortes et du SNDS.

7. Brique « Business Intelligence (BI) » qui est un ensemble d'outils et de méthodes visant à analyser les données pour l'aide à la prise de décision.

8. Bureaux virtuels pour accéder aux bulles sécurisées dans le respect des exigences de sécurité des postes de travail qui accèdent à des données sensibles. »

⁵² Note de présentation du projet France Cohortes, 21 mai 2019.

Graphique 3 : Vision cible de la plateforme de France Cohortes



Source : Inserm

Le projet est prévu pour être opérationnel dix-huit mois après son démarrage, soit au plus tôt à la fin de l'année 2020. Son coût prévisionnel de mise en place est de 1,95 M€, pour des coûts de maintenance estimés à 200 K€/an.

Ce projet qui se situe à l'interface d'autres infrastructures existantes ou en cours de création, elles-mêmes d'ambition et de caractéristiques voisines, soulève à court-moyen terme, compte tenu de ce positionnement, trois types de questionnements et de risques liés :

- aux conditions pratiques de transition entre les différents systèmes d'information déjà développés et utilisés par les cohortes auxquelles participe l'Inserm, qui présentent eux-mêmes des caractéristiques fonctionnelles et techniques diverses et des niveaux de maturité variables, et le nouveau système unique mutualisé qui doit leur succéder ;
- aux modalités d'articulation du nouveau système d'information avec les fonctionnalités de l'infrastructure informatique du CASD ;
- les modalités d'articulation technique et fonctionnelle à définir du nouveau système d'information avec la plateforme technologique en cours de développement dans le cadre du projet de Health data Hub.

ANNEXE 7 : LES GRANDES COHORTES ÉPIDÉMIOLOGIQUES EN POPULATION GÉNÉRALE : HISTORIQUE ET PANAROMA INTERNATIONAL

Dans le cadre de l'évaluation de la politique des cohortes en France, lancée sous l'égide du PIA, il était intéressant et logique de procéder à une comparaison avec les pratiques retenues par les principaux pays industrialisés ; cette analyse internationale met en particulier en lumière la spécificité et la place centrale qu'occupe, à cet égard, le monde anglo-saxon.

1 Une prédominance anglo-saxonne ancienne et solidement établie

Issu, dans son acception épidémiologique, du vocabulaire anglo-saxon, le terme « *cohort* » a été utilisé pour la première fois en 1939, dans une importante étude publiée dans l'*American Journal of Hygiene* et consacrée à l'épidémiologie de la tuberculose. Il sera largement repris ultérieurement, à compter des années 1950, à mesure que se sont développées de façon rapide et importante les études de ce type, au sortir de la Seconde Guerre mondiale.

1.1 Les cohortes épidémiologiques fondatrices

Pionnières, ces grandes cohortes fondatrices vont marquer les esprits par leurs résultats scientifiques retentissants, et vont contribuer, de manière décisive, à l'essor de l'épidémiologie et à l'acquisition par cette discipline d'un véritable statut scientifique.

Or, ces premières cohortes sont toutes nées, simultanément, au Royaume-Uni et aux États-Unis d'Amérique, témoignant de l'enracinement dans le monde anglo-saxon d'une puissante culture de santé publique et ouvrant la voie aux autres pays développés.

Passées à la postérité depuis, ces premières cohortes ont connu un retentissement considérable. Elles ont permis de dépasser les seules enquêtes « cas/témoins », permis d'acquérir des résultats substantiels et, ainsi, ont renouvelé la santé publique tout entière. La plus célèbre d'entre elles, la *Framingham Heart Study*, est devenue emblématique dans l'univers de l'épidémiologie.

- La Framingham Heart Study (États-Unis) : mise en place en 1949, son impact fut considérable ; elle reste encore aujourd'hui une référence et la cohorte la plus célèbre. Composée initialement de 5.000 personnes -à parité hommes/femmes-, elle fut complétée en 1971 par 5.000 participants supplémentaires -les enfants de la première cohorte-, et encore en 2000 par près de 4.000 petits-enfants des premiers inclus. Elle a permis, grâce aux 2.000 articles scientifiques auxquels elle a donné naissance à ce jour, de comprendre de façon approfondie la maladie cardio-vasculaire.
- La 1946 Birth Cohort (Royaume-Uni) : destinée à élucider les raisons de la baisse, inquiétante, des taux de fécondité britanniques, il s'agissait, à travers elle, d'évaluer l'efficacité des maternités. Pour cela, on y inclut toutes les femmes ayant accouché au cours de la même semaine de mars 1946, soit 13.687 femmes. Après que le baby-boom eut rendu son premier

objectif caduc, elle fut maintenue afin d'étudier la portée des conditions familiales et sociales sur le devenir des enfants. À partir d'elle, a été tiré une seconde cohorte de 5.000 enfants nés en mars 1946, suivis tout au long de leurs vies, tant sur le plan médical que psychologique, mais aussi scolaire. Première des cohortes de naissance, elle a ouvert la voie à ce type d'étude dans le monde entier. Au Royaume-Uni, elle fut prolongée par les cohortes de naissance de 1958, 1970 et 2000.

- La *British Doctors' Study* (Royaume-Uni) : eut, elle aussi, un retentissement considérable. Elle permit, en effet, de confirmer le rôle de facteur de risque du tabagisme dans l'apparition du cancer broncho-pulmonaire. Rôle établi par les auteurs dans une enquête rétrospective « cas/témoins », mais remis en question en raison de l'instrument épidémiologique utilisé. Aussi, en 1951, les auteurs de cette première enquête ont-ils lancé la cohorte en y incluant plus de 40.000 médecins (34.440 hommes et 6.194 femmes). Dès 1954, les premiers résultats obtenus, corroborés par ceux qui suivirent, établirent sans contestation possible le rôle cancérigène du tabac.
- L'*American Cancer Society Study* (États-Unis) : très proche de la précédente, et lancée au même moment, comportant 190.000 individus mâles, âgés de 50 à 69 ans et membres de cette association, elle en confirma les résultats.

1.2 Le développement des cohortes dans les années 1970

Avec la reconnaissance du caractère multifactoriel de l'étiologie des maladies chroniques, à partir des années 1970, vint le temps du développement des grandes cohortes.

Cet accroissement quantitatif s'est doublé d'un perfectionnement quantitatif de l'instrument cohorte : établissement de modèles causaux des maladies et raffinement progressif des méthodes statistiques utilisées pour traiter les données longitudinales. Est venue s'ajouter à ce perfectionnement la possibilité d'apparier les données de santé des sujets suivis avec les données des registres nationaux de maladie et de mortalité.

Parmi les très nombreuses cohortes ainsi mises en place au fil des décennies, on peut relever quelques exemples particulièrement significatifs.

- Les cohortes *Whitehall I et II* (Royaume-Uni) : mise en place en 1967-1969, la première Cohorte *Whitehall* (du nom de la rue célèbre de Londres où étaient situés les bureaux des personnes recrutées) était constituée de 19.018 fonctionnaires masculins. Son objectif était d'évaluer le risque de maladie cardio-vasculaire, sur le modèle de la *Framingham Study* américaine. Elle permit de mettre en évidence un gradient social de santé au regard de la maladie cardio-vasculaire. En 1985, *Whitehall II* vit le jour, avec un recrutement similaire, afin d'approfondir ces premiers résultats. Elle représente une contribution majeure à la recherche en épidémiologie sociale et est considérée, d'un point de vue international, comme l'une des meilleures sources de connaissance scientifique sur la question, importante, des déterminants sociaux de santé.
- La *Nurses' Health Study* (États-Unis) : cette étude, lancée en 1976, assure le suivi prospectif de 122.000 infirmières âgées de 30 à 55 ans, afin d'évaluer, à long terme, les effets de la contraception orale. Deux vagues de prélèvements sanguins (1989 et 2000) ont été collectés et conservés en vue d'examen ultérieurs. En 1989, une seconde vague de cette étude a inclus

116.000 femmes de 25 à 42 ans. Les deux phases de cette même étude explorent les facteurs de risque pour les principales maladies chroniques de la femme.

- La Physicians' Health Studies (États-Unis) : débuté en 1982, cet essai randomisé, visant à évaluer les efficacités respectives de l'aspirine et du bêta-carotène dans la prévention de la maladie coronarienne et du cancer, s'est transformé en une des plus grandes cohortes observationnelles américaines. Elle comportait 33.223 médecins hommes, âgés de 40 à 84 ans lors de l'inclusion. Elle a permis d'établir l'action préventive de l'aspirine à dose filée au regard de l'infarctus du myocarde.
- L'European Prospective Investigation Into Cancer and Nutrition (EPIC, Europe) : lancée dans les années 1990, sur la base de l'inclusion de 520.000 personnes dans dix pays d'Europe occidentale (Danemark, France, Allemagne, Royaume-Uni, Suède, Grèce, Italie, Pays-Bas, Norvège, Espagne), elle vise à étudier la relation entre alimentation, mode de vie, facteurs environnementaux et incidence des cancers. Des échantillons sanguins ont été prélevés chez 387.256 personnes suivies et stockés pour des analyses ultérieures.

2 Une diffusion internationale désormais considérable d'un instrument à la puissance renouvelée et aux objectifs évolutifs

Considérées désormais comme un outil épidémiologique fondamental, les cohortes en population générale ont connu, depuis leur apparition au lendemain de la Seconde Guerre mondiale, un essor tel qu'il est aujourd'hui impossible de les dénombrer avec précision. Elles sont devenues un instrument de recherche sanitaire et sociale très répandu de par le monde.

À ce développement rapide et marqué, est venue s'ajouter la prise en compte de changements scientifiques importants survenus dans les dernières décennies, qui ont contribué au développement de ce qu'il est convenu de désigner par le terme « méga-cohortes ».

2.1 L'émergence internationale des « méga-cohortes »

La nature des questions scientifiques ou médicales auxquelles les cohortes doivent apporter des réponses solides a beaucoup évolué. L'environnement scientifique a lui-même contribué, par ses changements, à cette redéfinition du rôle et du mode d'emploi des cohortes.

Cinq évolutions importantes ont ainsi conduit à l'apparition et à l'utilisation des « méga-cohortes » :

- le rôle des facteurs génétiques : l'identification progressive de diverses variantes génétiques associées aux maladies, ainsi que l'interaction potentielle de cette dimension génétique avec les facteurs environnementaux, rend nécessaire des cohortes de grande taille afin de réaliser des analyses suffisamment puissantes ;
- le vieillissement de la population dans les pays occidentaux et, phénomène lié, le développement des maladies chroniques et des co-morbidités, qui impliquent l'analyse combinée de multiples déterminants et de leurs interactions ;
- la prise de conscience récente de l'importance des facteurs environnementaux qui nécessite des études à un niveau géographique fin dans des localisations plurielles ;

- l'évolution rapide des connaissances et des techniques de biologie moléculaire (génomique, protéomique), qui nécessite la mise en place de larges biobanques ;
- le souci de détecter des risques de très faible ampleur : ce besoin est particulièrement aigu dans les domaines de la recherche sur l'origine et les causes des maladies de nature sociale, environnementale, professionnelle, nutritionnelle, génétique ou biologique. Il en va de même pour la pharmaco-épidémiologie. Grâce aux cohortes de très grande taille, il devient alors envisageable dépister des « signaux faibles » voire « très faibles » : effets indésirables des médicaments, effets des faibles doses de rayonnements ionisants ou des ondes électromagnétiques, effets potentiellement cancérigènes des téléphones portables...

Ce contexte scientifique nouveau rend nécessaire l'usage de cohortes de très grande ampleur, accompagné d'un suivi à long terme des personnes incluses et d'un phénotypage des maladies de haute qualité, afin d'assurer une puissance statistique suffisante, permettant de saisir les rôles respectifs des facteurs comportementaux et environnementaux, ainsi que leur interaction avec les caractères génétiques.

Ainsi, se sont mises en place, dans les années 2000, une nouvelle génération de cohortes en population, de très grande taille, les « méga-cohortes ». Leur nombre est estimé au début des années 2010 à plus de 100 de par le monde⁵³.

Quelle que soit leur diversité, ces grandes cohortes ont souvent des caractéristiques communes :

- les cohortes d'adultes s'intéressent en particulier aux grandes pathologies chroniques et dégénératives : maladies cardio-vasculaires, diabète et maladies métaboliques, cancers, démences... ;
- les cohortes de naissance sont centrées en particulier sur les facteurs de risque liés aux différents stades du développement de l'enfant ;
- l'étude de l'influence des facteurs génétiques est très présente, voire essentielle pour certaines cohortes ;
- le développement de biomarqueurs de détection précoce des pathologies est privilégié ;
- toutefois, nombre de facteurs de risque sont pris en compte : personnels et familiaux, comportementaux, sociaux, professionnels, environnementaux... ;
- ces cohortes incluent des recueils de données multiples mettant en œuvre des techniques diverses : questionnaires, entretiens, examen médical, appariement à des bases de données nationales, investigations complémentaires, radiologiques notamment ;
- la mise en place de biobanques y a été associée, destinées à stocker des échantillons biologiques divers (sérum, cellules, ADN).

La plupart de ces « méga-cohortes » sont gérées comme des infrastructures de recherche, largement ouvertes sur la ou les communautés scientifiques intéressées, notamment sous la forme d'appels publics à projets, offrant la possibilité à des chercheurs extérieurs de bénéficier d'un accès aux données collectées.

⁵³ Marie Zins et Marcel Goldberg, « Les nouvelles méga-cohortes en population en Europe », *ADSP*, n°78, mars 2012.

En dépit de leur grande taille, ces « méga-cohortes » ne sont pas toujours à même de fournir la puissance statistique nécessaire à l'élucidation des relations étiologiques complexes des maladies.

Il faut alors mettre en commun les données obtenues par plusieurs grandes cohortes en constituant un consortium de recherche ; ce regroupement de forces scientifiques est ainsi devenu la règle pour les études de génomique et tend à se généraliser pour l'ensemble des études épidémiologiques. Il existe, en particulier, un consortium international de ces grandes cohortes, le *100K Cohort Consortium* (IHCC), dont la cohorte française Constances fait partie.

C'est à l'occasion de l'émergence de ces « méga-cohortes » que l'Europe continentale a comblé une partie de son retard au regard du monde anglo-saxon.

2.2 Le rattrapage européen

Dans ce mouvement général de rattrapage européen, le Royaume-Uni a néanmoins conservé sa place de locomotive. Il offre, en effet, comparativement à nombre de pays européens un environnement scientifique et réglementaire propice à l'établissement et au développement de grandes cohortes de population. Ceci est particulièrement vrai au regard des questions de protection des données personnelles et d'appariement. Ainsi, par exemple, si les états nordiques possèdent de nombreuses cohortes appariées, celles prévoyant une participation directe des personnes incluses et le recueil d'échantillons biologiques sont rares : les « cohortes » scandinaves ont plutôt tendance à être, comme en Suède, des dispositifs de suivi longitudinaux fondés sur des données médicales et administratives que des cohortes cliniques au sens classique du terme.

A contrario, l'absence, aux États-Unis, d'un système national de santé unique, limite la collecte de données de santé par les études longitudinales.

Sans chercher à les énumérer toutes ici, voici les principales grandes cohortes de population européennes :

- La Million Women Study (Royaume-Uni) : entre 1996 et 2001, plus de 1,3 million de femmes ont été incluses dans cette étude. Étude ayant pour objectif l'exploration de l'influence des facteurs comportementaux et liés à la reproduction sur l'état de santé. Un de ses principaux résultats a été la mise en évidence du rôle de l'hormonothérapie comme traitement substitutif de la ménopause dans l'apparition du cancer du sein. Cette découverte a connu un grand retentissement et aurait permis, par les modifications thérapeutiques auxquelles elle a donné le jour, la prévention de la survenue d'environ 10.000 cancers du sein ;
- UK Biobank (Royaume-Uni) : cette étude a intégré, entre 2006 et 2010, 500.000 personnes âgées de 40 à 69 ans. Et cela afin d'évaluer l'interaction entre génétique, mode de vie et facteurs environnementaux dans l'apparition d'un large panel de maladies. Grâce à la collecte d'information et d'échantillons biologiques sur 500.000 personnes âgées de 40 à 69 ans, *UK Biobank* est devenue une ressource essentielle des politiques de prévention, de diagnostic et de traitement au Royaume-Uni. Les participants inclus seront suivis pendant les 30 prochaines années, notamment grâce à l'appariement avec les données issues du NHS et des registres nationaux de décès et de cancers. En 2012, *UK Biobank* a été ouverte (*open data*) aux chercheurs britanniques et étrangers ;

- CONOR, Cohort of Norway et MoBa (Norwegian Mother and Child Cohort Study, Norvège) : comportant respectivement 200.000 adultes et 270.000 mères, pères et leurs enfants, résultats considérables pour un pays peuplé seulement de 4,5 millions d'habitants. La méthode retenue repose sur des questionnaires auto-remplis et des examens biologiques dès la 17^{ème} semaine de grossesse, puis à la 30^{ème} semaine, aux 6^{ème} et 18^{ème} mois de l'enfant, et enfin à l'âge de 7 ans. Des prélèvements sanguins (sang du cordon, sang des parents aux 18 mois de l'enfant) et d'urine sont pratiqués. Elle vise à évaluer les rôles de l'environnement et des expositions à des facteurs de risque sur la santé des mères durant leur grossesse et sur celle de leurs enfants ; mais aussi les associations entre certaines expositions et le développement de maladies spécifiques. Les thèmes explorés sont la nutrition, les infections, les facteurs héréditaires et environnementaux, les traitements médicaux, les expositions aux risques professionnels.
- LifeGene (Suède) : comportant 500.000 sujets âgés de 0 à 45 ans, grâce d'abord à un échantillon aléatoire de personnes âgées de 18 à 45 ans, avec inclusion ultérieure de leurs enfants ;
- LifeLines (Pays-Bas) : sur le même modèle constitutif que la précédente, avec 165.000 personnes incluses ;
- GeNatCo, German National Cohort, (Allemagne) : 200.000 participants âgés de 20 à 70 ans.

ANNEXE 8 : L'EXPÉRIENCE DES GRANDES COHORTES ÉPIDÉMIOLOGIQUES AU ROYAUME-UNI

Le Royaume-Uni possède un très large éventail de grandes cohortes de population, pour ne pas dire unique. Ce dispositif scientifique, procurant une grande richesse d'informations phénotypiques, biologiques, mais aussi sociales, lui permet de conduire au mieux les études longitudinales, en termes de santé publique ou d'impact socio-économique tout au long de la vie, qu'il estime nécessaires.

Il est à noter que les cohortes mentionnées ici sont des cohortes en population, centrées sur les déterminants de l'état de santé et les facteurs de risques. Ces cohortes donnent lieu à une politique spécifique, qui, contrairement à ce qui a été pratiqué en France, n'inclut pas dans la même problématique les cohortes particulières de patients concernés par telle ou telle pathologie ou suivant tel ou tel type de traitement.

1 Une large et riche palette de « grandes cohortes » épidémiologiques

Depuis plus de 50 ans, le *Medical Research Council (MRC)* a financé une large palette de cohortes de population, diverses au regard des critères de taille, de localisation ou de durée de vie, comme des caractéristiques retenues pour l'inclusion des personnes (âge, sexe, origines ethniques, niveau socio-économique).

1.1 Éléments de présentation des cohortes épidémiologiques les plus importantes

Seront présentées ici, sans souci d'exhaustivité, les 34 plus importantes cohortes épidémiologiques britanniques ; 19 d'entre elles sont financées, en tout ou en partie, par le MRC, les 15 autres par l'initiative privée et les organisations philanthropiques.

Présentation des 34 cohortes les plus importantes

Intitulé de la cohorte	Organisme responsable	Principaux financeurs
11-16 and 16+ Study	MRC Social and Public Health Sciences Unit	Medical Research Council
1970 British Cohort Study (1970BCS)	Centre for Longitudinal Studies, Institute of Education	Economic and Social Research Council
Aberdeen Children of the 1950s (ACONF)	University of Aberdeen	Medical Research Council

Avon Longitudinal Study of Parents and Children (ALSPAC)	University of Bristol	Medical Research Council Wellcome Trust
Born in Bradford (BiB)	Bradford Royal Infirmary	
Boyd Orr Cohort	University of Bristol	British Heart Foundation
Breakthrough Generations Study	Breakthrough Breast Cancer	Breakthrough Breast Cancer
British Regional Heart Study (BRHS)	University College London	British Heart Foundation
British Women's Heart & Health Study (BWHHS)	London School of Hygiene and Tropical Medicine	Department of Health British Heart Foundation
Cognitive Function and Ageing Studies I (CFAS I)	University of Sheffield CFAS	Medical Research Council
Cognitive Function and Ageing Studies II (CFAS II)	University of Cambridge	Medical Research Council
Determinants of Adolescent Social well-being and Health (DASH)	MRC Social and Public Health Sciences Unit	Medical Research Council

English Longitudinal Study of Ageing (ELSA)	University College London, Institute for Fiscal Studies, University of Manchester, NatCen Social Research	UK Government Consortium (Department for Communities and Local Government, Department for Transport, Department for Work and Pensions, HM Revenue and Customs, Office for National Statistics, Department of Health National Institute Health Research, Office for National Statistics) National Institute on Aging (NIH)
---	---	--

European Prospective Investigation of Cancer Norfolk (EPIC Norfolk)	University of Cambridge,	Medical Research Council Cancer Research UK
European Prospective Investigation of Cancer Oxford (EPIC Oxford)	University of Oxford	Cancer Research UK
Gemini: health and development in twins	Department of Epidemiology and Public Health, University College London	Cancer Research UK
Growing up in Scotland (GUS)	ScotGen Social Research MRC Social and Public Health Sciences Unit Centre for Research on Families and Relationships	Scottish Government
Hertfordshire cohort study (HCS)	MRC Lifecourse Epidemiology Unit,	Medical Research Council
Lothian Birth Cohort 1936 (LBC1936)	University of Edinburgh;	Age UK Medical Research Council
Millennium Cohort (MCS)	Centre for Longitudinal Studies, Institute of Education	Economic and Social Research Council
Million Women Study	University of Oxford	Cancer Research UK Medical Research Council
MRC National Survey of Health and Development Cohort /1946 Birth Cohort (NSHD/1946BC)	MRC Unit for Lifelong Health and Ageing	Medical Research Council
National Child Development Study/1958 Birth Cohort (NCDS/1958BC)	Centre for Longitudinal Studies, Institute of Education	Economic and Social Research Council Medical Research Council Wellcome Trust
Newcastle 85+	Newcastle University	Medical Research Council Biotechnology and Biological Sciences Research Council Newcastle University

Rural Uganda General Population Cohort (GPC)	MRC/UVRI Uganda Research Unit on AIDS	Medical Research Council
Southall and Brent Revisited (SABRE)		Wellcome Trust /British Heart Foundation
Southampton Women's Survey (SWS)	MRC Lifecourse Epidemiology Unit	Medical Research Council / British Heart Foundation European Union
Twin Early Development Study (TEDS)	King's College London	Medical Research Council
Twins UK	King's College London	Wellcome Trust
UK Biobank	University of Oxford	Medical Research Council Wellcome Trust Chief Scientist Office Department of Health National Institute for Health Research British Heart Foundation Northwest Regional Development Agency Welsh Government
UK Women's Cohort Study (UKWCS)	University of Leeds	
Understanding Society	Institute for Social and Economic Research, University of Essex	Economic and Social Research Council
West of Scotland Twenty-07 Study (Twenty-07)	MRC Social and Public Health Sciences Unit	Medical Research Council
Whitehall II Study	University College London	Medical Research Council National Institute on Aging (NIH)

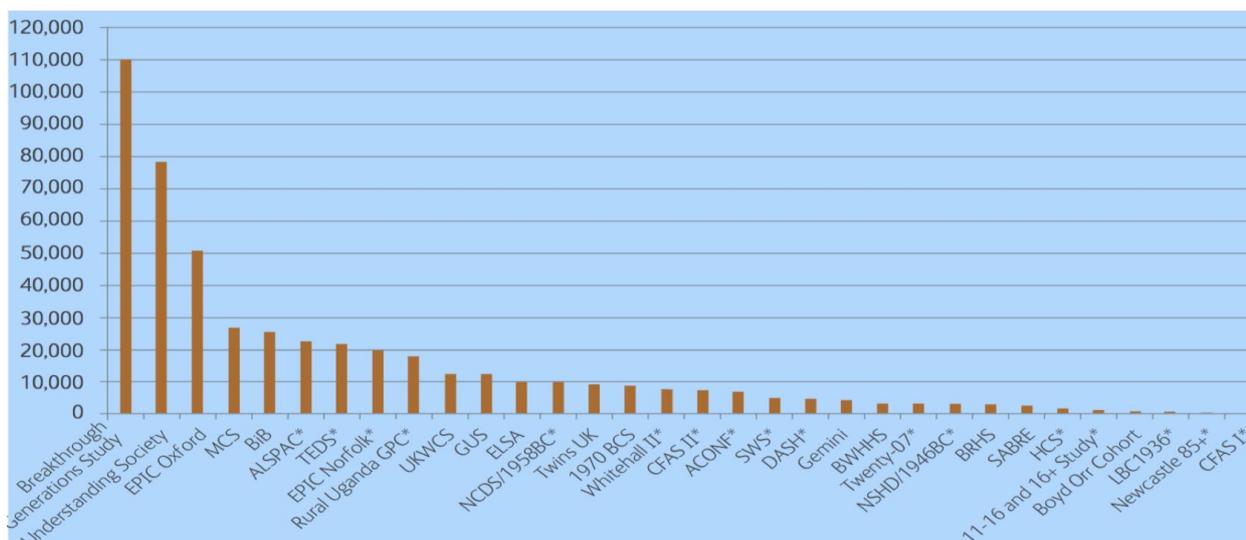
Source : *Maximising the value of UK population cohorts, MRC Strategic Review of the largest UK Population cohort studies*

Classification thématique des 34 principales cohortes

Cohortes de naissance :	NSHD/1946BC, NCDS/1958BC, 1970BCS, ALSPAC, SWS, MCS
Cohortes géographiques	11-16 and 16+ study, the Twenty-07 Study, ACONF, GUS, Rural Uganda GPC
Cohorte professionnelle (Fonctionnaires)	Whitehall II
Cohorte socio-économique	Understanding society
Gémellité	TEDS, Twins UK et Gemini
Minorités ethniques	DASH, SABRE et BiB
Cohortes d'âge	CFAS I and II, ELSA, UK Biobank, Boyd Orr, LBC1936, HCS, Newcastle 85+
Nutrition /questions sanitaires spécifiques	EPIC Oxford, EPIC Norfolk, Million Women Study, Breakthrough Generations Study, BHRS, BWHHS, UKWCS

Source : *Maximising the value of UK population cohorts, MRC Strategic Review of the largest UK Population cohort studies*

Graphique 1 : Distribution des cohortes britanniques par effectifs actuels, à l'exclusion de *Million Women Study* et de *UK Biobank*

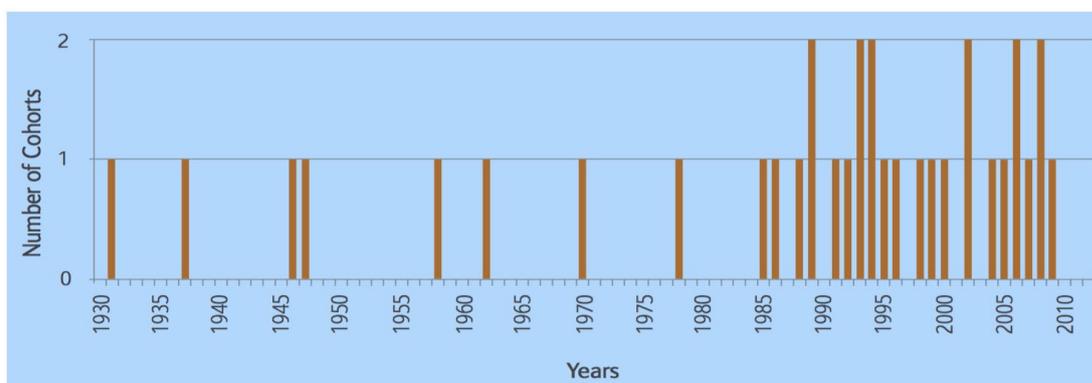


Source : *Maximising the value of UK population cohorts, MRC Strategic Review of the largest UK Population cohort studies*

Ces cohortes recrutent leurs participants en Angleterre, au Pays de Galles et en Écosse ; à noter que les cohortes *Millennium Cohort Study* et *Twins UK* élargissent leurs recrutements respectifs à l'Irlande du Nord. Une cohorte, financée par le MRC, recrute en dehors des frontières du Royaume-Uni : il s'agit de *Rural Uganda GPC*, incluant 18,000 personnes vivant dans le Sud-Ouest de l'Ouganda ; elle étudie la prévalence et l'incidence de l'infection par le VIH.

3 de ces 34 cohortes ne concernent que les femmes : *Million Women Study*, *Breakthrough Generations Study* and *UKWCS* ; elles s'attachent aux enjeux de santé liés à la pré-ménopause et à la ménopause elle-même. Une seule, *BRHS*, est exclusivement masculine.

Graphique 2 : Distribution 34 des cohortes britanniques en fonction de leurs dates respectives de lancement :



Source : *Maximising the value of UK population cohorts, MRC Strategic Review of the largest UK Population cohort studies*

1.2 Un focus sur trois cohortes particulièrement intéressantes

1.2.1 *The Million Women Study*

The Million Women Study est une très importante cohorte ; 1.320.000 femmes ont été incluses entre 1996 et 2001, soit une femme sur 4 de celles nées entre 1936 et 1950, avec un âge moyen lors de l'inclusion de 56 ans. Leur recrutement s'est fait par l'intermédiaire de 66 centres de dépistage du cancer du sein du NHS, en Angleterre et en Écosse, recouvrant la moitié de la population britannique.

L'objectif de cette ambitieuse opération était l'exploration de l'influence des facteurs comportementaux et thérapeutiques liés à la reproduction sur l'état de santé. Un de ses principaux résultats a été la mise en évidence du rôle de l'hormonothérapie comme traitement substitutif de la ménopause dans l'apparition du cancer du sein. Cette découverte, publiée dans la célèbre revue *Lancet* (9 août 2003), a connu un grand retentissement et aurait permis, par les modifications thérapeutiques auxquelles elle a donné le jour, la prévention de la survenue d'environ 10.000 cancers du sein. Des résultats comparables ont été également obtenus pour le cancer endométrial et celui de l'ovaire, établissant là aussi le facteur de risque représenté par le traitement substitutif de la ménopause.

1.2.2 *UK Biobank*

Contrairement à ce que son nom seul pourrait laisser penser, *UK Biobank* est une cohorte, de très grande taille, lancée en 2006 afin d'évaluer l'interaction entre génétique, mode de vie et facteurs environnementaux dans l'apparition d'un large panel de maladies : cancers, maladies cardiovasculaires, AVC, diabète, ostéoporose, arthrose, dépression, démence, troubles ophtalmologiques, entre autres.

Grâce à la collecte d'informations et au prélèvement de 36 types d'échantillons biologiques (sang et urine) sur les 500.000 personnes âgées de 40 à 69 ans, incluses entre 2006 et 2010, par 10 centres répartis sur l'ensemble du territoire britannique, *UK Biobank* est devenue une ressource essentielle des politiques de prévention, de diagnostic et de traitement au Royaume-Uni.

Les participants inclus seront suivis pendant les 30 prochaines années, notamment grâce à l'appariement avec les données issues du NHS et des registres nationaux de décès et de cancers. Un *Big Data Institute (University of Oxford)* est couplé à *UK Biobank*, pour le traitement (*data mining*) des informations collectées. En 2012, *UK Biobank* a été ouverte (*open data*) aux chercheurs britanniques et étrangers.

Son financement initial s'est élevé à quelque 65 millions de livres. Outre une large gamme de prélèvements biologiques, l'ensemble des personnes incluses ont toutes eu un génotypage réalisé. Par ailleurs, un programme d'imagerie (scanner corps entier) a été lancé sur 100 000 volontaires.

1.2.3 *The Determinants of Adolescent Social well-being and Health (DASH)*

The Determinants of Adolescent Social well-being and Health (DASH) est une étude concernant environ 5.000 jeunes gens, ayant pour objectif l'exploration de l'influence, à long terme, des conditions sociales sur l'état de santé et le bien-être des adolescents issus des minorités ethniques, et cela dans la zone de Londres.

Les résultats de cette étude ont notamment montré un meilleur état psychologique de ces adolescents par rapport à leurs camarades blancs. Les résultats sanitaires de DASH ont été largement utilisés par les établissements scolaires et les communautés locales.

2 Les cohortes britanniques : des caractéristiques liées à un contexte favorable, mais des enjeux et des stratégies en débat

2.1 Des caractéristiques et un contexte favorable

2.1.1 La coopération entre trois grands financeurs

Le financement de la recherche épidémiologique au Royaume-Uni est caractérisé par la coexistence et l'action complémentaire du *MRC* et de deux autres institutions. À cet égard, un fait remarquable doit être souligné : en accord avec son histoire et ses traditions culturelles, le Royaume-Uni voit l'initiative privée occuper une place de choix dans le paysage des cohortes épidémiologique. Paysage dans le lequel le *Wellcome Trust* joue un rôle central.

Fondée en 1936 et désormais dotée d'un fonds d'environ 15 milliards de livres, cette organisation caritative dédiée à la promotion de la recherche biomédicale, héritière des laboratoires pharmaceutiques éponymes, occupe, dans ce secteur, la seconde place mondiale en termes de puissance financière, derrière la Fondation Gates. Chaque année, le *Wellcome Trust* investit quelque 700 millions de livres, principalement au Royaume-Uni (même si des projets internationaux importants existent), ce qui en fait un acteur majeur de la recherche médicale dans ce pays.

Autre caractéristique du système anglais : le rôle important joué par l'*Economic and Social Research Council (ESRC)*, partenaire majeur du financement de la recherche et qui participe, en partenariat avec le *MRC* et le *Wellcome Trust*, à celui des grandes cohortes nationales. L'*ESRC* est une des 7 agences du *UK Research and Innovation (UKRI)*, organisme public central chargé de la stratégie britannique en matière de recherche. Son budget annuel avoisine les 202 millions de livres, dont 10% sont affectés aux cohortes et autres études longitudinales. Ses cohortes de naissance (*NCDS*, *BCS 70*, *MCS* et *LSYYPE*) sont hébergées dans le *Centre for Longitudinal Studies (CLS)* au sein de *UCL Institute of Education*. La longue tradition britannique de recherche en sciences sociales dans laquelle s'inscrit l'action de l'*ESRC* a fait naître, notamment grâce aux études longitudinales, le concept de "*nurture*", contraction des termes nature et culture, témoignant de l'intrication entre données biologiques et sociales.

Ces trois opérateurs majeurs du financement de la recherche épidémiologique au Royaume-Uni ont récemment introduit des critères objectifs pour guider leurs choix de soutien et leurs investissements, à l'image du *Wellcome Trust* et de sa *Longitudinal Population Studies strategy*.

2.1.2 Un environnement propice permettant de tirer parti de bases médico-administratives très riche

Comparativement à nombre d'autres pays européens, le Royaume-Uni offre un environnement scientifique et réglementaire propice à l'établissement et au développement de grandes cohortes de population. Ceci est particulièrement vrai au regard des questions de protection des données personnelles et d'appariement. Ainsi, la possibilité de chaîner les données de santé, de collecter et conserver les prélèvements biologiques, de même que certains examens radiologiques et des

informations génétiques, place ce pays dans une position excellente au regard des enjeux sanitaires et sociaux liés au développement de l'instrument « cohorte ». L'absence, aux États-Unis, d'un système national de santé unique, y limite la collecte de données de santé par les études longitudinales.

Indépendamment de ces grandes cohortes, le NHS est train de développer sa propre stratégie en matière d'utilisation massive des données de santé et d'ouverture des données aux chercheurs dans le cadre de programmes et des structures qui lui sont propres (*NHS Digital*).

Ainsi toutes les cohortes sur les 34 présentées précédemment, sauf une (*Gemini*), ont recours au chaînage des données.

Le prélèvement d'échantillons biologiques est réalisé par 31 des 34 cohortes. Il s'agit de sang, d'urine, de cellules buccales, de placenta, mais aussi de prélèvements *post-mortem* (cerveau, dents, muscles).

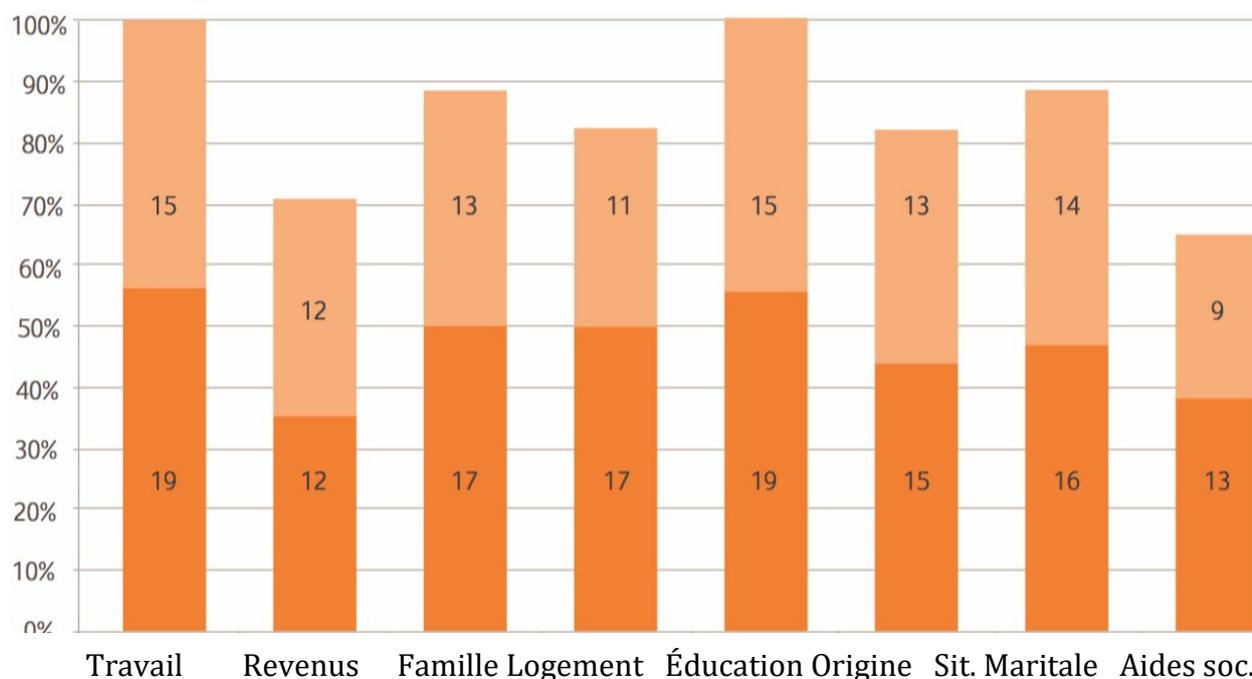
De même, 26 de ces 34 cohortes ont conduit des études génétiques ou métaboliques (*omics studies*), le plus souvent un génotypage (23 de ces 26 cohortes). Rappelons qu'avec *UK Biobank*, le Royaume-Uni dispose déjà d'une collection de 500.000 génotypes.

2.1.3 Un accent fort mis sur les liens entre caractéristiques socio-économiques et données de santé

Conformément à la tradition culturelle, politique et scientifique du Royaume-Uni, qui met au premier plan les recherches sur les déterminants de l'état de santé et le rôle joué par les inégalités sociales en la matière, le recueil des données de nature socio-économique est très répandu, comme en témoigne le tableau suivant.

Ces informations sont obtenues, le plus souvent, *via* des auto-questionnaires, mais des appariements avec les données de l'aide sociale sont aussi largement pratiqués.

Graphique 3 : Nombre de cohortes recueillant des données socio-économiques, sur les 34 retenues



Source : *Maximising the value of UK population cohorts, MRC Strategic Review of the largest UK Population cohort studies*

2.2 Des enjeux et des stratégies en débat

Ce contexte favorable n'empêche pas qu'aient émergé des interrogations sur la multiplicité et la profusion des cohortes existantes, qui ont des tailles, des populations d'intérêt et des variables définies de façons différentes, et peuvent coexister, sans forcément beaucoup d'échanges ni d'interopérabilité, sur des périodes souvent très longues.

Ce constat a conduit les trois grands financeurs à engager, chacun de son côté puis en commun, des réflexions sur la stratégie d'ensemble à adopter à l'avenir en matière d'études longitudinales afin d'en « maximiser l'impact »⁵⁴⁵⁵⁵⁶.

Les questions soulevées portent notamment sur l'évolution de la nature des besoins dans un contexte d'ouverture accrue des données médico-administratives et de développement des possibilités scientifiques liées à la génomique (et autres *-omics*).

Dans ce contexte, la préoccupation des commanditaires porte sur la nécessité de réaliser des harmonisations (de concepts, de variables, de modes de recueil...) permettant de tirer effectivement parti des appariements possibles, en articulation avec les projets propres au NHS et de faire dialoguer les cohortes entre elles. Leur souci est également de mettre en place un repérage (*discoverability*) et un accès facilité à leurs données ainsi qu'une évaluation effective de leurs utilisations.

3 Quelques enseignements ou sujets de réflexion pour la France

Un certain nombre d'enseignements tirés de l'expérience du Royaume-Uni -une nation occupant, répétons-le, une place de tête au regard de l'usage des cohortes parmi les pays développés-, peuvent éclairer les choix à venir et à faire en la matière dans notre propre pays. Ils découlent des pratiques déjà adoptées Outre-Manche comme des orientations nouvelles qui y sont d'ores et déjà envisagées ou en cours de planification.

L'ampleur, parfois considérable, et le spectre, très large, des grandes cohortes britanniques en population générale peuvent servir de modèle de référence, même si elle résulte d'une histoire et d'un contexte institutionnel spécifiques et d'une recherche traditionnellement fortement orientée sur les déterminants et les facteurs de risques en matière de santé.

À cet égard, la prolongation, annoncée en octobre 2018, de la cohorte phare *UK Biobank*, par une méga-cohorte de 5 millions de personnes doit être considérée avec la plus grande attention. Il s'agit d'un projet très ambitieux, annoncé par le gouvernement, faisant partie d'un vaste programme de prévention, l'*Accelerating Detection of Diseases (ADD) Programme*, prévoyant l'analyse génomique complète de 5 millions de britanniques, avec des finalités à la fois individuelles et collectives. Pour mémoire, *ADD Programme* sera doté de 79 millions de livres de financement public et de 160 millions de livres provenant des organisations caritatives et des industries intéressées.

Dans cette perspective, on peut aussi noter :

- le besoin, reconnu et intégré, d'une relance régulière de cohortes à la naissance, bien que la dernière tentative en la matière ait connu un échec majeur (en octobre 2015, huit mois après

⁵⁴ Medical Research Council, *Maximising the value of UK population cohorts*, 2014.

⁵⁵ Economic and Social Research Council, *Longitudinal Studies Strategic Review*, 2017.

⁵⁶ Wellcome Trust, *Longitudinal Population Studies Strategy*, 2017.

son lancement, un ambitieux projet de cohorte de population devant concerner 80.000 enfants nouveau-nés a dû, en effet, être abandonné en raison d'un recrutement insuffisant) ;

- l'expérience, ancienne et établie, d'une articulation forte entre les préoccupations et les recueils de données dans le domaine de la santé et ceux dans celui des sciences sociales, afin, notamment, d'explorer et de traiter les inégalités sociales de santé ;
- la combinaison, active et très efficace, de trois grands financeurs : le *MRC*, dans le domaine de la santé ; l'*Economic and Social Research Council (ESRC)*, couvrant le domaine des sciences humaines et sociales ; le *Wellcome Trust*, enfin, une grande fondation privée, si elle est propre au Royaume-Uni, peut néanmoins inspirer la France en termes de complémentarité des acteurs et d'accroissement des financements ;
- le caractère quasi systématique des appariements avec des BDMA, très riches en informations sociales et médicales. À cet égard, le Royaume-Uni dispose, avec le *NHS* d'un atout de premier ordre : un système unique de soins détenant des informations sanitaires longitudinales concernant toute la population, ouvrant, avec le chaînage, des possibilités considérables d'études cliniques, de santé publique et sociales.
- la nécessité de prioriser et de coordonner les décisions concernant les grandes cohortes en population : lancement, poursuite, financement. Ce qui suppose, au préalable, de disposer d'une vision globale du dispositif des cohortes. Comme l'indique expressément le *MRC*, "*it is important to take the overall UK cohort portfolio into account when making individual funding decisions*". C'est à cette fin que le choix a été fait, au Royaume-Uni, de créer une instance de supervision de l'ensemble des cohortes, l'*Expert Cross-Research Board Advisory Group*. Ce groupe a vocation à éclairer le *MRC*, mais aussi tous les autres financeurs, déjà actifs ou potentiels, dans leurs choix de soutien. L'intégration des considérations liées à la génomique, à l'épigénétique, au métabolisme ou à la microbiologie -ce que les Britanniques désignent sous le terme générique de *omics-*, mais aussi les données d'imagerie ou de technologies émergentes, est par ailleurs désormais prévue de façon systématique. Et cela afin de disposer d'un génotypage et d'un phénotypage des populations étudiées aussi complets que possible.

D'autres pistes sont en cours d'exploration Outre-Manche, qui peuvent aussi inspirer nos propres réflexions. L'ouverture aux technologies digitales en matière de recrutement et de suivi des cohortes est ainsi systématiquement envisagée. Les méthodes traditionnelles sont, en effet, coûteuses et potentiellement génératrices de taux élevés d'attrition. Les applications et l'usage de l'Internet offrent, à cet égard, des possibilités intéressantes pour l'inclusion des personnes et la collecte de données.

La connaissance de l'existence même des diverses cohortes, de même que l'accès aux données et aux échantillons qu'elles détiennent, représentent des enjeux décisifs et sont des sujets de préoccupation fortement mis en avant par les principaux financeurs C'est évidemment le cas au regard de la valorisation des travaux scientifiques et du développement de la recherche translationnelle. La pratique de l'accès libre et gratuit aux données, ou *open data*, est très répandue au Royaume-Uni. Ainsi, sur les 34 cohortes présentées dans cette annexe, 24 -dont toutes celles financées par le *MRC*- font partie de la structure mise en place aux fins de partage des données par le *MRC*, le *MRC Data Gateway Cohort Directory*. Un effort a été entrepris, compte tenu de la dispersion qui continue à exister en ce domaine, afin d'harmoniser les pratiques respectives de l'ensemble des cohortes dans ce domaine, l'objectif affiché étant celui d'une politique d'*open data* généralisée.