

Les cohortes pour les études et la recherche en santé

Tome 1 : Rapport

Mireille ELBAUM
Dr Aquilino MORELLE

Membres de l'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

SYNTHÈSE

Par lettre du 30 avril 2019, les directeurs de cabinet des ministres en charge des solidarités et de la santé, d'une part, de l'enseignement supérieur, la recherche et de l'innovation d'autre part, ont saisi l'inspection générale de l'éducation, du sport et de la recherche (IGÉSR) et l'inspection générale des affaires sociales (IGAS) d'une demande d'évaluation des cohortes épidémiologiques soutenues dans le cadre des grands programmes d'investissement dans la recherche, au premier chef le programme des investissements d'avenir (PIA) lancé en 2010. Les cohortes concernées sont trois études en population : E3N/E4N, RE-CO-NAI (comprenant elle-même deux cohortes d'enfants, ELFE et EPIPAGE2), et CONSTANCES, ainsi que dix cohortes de patients : CANTO (cancer du sein), CKD-Rein (insuffisance rénale chronique), COBLANCE (cancer de la vessie), CRYOSTEM (greffe de moelle), HEPATHER (hépatites B et C), HOPE-EPI (cancers de l'enfant), OFSEP (sclérose en plaques), PSY-COH (troubles bipolaires et schizophrénie), RADICO (maladies rares) ; s'y ajoute la cohorte MEMENTO (maladie d'Alzheimer), financée par le ministère chargé de la recherche.

L'instrument spécifique que constituent les « cohortes » occupe une place emblématique dans le monde de l'épidémiologie et de la santé publique. La démarche des « investissements d'avenir » l'a confortée en consacrant des moyens significatifs au soutien de ces projets identifiés comme des « cohortes », bien que leurs objectifs et leurs caractéristiques soient en réalité extrêmement différents. Dans le même temps, l'environnement des cohortes épidémiologiques, dont le nombre dépasse au total 250¹, comprend une variété d'autres instruments d'observation qui peuvent contribuer, dans le champ sanitaire et social, aux études et recherches à dimension longitudinale s'appuyant sur des données individuelles.

La mission a donc voulu adopter une vision large des enjeux d'observation et d'articulation des sources existant en ce domaine, tout en partant du « cœur de cible » que constituent les cohortes soutenues par le PIA. Elle a organisé une journée de travail avec chacune des équipes responsables de ces projets, leurs principaux partenaires et les chercheurs utilisant ces données. Elle a examiné les éléments disponibles sur les autres instruments d'étude et de recherche à portée longitudinale, ainsi que l'évolution en cours des bases de données médico-administratives. Elle a conduit une large palette d'entretiens visant à apprécier le bilan, les apports et l'utilisation des différentes cohortes, leurs perspectives de développement et les besoins ressentis en la matière, ainsi que les modes de pilotage et de coordination souhaités pour ce type d'outils.

La mission a par ailleurs voulu bénéficier d'un éclairage international, grâce à l'organisation d'un déplacement au Royaume-Uni, pays qui a une longue expérience et une ambition affirmée en matière de « très grandes cohortes » épidémiologiques, à dimension à la fois sanitaire et sociale.

Le rapport, qui s'appuie sur huit annexes dont l'une comprend des fiches de présentation des quatorze cohortes visitées par la mission, présente les résultats de ces travaux en partant d'un diagnostic de la situation actuelle (partie 1), puis d'une analyse des perspectives et incertitudes liées aux évolutions en cours (partie 2). Il envisage ensuite les actions de consolidation et de clarification nécessaires à court-moyen terme pour tirer le meilleur parti des cohortes déjà financées (partie 3). Il tente enfin de tracer, à plus long terme, les voies d'une politique plus ambitieuse et plus coordonnée

¹ Selon le portail Épidémiologie France.

de développement et de mobilisation des données longitudinales pour la recherche et les études en santé, en élargissant la réflexion à certains domaines sociaux ou médico-sociaux (partie 4).

La première partie du rapport met en évidence pour les quatorze cohortes examinées par la mission des éléments de bilan appréciables, mais aussi de réelles fragilités.

Elle part d'un rappel sur les spécificités de l'instrument que constituent les cohortes, définies comme « *le suivi longitudinal, à l'échelle individuelle, d'un groupe de sujets selon un protocole de recherche préétabli* ». Elle montre que, si cet instrument est considéré comme structurant pour l'expertise et la recherche épidémiologiques, certains de ses objectifs, centrés sur la mesure d'associations causales à partir de questions de recherche identifiées, se sont diversifiés et ont perdu de leurs spécificités. À côté des « cohortes », se sont en outre développés d'autres outils d'observation et de recherche à dimension longitudinale (bases de données médico-administratives, registres, enquêtes et panels du système statistique), avec lesquels la distinction n'est pas toujours aisée et qui, bien qu'ayant en principe des caractéristiques méthodologiques différentes, peuvent questionner l'intérêt de certaines cohortes et les articulations à envisager entre les différentes sources.

S'agissant du sous-ensemble constitué par les cohortes du PIA, le rapport rappelle qu'elles ont été financées dans le cadre de trois appels à projets distincts (« Cohortes », « Equipex » et « Infrastructures nationales Biologie-santé », qui leur ont permis de mobiliser des financements d'un montant significatif, atteignant au total 124 M€ sur la période 2011-2018 (dont 35 M€ pour l'infrastructure CONSTANCES)². Le rythme rapide et le recours à une démarche « par le bas » (*bottom-up*) pour la sélection des projets ont toutefois conduit à un caractère très disparate des dispositifs retenus, lesquels auraient mérité des approches différenciées selon les types de d'instruments : cohortes en population, cohortes de patients diagnostiqués pour certaines pathologies ou sujets à des prises en charge spécifiques, « plateformes » organisant des suivis longitudinaux à partir de données et méthodes partagées entre scientifiques et cliniciens.

En dépit de cette hétérogénéité, le bilan qui peut être tiré presque dix ans après le lancement du programme fait apparaître des résultats appréciables, détaillés pour chaque cohorte dans l'annexe 4 du rapport. Ils concernent, de façon transversale, le potentiel scientifique et le positionnement international des cohortes, largement reconnus par le jury international réuni par l'ANR en 2019, ainsi que leur impact en termes de structuration des milieux de recherche.

Une large part de ces acquis reste toutefois en devenir, du fait d'une maturité encore incomplète : le temps de mise en place et l'incomplétude de la plupart des projets, joints à un horizon scientifique long, impliquent des résultats seulement dans la durée. La visibilité et l'utilisation des cohortes par les acteurs du système de santé sont, en outre, loin d'être systématiques et varient fortement selon les types d'instruments et d'interlocuteurs.

Cet ensemble de cohortes présente par ailleurs de réelles fragilités, qui soit leur sont spécifiques, soit tiennent à un accompagnement resté lacunaire. Les premières peuvent faire suite à des difficultés scientifiques ou à des problèmes de gouvernance, tandis que les autres sont plus largement partagées et ont un caractère « systémique » : elles révèlent des problèmes généraux d'infrastructure technique concernant les biobanques et l'accès aux données médico-administratives, l'absence d'orientations communes dans les relations avec les milieux de recherche concernant l'ouverture et la mise à disposition des données, ainsi que des difficultés dans la gestion des ressources humaines et l'acquisition de compétences clés. Cette deuxième série de fragilités tient à la faiblesse du cadre commun et de l'accompagnement des projets mis en place par le PIA pour

² Hors crédits alloués à la cohorte MEMENTO.

encadrer le développement des cohortes et les mettre au service des communautés scientifiques comme du monde de la santé.

Au-delà de ces éléments de bilan, la deuxième partie du rapport souligne combien les perspectives et l’environnement qui sont ceux des cohortes demeurent, pour une large part, brouillés et incertains.

En premier lieu, la place des cohortes, et en particulier de celles soutenues par le PIA, est loin d’être clairement définie dans le paysage des outils statistiques et épidémiologiques portant sur les thèmes prioritaires en santé. D’autres cohortes importantes ont ainsi été développées et soutenues en dehors du PIA, sans que les articulations ou les complémentarités aient été forcément organisées entre ces différents instruments, pourtant tous financés à hauteur significative par des fonds publics. Plus généralement, au-delà des seules cohortes épidémiologiques, les différents domaines prioritaires de la santé publique sont couverts de façon fragmentée, dispersée et peu articulée par une diversité de dispositifs d’observation, qui, de façon directe ou indirecte, revêtent de plus en plus souvent une dimension longitudinale. Le rapport passe en revue plusieurs de ces domaines prioritaires des politiques de santé où le constat de cette faible articulation prévaut.

En second lieu, les relations des cohortes restent à concrétiser avec un système de données médico-administratives, lui-même en évolution forte et inachevée. La visibilité qu’ont les responsables de cohorte sur les potentialités et les modes d’accès à ces bases de données est aujourd’hui loin d’être assurée, ce qui induit attentisme et perplexité. Ainsi, en dépit des améliorations apportées par la création du système national des données de santé (SNDS), les changements intervenus entre 2016 et la mi-2019 dans ses modalités d’accès et d’utilisation ont conduit nombre de responsables de cohortes à retarder ou reprendre plusieurs fois les dossiers d’appariements projetés. Par ailleurs, le référentiel de sécurité défini par l’arrêté du 22 mars 2017, exigeant le stockage de ces données dans un environnement homologué ayant donné lieu à un audit préalable a également conduit certaines cohortes à demeurer dans l’attente ou à adopter des solutions temporaires, sans visibilité sur leur durée. Depuis la mi-2019, d’autres modifications de grande ampleur sont en outre introduites dans l’environnement technique des cohortes. Il s’agit d’abord de la création de la plateforme des données de santé (dite aussi *Health data Hub*), qui s’est vu conférer la mission d’effectuer en propre la mise à disposition des données d’un SNDS qui doit être élargi à un vaste ensemble d’informations cliniques (incluant les données des cohortes), ainsi que la réalisation directe de traitements pour le compte de tiers. D’un autre côté, dans le cadre de son projet « France Cohortes » (*cf. infra*), l’Inserm entend développer une plateforme informatique commune aux cohortes auxquelles l’organisme participe. Le projet ambitionne de fournir un système d’information mutualisé qui soit conforme au référentiel de sécurité du SNDS, décharge les chercheurs de la gestion informatique des cohortes, soit interopérable avec d’autres systèmes d’information, et doté de bulles sécurisées permettant aux chercheurs de traiter les données en toute sécurité. La mise en œuvre de ces deux projets doit avancer de façon concomitante au cours des années 2020 et 2021, sans que le contenu précis des services offerts de part et d’autre aux responsables de cohortes et, plus largement, aux chercheurs souhaitant bénéficier d’un accès à ces données, ait encore été précisé ; il en est de même des modalités de leur tarification et de leur articulation, entre eux et avec une autre plateforme développée par l’Inca.

La place des cohortes reste également à établir dans la restructuration en cours de la politique de recherche en santé. Les cohortes ont en effet souvent un caractère hybride, tenant à la fois du « projet de recherche » et de l’« infrastructure de recherche³ ». De ce fait, elles s’insèrent difficilement

³ Selon la définition de l’Inserm, « les infrastructures de recherche correspondent à des équipements, des ressources et des services utilisés par les scientifiques pour mener leurs recherches et favoriser l’innovation dans leur domaine ».

dans les structures traditionnelles de la recherche, qui distinguent, le plus souvent, les activités de recherche et de service. La dynamique du PIA ayant, quant à elle, été centrée sur la notion de « projet » porté par des établissements « partenaires », dont un « coordonnateur », il en ressort un paysage particulièrement complexe, avec des cohortes portées par des opérateurs privés (CANTO, PSY-COH) ou publics aux statuts divers (unité mixte de service –UMS- pour CONSTANCES et RECONAI, équipes internes à un laboratoire de recherche pour CKD REIN et E3N/E4N, association pour CRYOSTEM, fondations pour OFSEP et PSY-COH), parfois très intégrées à des unités de recherche ou, au contraire, sans liens autres que scientifiques avec les structures académiques. Envisager l'insertion des cohortes dans le paysage de la recherche, avec le renouvellement des grands laboratoires d'épidémiologie, mais aussi une réflexion en cours sur les contours et la gouvernance des « infrastructures de recherche », constitue donc un enjeu important pour apporter à ces cohortes un accompagnement dans la durée.

D'un autre côté, la promotion par le PIA d'un « modèle économique » tablant sur une large participation des partenaires privés pèse, avec le recul, sur l'inscription des cohortes dans la durée. Ce modèle, fondé sur la recherche d'un « effet de levier » aux financements publics, est la plupart du temps peu adapté à leur logique et à leurs objectifs scientifiques, notamment s'agissant des cohortes en population centrées sur les déterminants de santé et les facteurs de risque. Les responsables de projets ont néanmoins tenté d'y satisfaire par deux démarches complémentaires. La première a consisté à rendre payant l'accès aux données pour les autres équipes de chercheurs, avec des redevances dont les tarifs peuvent varier selon le type de relations engagées avec l'équipe responsable de la cohorte (collaboration ou prestation de service pour l'accès aux données). Ce principe, qui ne fait pas consensus, donne lieu à des applications diverses, et ne conduit en tout état de cause qu'à une contribution modeste au financement des cohortes. La seconde démarche parie sur la mobilisation de financements privés qui peuvent eux-mêmes revêtir plusieurs formes (outre le mécénat, des partenariats public/multiprivés avec un ensemble d'industriels ou des partenariats spécifiques au projet). Le bilan fait par la mission montre que les financements obtenus proviennent d'un nombre réduit de partenaires et restent globalement modestes, bien que non négligeables. Surtout, les cohortes en population en bénéficient très peu, les contributions des industriels allant pour l'essentiel aux cohortes de patients dont le champ correspond à des opportunités particulières de développement de produits ou de thérapies, dans un créneau de temps limité. La viabilité d'un modèle économique qui tablerait sur un financement important et durable des grandes cohortes nationales par des partenaires privés pose donc des questions, qui ne devraient pas être laissées à la seule initiative des porteurs de projet, mais abordées dans un cadre global.

Enfin, et surtout, l'absence actuelle d'une politique ambitieuse et coordonnée de mobilisation des données longitudinales pour la recherche en santé ne fournit ni priorités ni perspectives claires pour guider leur évolution.

Ce manque d'ambition ressort d'abord de l'examen des expériences étrangères, qui, en matière de grandes cohortes en population générale, ont été fécondes et conduites sur longue période. Elles se sont ainsi considérablement développées partir des années soixante-dix, avec la reconnaissance du caractère multifactoriel de l'étiologie des maladies chroniques. Cet accroissement s'est doublé d'un perfectionnement qualitatif : suivi à long terme d'un nombre massif de participants, phénotypage des maladies de haute qualité visant à saisir les rôles respectifs des facteurs comportementaux et environnementaux, ainsi que leur interaction avec les caractères génétiques, appariements avec les registres ou données administratives. **L'expérience du Royaume-Uni est en particulier précieuse à analyser,** dans la mesure où ce pays possède un éventail unique de grandes cohortes de population, avec par exemple la cohorte phare *UK Biobank* qui suit 500.000 personnes âgées de 40 à 69 ans incluses entre 2006 et 2010, et l'émergence d'un nouveau projet de méga-cohorte de cinq millions de participants. Le financement de la recherche épidémiologique y est

caractérisé par l'action complémentaire du *Medical Research Council (MRC)* et de deux autres institutions, le *Wellcome Trust*, organisation caritative dédiée à la promotion de la recherche biomédicale et l'*Economic and Social Research Council (ESRC)*, autre partenaire majeur du financement de la recherche. Le Royaume-Uni offre aussi un environnement scientifique et réglementaire particulièrement favorable aux cohortes, avec la possibilité d'appariements quasi systématiques à des bases de données administratives très riches en informations sociales comme médicales, en provenance notamment du *National Health System (NHS)*. Dans ce cadre, et tout en conservant la préoccupation d'une articulation forte entre les données afférentes à la santé et aux sciences sociales, les grands acteurs ressentent cependant la nécessité de mieux prioriser et coordonner les décisions relatives aux cohortes, d'harmoniser leurs pratiques et d'aller vers une politique d'*open data* généralisée.

Au regard de ces expériences, les projets français de développement d'une politique plus ambitieuse et coordonnée de recherche en santé publique, qui ont donné lieu à plusieurs tentatives, sont restés peu opérants, et se sont heurtés tant à une faiblesse de moyens qu'à un pilotage incertain. L'absence de mise en œuvre des orientations esquissées en la matière a laissé pendante la question de l'articulation entre priorités de fond et outils d'observation, qui demeure traitée de façon cloisonnée par une variété d'acteurs ayant tendance à promouvoir leurs propres dispositifs. La dissociation s'est, par exemple, approfondie entre les instruments d'observation promus, en matière de santé, sous l'égide du système statistique public et les cohortes épidémiologiques relevant du monde de la recherche, avec, entre les deux, peu d'échanges réciproques.

S'agissant plus spécifiquement des cohortes épidémiologiques, les tentatives de coordination et de mutualisation sont demeurées embryonnaires. Elles ont donné lieu à des initiatives intéressantes, mais qui sont restées ponctuelles et d'ampleur limitée (portail Aviesan, club Cohortes de l'Iresp, *Cohort Innovation Day* organisé avec l'Ariis).

Dans ce contexte, la relance qui vient d'être récemment opérée, au travers du projet France Cohortes porté par l'Inserm et centré sur la mutualisation et l'accompagnement des cohortes, est une initiative intéressante mais qui comporte encore nombre d'imprécisions. Son périmètre (les seules cohortes financées par le PIA ou au-delà), son objet (par-delà la seule question des systèmes d'information), les appuis qui seront proposés dans des domaines comme les biobanques, l'accompagnement juridique et administratif des projets ou les actions de communication, son calendrier et l'ensemble des moyens mis en œuvre doivent encore être précisés, et peuvent poser des questions de priorités. Le projet laisse en outre entières les questions de l'animation scientifique des cohortes et de la définition d'orientations et de règles communes en matière de relations avec les différents partenaires privés ou publics ainsi que de mise à disposition des données des communautés académiques.

La troisième partie du rapport envisage les actions à conduire à court-moyen terme, pour répondre aux besoins de consolidation et de clarification des cohortes soutenues dans le cadre du PIA, et tirer ainsi tout le parti possible des investissements réalisés depuis dix ans.

Un premier besoin concerne la consolidation de la gouvernance et du pilotage de certaines cohortes, dont l'examen montre des défauts de fonctionnement ou des limites qui pèsent sur la poursuite du projet. **La consolidation doit aussi concerner le potentiel scientifique des cohortes dont il importe, en les inscrivant dans un temps long, qu'elles puissent aller jusqu'au bout de la production de leurs résultats.** Cela implique la poursuite des suivis et de l'exploitation des informations recueillies jusqu'au terme des périodes scientifiquement pertinentes, la préservation et la meilleure utilisation possible des échantillons biologiques collectés ainsi que la réalisation effective des analyses génomiques et des diverses recherches envisagées dans le domaine des « –

omics »⁴, dans l'objectif de tirer le meilleur parti des prélèvements effectués « au long cours » auprès des participants. Il est également indispensable que les appariements avec le SNDS ne soient pas de nouveau retardés et trouvent rapidement des voies opérationnelles, fussent-elles temporaires.

Ces orientations impliquent qu'un « socle » récurrent de ressources publiques soit effectivement garanti, à hauteur suffisante, à l'ensemble de ces cohortes pour les années à venir. Ces financements devraient principalement viser la consolidation des dispositifs, sans forcément inclure une réouverture de leurs inclusions ou un élargissement de leur spectre d'intérêt. Les décisions en cours d'approbation pour la période 2020-2024 semblent globalement aller dans ce sens, mais sans être aisément corrélables au contenu des demandes d'aide ou aux perspectives de cofinancement présentées par les porteurs de projets. L'attribution de ces enveloppes devrait en outre être conditionnée au fait que ces projets avancent effectivement, en dépassant d'éventuels blocages qui ont pu être constatés. À l'opposé, la perspective d'arrêter le financement de certains projets peut paraître problématique au vu des efforts déjà consentis et dans la mesure où, d'après les constatations de la mission, aucun d'eux n'a réellement failli.

Le besoin de stabilisation concerne aussi certains personnels recrutés sur CDD et l'apport de compétences, notamment dans les métiers ayant trait à la gestion et l'exploitation des données. Plusieurs de ces métiers pourraient donner lieu à mutualisation, notamment dans le cadre du projet France Cohortes, qui peut offrir, en ce domaine, une perspective intéressante.

Un dernier volet d'une politique visant à tirer le meilleur parti des cohortes existantes consiste à procéder à des clarifications dans leur environnement immédiat, dans la mesure où responsables, chercheurs et financeurs ont besoin d'être au clair sur plusieurs points clés. Il s'agit d'abord des possibilités qui seront réellement offertes au fil du temps, en termes d'informations cliniques, par les données médico-administratives enrichies : des perspectives tangibles devraient être clairement affichées en la matière, par exemple quant à l'intégration, évoquée depuis plusieurs années, des résultats des examens biologiques réalisés en ville. De même, la politique des registres de morbidité ou de patients est un élément important de l'univers des cohortes et méritera d'être redéfinie rapidement par le ministère chargé de la santé, suite aux travaux lancés au sein du HCSP.

Enfin, les responsables de cohortes ont besoin d'être éclairés au plus vite, en matière de systèmes d'information, sur le rôle respectif et la répartition des fonctions qui s'opérera entre le projet France Cohortes, le *Health Data Hub* et la plateforme de l'Inca. Cet éclairage doit à la fois porter sur l'articulation technique et fonctionnelle des nouveaux systèmes, les lieux et structures destinés à réaliser les appariements entre les données de chaque cohorte et celles du SNDS, les modalités de stockage sécurisé des informations qui en découlent, ainsi que, point important, sur la nature et le coût des services qui seront proposés à cette occasion.

La quatrième partie du rapport esquisse enfin les voies d'une politique à moyen-long terme de développement et de mobilisation des données longitudinales dans les champs sanitaire et social, où les cohortes viendraient s'inscrire de façon cohérente, en fonction de leurs apports propres. L'horizon envisagé est de dix ans à quinze ans dans la mesure où, à cette échéance, beaucoup de cohortes « patients » du PIA auront fini leur vie naturelle, et où, s'agissant des grandes cohortes en population, les enfants suivis par ELFE auront plus de vingt ans alors que l'étude CONSTANCES, si elle n'est pas ré-ouverte, n'aura dans son échantillon que peu d'individus de moins de 40 ans.

⁴ Qui recouvrent métabolomique, protéomique, microbiomique.

Cette politique devrait reposer selon la mission sur deux soubassements préalables : la définition en amont de priorités programmatiques reconnues et financées pour la recherche en santé et la mise en place d'une coordination spécifique en matière d'outils d'observation, pour lesquelles les pouvoirs publics devraient assumer la fonction d'impulsion stratégique qui leur revient en propre. Ce programme devrait être assorti d'une cartographie en amont des outils d'observation déjà disponibles pour les principales priorités thématiques et de ceux qui seraient nécessaires pour les éclairer dans la durée, sur la base d'une réflexion prospective. Les dispositifs de suivi longitudinal sont appelés à prendre une place essentielle dans cette démarche, sachant que, en leur sein, les besoins spécifiques en cohortes devraient être établis en fonction des informations déjà disponibles dans les bases de données administratives et des manques repérés en la matière, en s'adaptant au fil du temps selon l'évolution des différentes sources. Ce pilotage devrait associer l'ensemble des parties prenantes (ministères, système statistique public, organismes de recherche, caisses de sécurité sociale, agences sanitaires, fédérations d'établissements...), de façon à prendre en compte leurs besoins et favoriser leur implication, y compris au travers de soutiens humains, informatiques ou logistiques.

Il reviendra à ce programme et à ces instances de coordination de définir précisément les projets à développer et les moyens à mobiliser pour ces différents outils d'observation, sachant que, **s'agissant spécifiquement des cohortes épidémiologiques, trois grands besoins mériteront d'être pris en compte. Un premier besoin concerne les cohortes en population, pour lequel la mission préconise que la France dispose en permanence de deux grandes infrastructures :** une cohorte d'enfants à la naissance, renouvelée régulièrement et une cohorte d'adultes à large spectre, visant de l'ordre de 500 000 participants. Cette cohorte devrait dépasser les limites actuelles du dispositif CONSTANCES en visant une représentativité statistique régulièrement entretenue, en recourant aux technologies digitales, en s'articulant avec les bases de données administratives, y compris socio-fiscales et en mobilisant une variété de ressources en termes de résultats d'exams cliniques, prélèvements biologiques, d'imagerie, de génétique et de « -omics ». Ces deux cohortes auraient vocation à être considérées comme des infrastructures de recherche à portée nationale, et gérées comme telles, sans dépendre d'une équipe ou d'un centre de recherche particuliers, mais d'un groupement scientifico-administratif créé à cet effet. Dans ce cadre, une offre de services devrait être prévue en direction des utilisateurs, dont le coût devrait être pour l'essentiel pris en compte dans le budget de fonctionnement de ces structures.

Un deuxième besoin concerne les cohortes de patients exposés, atteints ou traités pour des pathologies spécifiques, dont le maintien d'un portage et d'un financement ciblé restera indispensable à l'avenir, même s'il convient d'en prioriser soigneusement le contenu au regard de l'enrichissement progressif des bases de données médico-administratives ; ces cohortes devraient à cet égard répondre à quelques priorités spécifiques : assurer un suivi régulier et systématique de la prise en charge des patients atteints de maladies chroniques, asseoir la présence française à l'international à travers une politique scientifique de « niches » ; suivre les répercussions sanitaires d'éventuelles catastrophes environnementales ou industrielles, mettre en place un suivi particulier des traitements innovants susceptibles d'apparaître dans des horizons très courts.

Un troisième type de besoins de suivis longitudinaux, pouvant ou non prendre la forme de cohortes, concerne enfin d'autres domaines que la seule épidémiologie : ils portent notamment, dans le domaine de la régulation du système de santé, sur l'organisation des parcours de soins et la couverture sociale en matière de santé ; ils concernent également, dans les domaines sociaux et médico-sociaux, le suivi des enfants et des adolescents handicapés ou pris en charge par l'aide sociale à l'enfance, ainsi que de certaines populations exposées à des difficultés particulières (chômeurs, migrants...).

Un dernier pilier indispensable d'une politique à moyen-long terme de données longitudinales en santé concerne l'accompagnement des responsables scientifiques de ces dispositifs et des chercheurs désireux d'y accéder. Cet accompagnement devrait porter sur trois axes : la mise en place d'infrastructures « support » dans le domaine des biobanques et de la structuration des bases de données issues d'appariements avec les données administratives ; la prise en compte d'une variété de besoins en matière d'appui juridique, budgétaire, scientifique et de communication ; la définition harmonisée de modes de partenariat avec le monde économique, adaptée à la diversité des types de cohortes, laquelle ne peut dépendre exclusivement des porteurs du projet ou d'une structure de valorisation, mais doit incomber aux pouvoirs publics. Une doctrine commune, en vue d'une maîtrise collective permettant une utilisation transparente et partagée, semble à cet égard devoir être élaborée, en y associant Inserm Transfert, la HAS et l'ANSM, sous l'égide des deux ministères concernés.

Le rapport présente vingt recommandations visant à la mise en œuvre de ces orientations et concernant à la fois le cadre d'analyse et d'élaboration de la politique des cohortes, les actions de consolidation et de clarification à mener à court- moyen terme, et la politique de données longitudinales à développer à moyen-long terme.

RECOMMANDATIONS DE LA MISSION

n°	Recommandation	Priorité	Autorité responsable	Échéance
Thème 1 Le cadre d'analyse et d'élaboration de la politique des cohortes				
1	<p>Resituer les cohortes dans un cadre d'analyse adapté :</p> <ul style="list-style-type: none"> • distinguer, dans les politiques mises en œuvre, les grandes cohortes en population, les études longitudinales de patients concernés par certaines pathologies ou traitements, et les «plateformes » organisant des suivis longitudinaux à partir de méthodes et/ou de sources partagées ; • englober, dans la définition de ces politiques, les cohortes en population financées par des voies autres que le PIA, et les dispositifs de suivi longitudinal placés sous l'égide des organismes de protection sociale ou des agences sanitaires. 	1	Ministères chargés de la santé et de la recherche	2020
2	Actualiser, compléter et maintenir à jour le recensement du portail Epidémiologie France sur les bases de données individuelles en santé utiles pour l'expertise et la recherche en santé publique.	1	Inserm Aviesan	2020
3	Organiser un partage d'informations et une réflexion autour des principales expériences étrangères en matière de cohortes épidémiologiques.	2	Ministères chargés de la santé et de la recherche	2020-2021
4	Considérer, dans la réflexion en cours sur la politique de structuration de la recherche, les cohortes en population comme des « infrastructures » à vocation nationale.	1	Ministère chargé de la recherche	2020
5	Remettre en question l'idée, s'agissant des cohortes soutenues par le PIA, d'un modèle économique visant un relais pérenne par des partenaires privés et d'un mode de pilotage différencié de celui des autres projets ou infrastructures du même type.	1	Secrétariat général pour l'investissement Ministères chargés de la santé et de la recherche	2020

Thème 2 Les actions de consolidation et de clarification à mener à court- moyen terme vis-à-vis des cohortes du PIA				
6	<p>S'assurer d'un mode de pilotage et de fonctionnement consolidé pour l'ensemble des cohortes :</p> <ul style="list-style-type: none"> en initiant une redéfinition et/ou une clarification des modes de pilotage propres à certaines cohortes; en demandant à plusieurs de ces cohortes d'améliorer le mode de fonctionnement de leurs instances de gouvernance, sans forcément en alourdir le pilotage <p>(Cf. annexe 4)</p>	1	<p>Secrétariat général pour l'investissement</p> <p>Agence nationale de la recherche</p>	2020
7	Revoir les modalités d'évaluation des cohortes en organisant des visites de sites par des équipes d'évaluateurs d'origine à la fois scientifique et administrative	2	<p>Secrétariat général pour l'investissement,</p> <p>Agence nationale de la recherche</p>	2022
8	<p>Tirer parti au plus vite des possibilités déjà offertes par le SNDS :</p> <ul style="list-style-type: none"> en favorisent la réalisation rapide des appariements prévus avec le SNDS, au besoin en recourant à des conditions temporaires de stockage sécurisé de données, évitant d'attendre la mise en place opérationnelle effective de France Cohortes ou du <i>Health data Hub</i> ; en poursuivant et amplifiant, dans le cadre d'une coordination <i>ad hoc</i>, les travaux initiés dans le cadre des projets ReDSiam afin de développer au plus vite l'identification robuste des pathologies dans le SNDS. 	1	<p>Inserm</p> <p>Cnam,</p> <p>Responsables de cohortes</p>	2020
9	<p>Stabiliser un socle de financement public pluriannuel permettant aux cohortes ayant bénéficié de financements publics de tirer le meilleur parti des investissements réalisés :</p> <ul style="list-style-type: none"> en mettant en place les analyses génomiques (ou autres <i>-omics</i>) permises par les prélèvements collectés ; en poursuivant les suivis et l'exploitation des informations recueillies jusqu'au terme de la période scientifiquement pertinente ; en employant un noyau significatif de personnels permanents. 	1	<p>Ministère chargé de la recherche</p> <p>Secrétariat général pour l'investissement</p> <p>Agence nationale de la recherche</p>	2020
10	Étudier les conditions d'une possible mobilisation de données socio-économiques détenues par les services statistiques publics pour fiabiliser et enrichir les données des cohortes en population.	2	<p>Insee</p> <p>CNAF</p> <p>Inserm</p>	2021

11	<p>Préciser rapidement à l'intention des responsables de cohortes les évolutions prochaines des bases de données médico-administratives et le rôle des plateformes d'appui qui leur seront proposées :</p> <ul style="list-style-type: none"> • en présentant aux responsables de cohortes un schéma clair des données actuellement et prochainement accessibles <i>via</i> les données médico-administratives, et précisant l'horizon réaliste de leur mise à disposition ; • en clarifiant les fonctions et rôles précis qui seront exercés en matière d'appariements avec le SNDS, de stockage sécurisé des données, et d'ouverture aux chercheurs académiques respectivement par France Cohortes, l'Inca et le <i>Health Data Hub</i> ; • en définissant et en priorisant à l'intention des responsables de cohortes l'offre de services du futur dispositif France Cohortes, incluant le renforcement des fonctions juridiques, administratives et budgétaires et l'animation à visée scientifique ou de communication. 	1	<p>Ministères chargés de la santé et de la recherche</p> <p>Inserm</p> <p><i>Health data Hub</i></p>	2020-2021
----	---	---	--	-----------

Thème 3 La politique de données longitudinales à développer à moyen-long terme				
12	<p>Définir à l'horizon de cinq ans des priorités s'inscrivant dans les stratégies nationales de santé et de recherche et un programme à moyen-long terme insérant les dispositifs de suivi longitudinaux dans une politique ambitieuse de recherche en santé et recensant, par grande thématique, les outils d'observation disponibles et nécessaires.</p>	1	Ministères chargés de la santé et de la recherche	2021
13	<p>Associer les parlementaires et les représentants de la société civile à l'élaboration et au choix de ces priorités de programmation.</p>	1	Ministères chargés de la santé et de la recherche	2021-2022
14	<p>Assurer un pilotage partenarial des outils d'observation relevant des priorités ainsi définies :</p> <ul style="list-style-type: none"> • en mettant en place une instance opérationnelle de coordination des dispositifs d'observation et de suivi longitudinal en santé associant les ministères concernés, les organismes de recherche, le système statistique public, les agences sanitaires et les caisses de sécurité sociale ; • en expérimentant des modes d'appels à projets articulant des priorités thématiques fixées par les pouvoirs publics et des recommandations de méthode issues du monde de la recherche ; • en évaluant régulièrement les dispositifs ou infrastructures ayant fait l'objet d'un soutien ; • en veillant à l'adaptation des projets en fonction de l'évolution des techniques ou des données disponibles. 	1	<p>Ministères chargés de la santé et de la recherche</p> <p>Agence nationale de la recherche</p>	2021-2022

15	<p>Remettre en place à l’horizon de dix ans une à deux grandes cohortes en population générale :</p> <ul style="list-style-type: none"> • en programmant sur le long terme l’existence, de façon complémentaire aux bases de données médico-administratives, d’une grande cohorte d’adultes et d’une cohorte d’enfants à la naissance ; • en donnant à ces grandes cohortes une ambition affirmée, en visant une représentativité en population générale, une articulation avec les systèmes médico administratifs et une variété de ressources en termes de prélèvements biologiques, d’imagerie, de génétique et de <i>-omics</i> ; • en prévoyant pour ces cohortes un mode de pilotage spécifique reconnaissant leur fonction d’infrastructure de recherche à portée nationale. 	1	<p>Ministères chargés de la santé et de la recherche</p> <p>Agence nationale de la recherche</p>	2030
16	<p>Définir des priorités et des modes opératoires concernant les dispositifs de suivi longitudinal à lancer concernant des pathologies ou des traitements spécifiques :</p> <ul style="list-style-type: none"> • en incluant dans ces priorités le suivi de certaines maladies chroniques, des traitements innovants et coûteux ainsi que des domaines de « niche » où la recherche française est en position d’excellence ; • en prévoyant pour les cohortes relevant de ces priorités des circuits de financement spécifiques. 	1	<p>Ministères chargés de la santé et de la recherche</p> <p>Agence nationale de la recherche</p>	2021
17	<p>Intégrer aux priorités nationales concernant les besoins d’observation et de suivi longitudinal à long terme :</p> <ul style="list-style-type: none"> • dans le domaine de la santé, les questions d’organisation des parcours de soins et de couverture des frais de santé ; • dans les domaines social et médico-social, le devenir des enfants handicapés et de ceux pris en charge par l’aide sociale à l’enfance. 	1	Ministères chargés de la santé et de la recherche	2020-2021
18	<p>Assurer aux dispositifs longitudinaux prioritaires un environnement et des mécanismes d’accompagnement adaptés :</p> <ul style="list-style-type: none"> • en entreprenant une structuration au niveau national des biobanques destinées à la recherche sur le territoire français ; • en consolidant et développant l’appui aux responsables de cohortes en matière juridique, budgétaire, et de communication ; • en soutenant, dans les domaines de l’épidémiologie et de la santé publique, l’attractivité des métiers dans les domaines <i>du data management</i> et des biostatistiques. 	1	<p>Ministères chargés de la santé et de la recherche</p> <p>Inserm</p>	2020-2021

	<ul style="list-style-type: none"> en mettant en place, <i>via</i> le <i>Health data Hub</i> ou le CASD, des processus d'appariement systématique entre les données des cohortes épidémiologiques, notamment en population, et un vaste champ de données administratives, non seulement en santé, mais aussi médico-et fiscal-social (Insee, CNAF...). 			
19	<p>Clarifier et faciliter les processus d'accès des chercheurs académiques aux données des études épidémiologiques financées sur fonds publics :</p> <ul style="list-style-type: none"> en définissant des règles homogènes d'accès aux données valables pour l'ensemble des sources et des dispositifs ; en prévoyant soit un accès gratuit à ces infrastructures, soit une participation financière limitée au coût de la prestation de service réalisée pour chaque projet. 	1	Ministères chargés de la santé et de la recherche Inserm	2020-2021
20	<p>Établir une doctrine claire, publique et partagée entre organismes de recherche et autorités sanitaires (HAS, ANSM) concernant les partenariats entre firmes industrielles et dispositifs de suivi longitudinal financés sur fonds publics, et s'appliquant quel qu'en soit l'organisme porteur.</p>	1	HAS ANSM Inserm Transfert	2020-2021

SOMMAIRE

SYNTHESE	3
RECOMMANDATIONS DE LA MISSION	11
RAPPORT	19
1 DES COHORTES DU PIA A L'UTILISATION DES DONNEES LONGITUDINALES POUR LA RECHERCHE EN SANTE : UNE IMPULSION APPRECIABLE, DES ACQUIS EN DEVENIR ET DE REELLES FRAGILITES	21
1.1 UN INSTRUMENT STRUCTURANT DONT LES SPECIFICITES S'ESTOMPENT FACE DES DISPOSITIFS VOISINS OU CONCURRENTS	21
1.1.1 <i>Les cohortes : un instrument structurant pour l'expertise et la recherche épidémiologiques</i>	21
1.1.2 <i>Des caractéristiques qui tendent à perdre de leurs spécificités</i>	22
1.1.3 <i>Des dispositifs de suivi longitudinal proches ou concurrents.....</i>	23
1.2 LES COHORTES DU PIA : UN ENSEMBLE HETEROGENE, QUI AURAIT JUSTIFIE DES APPROCHES DIFFERENCIEES SELON LES TYPES D'INSTRUMENTS, MAIS QUI PRESENTE DES ELEMENTS DE BILAN APPRECIABLES	24
1.2.1 <i>Des dispositifs à caractère divers, sélectionnés à travers trois appels à projets du PIA et mobilisant des financements significatifs.....</i>	24
1.2.2 <i>Une typologie nécessaire pour asseoir la pertinence des approches et propositions à envisager.....</i>	28
1.2.3 <i>Des éléments de bilan appréciables en termes de potentiel scientifique, de positionnement international et de structuration des milieux de recherche.....</i>	30
1.3 UNE MATURETE ENCORE INCOMPLETE ET UNE UTILISATION TRES VARIABLE PAR LES ACTEURS DU SYSTEME DE SANTE.....	31
1.3.1 <i>Des délais de construction et des horizons temporels longs impliquant une maturité et des acquis encore en devenir.....</i>	32
1.3.2 <i>Une visibilité et des utilisations très variables par les acteurs du système de santé</i>	32
1.4 DE REELLES FRAGILITES, SPECIFIQUES AUX PROJETS OU TENANT A DES DEFICITS GENERAUX D'ACCOMPAGNEMENT	35
1.4.1 <i>Des fragilités spécifiques à certains projets, suite à des difficultés scientifiques ou à des problèmes de gouvernance.....</i>	35
1.4.2 <i>Des difficultés plus largement partagées, traduisant des déficits généraux d'accompagnement en termes d'infrastructures techniques, de politique scientifique et de gestion des compétences.....</i>	35
2 DES PERSPECTIVES ET UN ENVIRONNEMENT BROUILLES ET INCERTAINS	39
2.1 UNE PLACE INCERTAINE DES COHORTES DU PIA DANS LE PAYSAGE DES OUTILS STATISTIQUES ET EPIDEMIOLOGIQUES PORTANT SUR LES THEMES PRIORITAIRES EN SANTE	39
2.1.1 <i>D'autres cohortes épidémiologiques soutenues de façon séparée par les organismes publics</i>	39
2.1.2 <i>Des articulations insuffisamment maîtrisées entre les dispositifs d'observation des grands domaines de la santé publique.....</i>	41
2.2 DES INCERTITUDES SUR LE CONTENU ET LES MODALITES DES LIENS AVEC UN SYSTEME DE DONNEES MEDICO-ADMINISTRATIVES LUI-MEME EN DEVENIR.....	42
2.2.1 <i>Des évolutions du SNDS.....</i>	42
2.2.2 <i>... aux nouveaux environnements créés par le Health data Hub et le projet France Cohortes.....</i>	43
2.3 DES INCERTITUDES SUR L'INSERTION DES DIVERS TYPES DE COHORTES DANS LA STRUCTURATION DU PAYSAGE PARTICULIEREMENT COMPLEXE DE LA RECHERCHE EN SANTE	45
2.4 DES DOUTES SUR LA VIABILITE D'UN « MODELE ECONOMIQUE » TABLANT SUR UNE LARGE PARTICIPATION DES PARTENAIRES PRIVES	47
2.5 L'ABSENCE D'UNE POLITIQUE AMBITIEUSE ET COORDONNEE DE MOBILISATION DES DONNEES LONGITUDINALES POUR LA RECHERCHE EN SANTE	49
2.5.1 <i>Un manque d'ambition global au regard de certaines expériences étrangères</i>	50
2.5.2 <i>En France, des projets de développement d'une politique coordonnée de recherche en santé publique restés peu opérants.....</i>	53
2.5.3 <i>Des tentatives embryonnaires de coordination et de mutualisation des cohortes.....</i>	54
2.5.4 <i>Une relance récente, mais encore imprécise, au travers du projet Inserm France Cohortes, dont les contours et les priorités peuvent toutefois être discutés.....</i>	55

3	À COURT-MOYEN TERME, DES BESOINS DE CONSOLIDATION ET DE CLARIFICATION	57
3.1	DES BESOINS DE CONSOLIDATION DANS PLUSIEURS DOMAINES	57
3.1.1	<i>Des consolidations en termes de pilotage et de gouvernance.....</i>	57
3.1.2	<i>Des consolidations en termes d’horizon et de potentiel scientifiques</i>	58
3.2	DES BESOINS RECURRENTS D’UN « SOCLE » DE RESSOURCES PUBLIQUES DANS LES ANNEES A VENIR.....	59
3.2.1	<i>En termes de financement.....</i>	61
3.2.2	<i>En termes de stabilisation et de consolidation de certains personnels.....</i>	61
3.3	DES BESOINS DE CLARIFICATION CONCOMITANTS DE L’ENVIRONNEMENT IMMEDIAT DES COHORTES	62
4	À MOYEN-LONG TERME, LA NECESSITE D’UNE POLITIQUE PLUS AMBITIEUSE ET COORDONNEE DE MOBILISATION DES DONNEES LONGITUDINALES POUR LA RECHERCHE EN SANTE.....	65
4.1	LA NECESSAIRE CONSTRUCTION DE PRIORITES PROGRAMMATIQUES ET D’UNE COORDINATION D’ENSEMBLE DES OUTILS D’OBSERVATION.....	65
4.1.1	<i>Le besoin en amont de priorités programmatiques pour la recherche en santé.....</i>	65
4.1.2	<i>La nécessité en aval d’une coordination spécifique en matière d’outils d’observation, assise sur une gouvernance adaptée et des mécanismes d’évaluation.....</i>	66
4.2	LA NECESSAIRE PRISE EN COMPTE DE TROIS GRANDS TYPES DE BESOINS	67
4.2.1	<i>L’installation à l’horizon de dix ans d’une à deux grandes cohortes en population, dotées de plusieurs caractéristiques</i>	67
4.2.2	<i>La définition de priorités et de modes d’appui adaptés aux dispositifs de suivi longitudinaux centrés sur des pathologies ou des patients traités</i>	69
4.2.3	<i>Des besoins de suivis longitudinaux à long terme, pouvant ou non prendre la forme de cohortes, dans d’autres domaines que la seule épidémiologie.....</i>	70
4.3	DES BESOINS D’ACCOMPAGNEMENT A PRENDRE EN COMPTE DE FAÇON PARALLELE ET ESSENTIELLEMENT MUTUALISEE	72
4.3.1	<i>Des infrastructures « support » à mettre en place pour l’ensemble des dispositifs.....</i>	72
4.3.2	<i>Le renforcement de l’appui aux porteurs de cohortes en matière juridique, budgétaire, de communication et d’animation scientifique.....</i>	73
4.3.3	<i>Un pilotage plus cohérent et harmonisé des relations avec les industriels.....</i>	74
	LETRE DE MISSION	77
	LISTE DES ANNEXES	79
	LISTE DES PERSONNES RENCONTREES HORS VISITES DES COHORTES.....	81
	SIGLES UTILISES	105

RAPPORT

Introduction

Par lettre du 30 avril 2019, les directeurs de cabinet des ministres en charge des solidarités et de la santé, d'une part, de l'enseignement supérieur, la recherche et de l'innovation d'autre part, ont saisi l'inspection générale de l'éducation, du sport et de la recherche (IGÉSR) et l'inspection générale des affaires sociales (IGAS) d'une demande d'évaluation des cohortes épidémiologiques soutenues dans le cadre des grands programmes d'investissement dans la recherche, au premier chef le programme des investissements d'avenir (PIA) lancé en 2010. Les cohortes concernées sont trois études en population : E3N/E4N, RE-CO-NAI (comprenant elle-même deux cohortes d'enfants, ELFE et EPIPAGE2), et CONSTANCES, ainsi que dix cohortes de patients : CANTO (cancer du sein), CKD-Rein (insuffisance rénale chronique), COBLANCE (cancer de la vessie), CRYOSTEM (greffe de moelle), HEPATHER (hépatites B et C), HOPE-EPI (cancers de l'enfant), OFSEP (sclérose en plaques), PSY-COH (troubles bipolaires et schizophrénie), RADICO (maladies rares)⁵ ; s'y ajoute la cohorte MEMENTO (maladie d'Alzheimer), non financée par le PIA, mais, initialement, par la Fondation Alzheimer, puis le ministère chargé de la recherche.

Cette lettre demande aux inspections de réfléchir à la fois aux modalités à envisager pour leur consolidation, leur pilotage, leur gouvernance, leur cadre de gestion et leur modèle économique (financement et partenariats public-privé), et à la stratégie nationale à poursuivre en matière de cohortes. Cette mission a été confiée à MM. Bertrand MINAULT et Henri RIBIERAS pour l'IGÉSR, et au Dr Aquilino MORELLE et à Mme Mireille ELBAUM pour l'IGAS.

L'instrument spécifique que constituent les « cohortes », objet de la lettre de mission, occupe une place particulière, et même emblématique, dans le monde de l'épidémiologie et de la santé publique. La démarche des « investissements d'avenir » a contribué à illustrer et conforter cette place, en consacrant des moyens significatifs au soutien de quatorze projets, directement issus de propositions d'équipes de recherche et identifiés sous le vocable de « cohortes », bien que leurs objectifs et leurs caractéristiques soient en réalité extrêmement différents.

Dans le même temps, l'environnement des cohortes épidémiologiques, dont le nombre est par ailleurs très élevé, est depuis plusieurs années en forte évolution et comprend une variété d'autres instruments d'observation (bases de données, registres, enquêtes...) qui peuvent aussi contribuer, dans le champ sanitaire et social, aux études et recherches prenant appui sur des données individuelles « en vie réelle ».

L'objet « cohortes » se trouve ainsi à la croisée de trois « mondes », ceux de la recherche, de la santé publique et de la statistique. Ce constat a conduit la mission à considérer les cohortes, soutenues ou non par le PIA, comme l'une des catégories d'instruments mobilisables pour la recherche et l'expertise en santé et à adopter une vision large d'une part des enjeux d'observation existant en ce domaine et d'autre part des articulations à envisager entre les différents types d'outils.

L'analyse des cohortes soutenues par le PIA a constitué, en tant que « cœur de cible » de la mission, le point de départ lui permettant d'identifier ces enjeux et ces articulations. Une journée de travail a

⁵ Le financement du projet I SHARE, portant sur la santé des étudiants, a été interrompu en janvier 2018.

⁶ Le portail Épidémiologie France en recensait 251 en juillet 2019.

ainsi été organisée avec chacune des équipes responsables de ces projets, leurs principaux partenaires (institutionnels, associatifs, privés) et les chercheurs utilisant les données qui en sont issues. La description des objectifs, des caractéristiques scientifiques, des principaux résultats, des modalités de pilotage et du « modèle économique » de chacune des cohortes figure, accompagnée d'une appréciation de la mission sur leurs perspectives futures, en annexe à ce rapport (*cf.* annexe 4).

La mission a également examiné les éléments d'information disponibles sur un ensemble d'autres instruments d'étude et de recherche à portée longitudinale, ainsi que l'évolution en cours des bases de données médico-administratives (Système national des données de santé - SNDS - et *Health data Hub*- HDH- nouvellement créé).

Elle a conduit une large palette d'entretiens, tant avec les responsables des ministères chargés de la santé et de l'enseignement supérieur et de la recherche, qu'avec l'Agence nationale de la recherche (ANR), les organismes du système statistique public (Insee, Drees, Dares), les universités ou organismes porteurs de ces dispositifs (Inserm, Ined, ANRS) et les unités de recherche concernées. Elle s'est entretenue avec la Caisse nationale d'assurance maladie (Cnam), la CNSA et les agences sanitaires, qui sont pour certaines cohortes des partenaires institutionnels majeurs, tout en développant leurs initiatives propres. La mission a aussi dialogué avec des organismes ayant des missions de synthèse ou de coordination dans le domaine de la santé publique (Haute autorité de santé - HAS, Haut conseil de la santé publique - HCSP, Institut de recherche en santé publique - Iresp, Institut national des données de santé -INDS), ainsi qu'avec des représentants de fondations, associations ou entreprises intéressées par le développement des études longitudinales en santé (Ligue contre le cancer, Les Entreprises du Médicament - LEEM, Alliance pour la recherche et l'innovation des industries de santé - Ariis, laboratoires pharmaceutiques...). Elle a enfin consulté une série de personnalités qualifiées dans le domaine de l'épidémiologie ou de la recherche en santé publique et les promoteurs d'autres instruments aux caractéristiques voisines.

Ces entretiens lui ont permis de recueillir une diversité de points de vue sur le bilan, les apports et l'utilisation des différentes cohortes, sur leurs perspectives de développement et les besoins ressentis en la matière, ainsi que des appréciations sur les modes de pilotage et de coordination souhaitables pour ce type d'outils.

La mission a par ailleurs souhaité éclairer la situation française en matière de cohortes par une vision internationale. Elle a, dans ce but, effectué un déplacement au Royaume-Uni, pays qui a une expérience longue et une ambition affirmées en matière de « très grandes cohortes » épidémiologiques, incluant des effectifs massifs et ayant une dimension à la fois sanitaire et sociale.

Le présent rapport, qui s'appuie sur huit annexes méthodologiques ou descriptives, présente les résultats de ces travaux en partant d'un diagnostic de la situation actuelle (partie 1), puis d'une analyse des perspectives et incertitudes liées aux évolutions en cours (partie 2). Il envisage ensuite les actions de consolidation et de clarification nécessaires à court-moyen terme pour tirer le meilleur parti possible des cohortes déjà financées (partie 3). Il tente enfin de tracer, à plus long terme, les contours d'une politique plus ambitieuse et plus coordonnée de développement et de mobilisation des données longitudinales pour la recherche et les études en santé, en élargissant la réflexion à certains domaines sociaux ou médico-sociaux (partie 4).

1 Des cohortes du PIA à l'utilisation des données longitudinales pour la recherche en santé : une impulsion appréciable, des acquis en devenir et de réelles fragilités

La réflexion sur la politique à mener en matière de cohortes épidémiologiques implique de partir d'un diagnostic, tant sur les spécificités de l'instrument que constituent les cohortes (point 1.1), que sur le sous-ensemble constitué par les cohortes du PIA qui, réunissant des dispositifs à caractère divers, fait néanmoins apparaître des éléments de bilan appréciables (point 1.2). Leurs acquis sont toutefois largement encore en devenir, du fait d'une maturité inachevée, et leurs utilisations par les acteurs du système de santé demeurent inégales (point 1.3). Ces cohortes présentent, en outre, de réelles fragilités, qui soit leur sont spécifiques, soit tiennent à des déficits globaux d'accompagnement (point 1.4).

1.1 Un instrument structurant dont les spécificités s'estompent face des dispositifs voisins ou concurrents

1.1.1 Les cohortes : un instrument structurant pour l'expertise et la recherche épidémiologiques

Les cohortes sont, avec les études de cas-témoins, un instrument structurant pour 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* ».

Les cohortes peuvent être définies comme « *le suivi longitudinal, à l'échelle individuelle, d'un groupe de sujets selon un protocole de recherche préétabli* » (cf. annexe 1). Leur principe est donc 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* »⁷.

Le portail Épidémiologie France recensait, en juillet 2019, 251 études de cohorte de source française répondant à cette définition. Leurs objectifs peuvent être « *de décrire les circonstances, la survenue et l'évolution des maladies* », « *d'argumenter l'éventuelle causalité des liens mesurés* »⁸ entre des déterminants et une problématique de santé, ou 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 privilégiant la recherche d'une « *validité interne* »⁹ 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 de collecter, généralement sur la base du volontariat, une diversité de données détaillées à différents moments du temps (examens cliniques, réponses à des questionnaires ou auto-

⁷ 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).

⁸ Source : Institut de recherche en santé publique (Iresp).

⁹ C'est-à-dire la fiabilité des résultats sur la population étudiée.

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

Les problèmes de méthode auxquels sont confrontées les cohortes épidémiologiques sont traditionnellement de plusieurs types : tailles d'échantillons devant permettre des estimations avec la précision souhaitée, biais de sélection liés aux modes de recrutement des participants et à leur volontariat, attrition au cours du temps en raison des « perdus de vue », arbitrage entre qualité, variété et champ des informations recueillies (*cf.* annexe 1).

Eu égard à ces questions de méthode, les cohortes peuvent avoir des pratiques très différentes, évoquées dans l'annexe 1. 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 (*cf. infra* point 1.2.2), mais d'autres choix méthodologiques contribuent aussi à les différencier.

1.1.2 Des caractéristiques qui tendent à perdre de leurs spécificités

La revue effectuée par la mission, qui s'est intéressée à un ensemble de cohortes relevant ou non du PIA, a montré que certains de leurs objectifs premiers, centrés sur la mesure d'associations causales entre facteurs de risque et maladies, à partir de questions de recherche identifiées, se sont diversifiés et ont perdu de leurs spécificités (*cf.* annexe 1). La mission a ainsi noté que :

- la demande des pouvoirs publics d'instruments contribuant à la surveillance épidémiologique et aux politiques de santé publique a conduit certaines cohortes, notamment en population, à rechercher davantage la représentativité statistique, au besoin par des réouvertures périodiques (cas des cohortes COSET de Santé publique France) ;
- d'autres cohortes, notamment de patients, prennent appui sur des centres de diagnostic ou de soins, dont il s'agit avant tout d'harmoniser et d'améliorer les pratiques ;
- les données recueillies pour le suivi à long terme de l'état de santé des personnes incluses sont utilisées de façon différente au fur et à mesure de l'avancée en âge des participants, avec des problématiques autres de celles qui avaient été pensées au départ (cas de la cohorte E3N) ;
- le développement des maladies chroniques et le vieillissement de la population impliquent des approches multifactorielles qui peuvent rendre plus difficile l'établissement de causalités et conduisent à des modélisations sophistiquées, parfois critiquées pour leur côté « boîte noire »¹⁰;
- un nouvel objectif mis en avant est celui du développement d'une médecine « prédictive », visant à repérer des individus « à risque », principalement en vue de politiques de prévention, mais aussi avec des implications socio-économiques non encore complètement maîtrisées.

Cette diversité d'objectifs et d'utilisations potentielles conduit à poser la question de la « nature scientifique » des cohortes, qui peuvent être considérées soit comme un instrument particulier associé à des projets de recherche, soit comme des « infrastructures » à vocation plus générale pour le monde de la recherche, posant donc la question des critères d'évaluation de ces instruments.

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

1.1.3 Des dispositifs de suivi longitudinal proches ou concurrents

À côté des « cohortes », se sont par ailleurs développés d'autres instruments d'observation et de recherche à dimension longitudinale, avec lesquels la distinction et l'articulation ne sont pas toujours aisées. L'instrument « cohortes » s'inscrit désormais dans un paysage multiple de données individuelles potentiellement utilisables pour la recherche et les études en santé. Ces divers instruments ont en principe des caractéristiques méthodologiques différentes, mais ils peuvent, pour partie, questionner l'intérêt de certaines cohortes, ainsi que les articulations à envisager entre les différentes sources.

Ces sources de données potentiellement « concurrentes » ou « articulables » avec les cohortes sont notamment :

- les bases de données administratives ou médico-administratives (BDMA), qui peuvent avoir des origines diverses (systèmes d'information des hôpitaux et des organismes de protection sociale, état civil...).

La principale est aujourd'hui le système national des données de santé (SNDS) qui regroupe, après anonymisation et 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). Les pouvoirs publics ont décidé de l'enrichir en constituant une plateforme de données de santé (*Health data hub*-HDH) intégrant un ensemble de plus en plus large d'informations d'origine diverse (recueils des MDPH, dossiers médicaux, résultats d'analyses biologiques et d'imagerie...), et suscitant des questions sur son articulation, présente et future, avec les cohortes épidémiologiques.

Les recoupements et interactions potentiels sont à cet égard multiples, évolutifs et propres à chaque thématique traitée (cf. annexe 1). Utiliser pleinement les potentialités des BDMA implique en tout cas d'identifier de façon pertinente et robuste les pathologies dans ces bases, avec une qualité d'information médicale suivie et attestée. Il faut aussi être conscient des limites explicatives de ce type de données massives qui ne permettent souvent pas de répondre, en l'absence de recueil d'informations pensé à cet effet, aux recherches de causalités ou d'interactions complexes.

- les registres de morbidité ou de patients, qui recueillent aussi, avec une visée exhaustive, des données individuelles intéressant un ou plusieurs événements de santé. Il peut s'agir de registres populationnels (par pathologie et/ou par zone géographique) ou de registres constitués à partir d'une patientèle, dans certains établissements hospitaliers ou centres de soins. S'agissant des registres de morbidité, 61 étaient répertoriés en France en 2017, dont environ la moitié sont des registres de cancers, dont douze spécialisés sur un type de cancer particulier et 17 registres généraux.

Ces registres 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. Cette situation 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 des registres, à l'horizon de la fin 2020.

La distinction et l'articulation entre registres et cohortes suscite également des questions particulières, dans la mesure où nombre de registres ont eux-mêmes une dimension longitudinale et des recoupements divers avec des cohortes constituées à proximité d'eux ou en leur sein (cf. annexe 1).

- les enquêtes et panels statistiques mis en œuvre sous l'égide du système statistique public (Insee, Drees, Dares, Depp), des agences sanitaires, ou d'organismes ou équipes de recherche.

Ces enquêtes se différencient par une recherche systématique de représentativité statistique des populations d'intérêt, afin d'estimer des indicateurs statistiques en population et d'analyser divers aspects de l'état de santé des individus en fonction de leurs caractéristiques et comportements. Elles peuvent être, soit ponctuelles (comme l'enquête Santé et itinéraires professionnels), soit répétées à un rythme régulier, comme les enquêtes européennes sur la santé (EHIS), l'enquête Conditions de travail ou les enquêtes sur la nutrition de SpF 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 leur est de plus en plus souvent donnée, soit par l'introduction de questions rétrospectives, soit par l'appariement avec les bases de données médico-administratives¹¹, soit par leur réalisation, totale ou partielle, sous forme de « panels », c'est-à-dire d'échantillons aléatoires suivis de façon longitudinale et réalimentés de façon à conserver leur représentativité¹².

Les apports spécifiques des cohortes et des autres types d'instruments incorporant une dimension longitudinale méritent donc d'être appréciés au cas par cas pour chaque thématique d'intérêt, dans le cadre d'une confrontation destinée à la fois à examiner leur complémentarité et à envisager leurs articulations (*cf.* partie 4).

1.2 Les cohortes du PIA : un ensemble hétérogène, qui aurait justifié des approches différenciées selon les types d'instruments, mais qui présente des éléments de bilan appréciables

Le déploiement du PIA1, et notamment de l'action Santé Biotechnologies, s'est fait de manière très rapide, puisque l'ensemble des projets de cohortes a été sélectionné en moins d'un an, de janvier 2011 (résultats de l'appel à projets Cohortes) à décembre 2011 (résultats de l'appel à projets Equipex deuxième vague) (point 1.2.1). Ce rythme rapide, le recours à une démarche « par le bas » (*bottom-up*) et un calibrage incertain des financements initiaux expliquent le caractère disparate des projets finalement retenus et financés, qui auraient justifié des approches différenciées selon les types de d'instruments (point 1.2.2), mais dont le bilan comporte néanmoins des acquis appréciables (point 1.2.3).

1.2.1 Des dispositifs à caractère divers, sélectionnés à travers trois appels à projets du PIA et mobilisant des financements significatifs

Considérées comme « *un des instruments majeurs de la recherche médicale en santé publique et en sciences humaines et sociales* »¹³, les cohortes ont fait l'objet du deuxième appel à projet du PIA1 lancé le 28 juin 2010 dans le cadre du programme des investissements d'avenir. Il s'agissait de financer

¹¹ C'est le cas des enquêtes EHIS, dont l'appariement prévu avec le SNDS permettra une vision de cinq ans en rétrospectif et cinq ans en prospectif, et sera complété par un appariement avec les données socio-fiscales.

¹² C'est le cas par exemple de l'enquête *Share* qui recueille tous les deux ans des données sur la santé et les conditions de vie d'un panel de personnes de plus de 50 ans, ou de l'enquête Conditions de travail de la Dares, qui est réalisée tous les trois ans auprès de 35 000 actifs.

¹³ Texte de l'appel à projets, disponible sur le site de l'ANR, <https://anr.fr/fileadmin/aap/2010/ANR-AAP-COHORTES-2010.pdf>

des cohortes en population générale ou de malades, existantes (dans le but de les pérenniser et d'élargir leur champ d'investigation) ou nouvelles (dans des champs encore peu étudiés), pour une période minimale de six années. À cette fin, était prévu un financement initial reposant sur une dotation non consommable de 200 M€, permettant le versement d'une aide annuelle aux projets retenus pour une période ne pouvant excéder dix années, auquel s'est ajoutée une dotation consommable de 10 M€.

À l'issue de l'appel à projets, le jury international a proposé de financer dix cohortes (*cf.* annexe 2) pour un montant total de 66 M€¹⁴, équivalent aux intérêts versés à partir de la dotation non consommable et à la dotation consommable. À l'exception d'E4N, qui prolonge une cohorte, appelée E3N, lancée au début des années 1990, il s'agit de nouveaux projets de cohortes de patients, sélectionnées à partir de propositions émanant directement d'équipes de recherche, s'appuyant sur un organisme coordinateur.

Les cohortes en population générale n'ont toutefois pas été retenues dans le cadre de cet appel à projets pour des raisons de dépassement de l'enveloppe financière¹⁵. D'où la mobilisation de financements nouveaux dans le cadre, d'une part, de l'appel à projets Infrastructures nationales en biologie santé (INBS)¹⁶ pour financer CONSTANCES (35 M€) et, d'autre part, de la deuxième vague de l'appel à projet Equipex¹⁷, pour financer les deux cohortes d'enfants de RECONAI (13 M€) et la cohorte de patients atteints par une hépatite B ou C portée par l'ANRS, HEPATHER (10M€)¹⁸.

Cet ensemble de projets a donné lieu en 2017 à une évaluation à mi-parcours par un jury international qui, à la suite d'un dialogue entre les responsables de projets et le comité de pilotage a conduit à un certain nombre d'aménagements ou réallocations de crédits. Ont ainsi été redéployés les crédits libérés, au sein de l'action Cohortes, par l'arrêt du financement de I-SHARE au début 2018, ou, plus récemment, par transfert de crédits en provenance de l'action Instituts hospitalo-universitaires (suite à l'arrêt du projet ICAN). Par ailleurs, les modalités d'utilisation des crédits ont été assouplies pour l'année de transition 2019, le règlement financier permettant désormais de couvrir les dépenses liées à des engagements antérieurs.

Au total, et compte tenu de l'arrêt début 2018 du projet I-SHARE, ce sont 12 projets, portés par des équipes principalement franciliennes, qui ont été financés par l'entremise du PIA à hauteur de 124 M€, sur la période précédant l'évaluation finale de 2019 (tableau1).

¹⁴ Non compris le financement de I-SHARE.

¹⁵ Dans son discours de présentation des résultats, le 25 janvier 2011, la ministre Valérie Pécresse indique qu' « il est maintenant de notre devoir de donner une suite, autant que faire se peut, aux projets de qualité qui n'ont pas été retenus. Car il n'est pas question, d'aucune manière que ce soit de décourager les ambitions qui sont nées ici ou là ces derniers mois. Je pense en particulier aux projets de renforcement de cohortes existantes qui continueront à trouver dans ce ministère un soutien et un appui précieux dans les mois qui viennent, que ce soit au travers des financements courants dont ils bénéficient déjà, ou dans le cadre d'une articulation avec les autres programmes du plan d'investissements d'avenir. »

¹⁶ Clos en octobre 2010. Les cohortes ne relèvent pas explicitement du champ de l'appel à projets. Cf <https://anr.fr/fileadmin/aap/2010/ANR-AAP-INFRASTRUCTURES-2010.pdf>

¹⁷ Clos en septembre 2011.

¹⁸ La sélection d'HEPATHER est annoncée dans le cadre des réponses de la deuxième vague de l'appel à projet Laboratoire d'excellence de février 2012 au titre d'équipement complémentaire aux résultats de l'appel à projet Equipex annoncés en décembre 2011.

Tableau 1 : Les cohortes financées au travers du PIA

Nom de la cohorte	Objet	Etablissement coordinateur	Effectifs suivis	Financements PIA attribués 2011-2018 (K€)
CANTO	Apprécier la toxicité des traitements chez les patientes porteuses d'un cancer du sein localisé	Fédération Unicancer	12 000	13 100
CKD REIN	Appréhender les facteurs d'évolution des maladies rénales chroniques	Université Paris Sud	3 000	5 861
COBLANCE	Catégoriser et identifier les causes d'évolution du cancer de la vessie	Inserm	1 120	3 945
CRYOSTEM	Comprendre et prévenir les complications de la greffe de moelle	Société française de greffe de moelle et de thérapie cellulaire, puis association CRYOSTEM	5 000	5 238
E3N/E4N	Identifier les facteurs favorisant les principales maladies chroniques en suivant des familles sur trois générations	Université Paris Sud	80 000 femmes depuis 1990 18 000 conjoints 10 000 enfants	7 507
HOPE EPI	Mutualiser la collecte d'informations sur les cancers de l'enfant et progresser sur l'identification des risques environnementaux et génétiques	Université Paris 5	25 000	6 274
HEPATHER	Évaluer les prises en charge des patients infectés par le virus des hépatites B et C	ANRS	21 000	10 010
OFSEP	Construire un outil épidémiologique de suivi de la sclérose en plaques	Université Claude Bernard Lyon 1	62 000	9 768
PSY-COH	Suivre les patients atteints de schizophrénie et de troubles bipolaires	Fondation Fondamental	800 (troubles bipolaires) et 400 (schizophrénie)	3 846
RADICO	Construire une plateforme mutualisée permettant de développer des cohortes de patients atteints de maladies rares	Inserm	13 000 (pour les 13 cohortes)	10 513
RECONAI	Constituer une plateforme de suivi sur longue périodes de deux cohortes d'enfants à la naissance, dont l'une de prématurés	Ined	18 000 (cohorte ELFE) et 8 000 enfants nés prématurés (cohorte EPIPAGE2)	13 018
CONSTANCES	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,...)	UVSQ	200 000	35 000

Source : ANR, résultats des appels à projets du PIA.

Les tableaux de suivi de l'ANR montrent que le niveau de consommation des crédits constaté à la fin 2018 s'établit en moyenne à 69%, avec une accélération marquée en fin de période et des écarts importants (allant de 55,5 à 87% selon les projets) reflétant pour une large part leurs délais de mise en place et de construction (cf. annexe 2 et point 1.3.1.).

Une analyse de ces financements, un panorama plus détaillé des cohortes et une fiche de présentation de chacun des projets font l'objet des annexes 2, 3 et 4 de ce rapport.

La relative souplesse dans l'usage des crédits du PIA a également été utilisée pour programmer le financement, toujours dans le cadre de l'action Santé- biotechnologies, d'une nouvelle cohorte Autisme, décidée sur la base de l'article 92 de la stratégie nationale pour l'autisme d'avril 2018¹⁹, et dont la définition et le lancement sont prévus courant 2020 suite à un appel à manifestation d'intérêt émis en septembre 2019 (encadré 1).

Encadré 1

L'appel à manifestation d'intérêt sur une cohorte Autisme

Décidé dans le cadre de la stratégie nationale pour l'autisme au sein des troubles du neuro-développement présenté en avril 2018, le choix de financer une cohorte vise à permettre de mieux comprendre l'hétérogénéité des troubles du spectre autistique en s'appuyant sur «*une grande base de données multidimensionnelles recueillies prospectivement sur une longue durée et portant sur les caractéristiques comportementales, cliniques, cognitives et génétiques d'enfants ayant un diagnostic formel de TSA, ainsi que sur leurs expositions précoces au cours du développement* »²⁰.

À la différence des appels à projets précédents du PIA, la démarche s'effectue en trois temps visant à organiser une interaction entre les demandes des pouvoirs publics et les milieux de recherche : demande d'une note de cadrage à deux chercheuses, appel à manifestation d'intérêt permettant de faire état des projets susceptibles d'émerger, puis appel à projet proprement dit, doté de 6 M€, pour une cohorte prévue sur période initiale de cinq ans pouvant se prolonger jusqu'à dix ans.

Source : *Appel à manifestation d'intérêt de juillet 2019.*

En revanche, il n'a pas été fait appel au PIA pour le financement de la cohorte MEMENTO, mise en place dans le cadre du plan Alzheimer et financée par la fondation éponyme (à hauteur de 15,6 M€), puis sur crédits budgétaires du ministère chargé de la recherche²¹ (1,73 M€, auxquels s'ajoutent près de 1,3 M€ des provenant d'autres partenaires publics²²) (encadré 2).

¹⁹ https://handicap.gouv.fr/IMG/pdf/strategie_nationale_autisme_2018.pdf

²⁰ AMI du 26 juillet 2019, disponible sur le site de l'ANR, <https://anr.fr/fileadmin/aap/2019/ami-ia-cohorte-autisme-2019-v2.pdf>

²¹ Il a toutefois été décidé que MEMENTO serait évaluée selon le même dispositif que les cohortes du PIA.

²² Les « autres financements publics » correspondent 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. cf. annexe 2.

Encadré 2
La cohorte MEMENTO

MEMENTO est une cohorte de 2300 personnes, non malades au début de l'étude, mais présentant des troubles de la mémoire à un stade précoce et éventuellement susceptibles de développer ensuite une maladie d'Alzheimer ou une démence connexe. Décidée dans le cadre du plan Alzheimer 2008-2012, la cohorte entend « *mesurer, de façon répétée chez des patients avec des signes précoces, une série de biomarqueurs (imagerie par résonance magnétique (IRM) du cerveau, analyses génétiques, imagerie moléculaire en TEP, liquide céphalo-rachidien extrait par la ponction lombaire) et de facteurs de risque (habitudes de vie, traitements médicamenteux, qualité de vie) et de suivre parallèlement l'évolution de leurs performances neurocognitives* ». ²³ Les patients ont été recrutés entre 2011 et 2014 dans les centres mémoire de ressources et de recherche et font l'objet d'un suivi tous les six mois.

Source : Dossier d'évaluation 2019.

Les financements programmés par le PIA ayant un terme prévu en 2020, une nouvelle évaluation des projets de l'action Santé-biotechnologies s'est tenue en septembre 2019, conduisant, sur la base de l'évaluation faite par le jury international, à de nouvelles propositions de décisions d'allocation formulées par le comité de pilotage. Dans ce cadre, les cohortes (qu'elles soient financées sur les actions Infrastructures, Equipex ou Cohortes) bénéficieront d'un financement de 26,2 M€ pour la période 2020-2024. Le comité de pilotage a en outre transmis aux services du Premier ministre une proposition de ventilation, en cours de finalisation.

1.2.2 Une typologie nécessaire pour asseoir la pertinence des approches et propositions à envisager

Un premier examen des quatorze cohortes soutenues par le PIA, dont le panorama figure sous forme de tableau synoptique en annexe 3, montre que, au-delà même des écarts extrêmes entre le nombre des participants qu'elles suivent (compris entre 1 300 et 200 000), elles ont des objectifs et des caractéristiques très différents : cohortes pluridimensionnelles en population *versus* cohortes spécialisées de personnes atteintes et/ou traitées au titre d'une ou plusieurs pathologies ; cohortes « d'amont » cherchant à évaluer les effets des différents facteurs ou expositions ou cohortes « d'aval » centrées sur l'impact de thérapeutiques ou de politiques publiques ; cohortes fermées, encore ouvertes ou aspirant à l'être à nouveau ; cohortes constituant des projets spécifiques lancés à partir de 2010-2011 ou consolidées dans le cadre de plateformes de recherche favorisant leur utilisation partagée.

Or, il est apparu à la mission que, selon les cas de figure, les questions relatives à la pérennisation, à l'accessibilité, l'ouverture et l'interopérabilité, ainsi qu'à la gouvernance et au financement de ces différents projets n'ont pas vocation à trouver les mêmes réponses : une cohorte suivant de façon fermée des patients diagnostiqués ou traités à un moment donné, et proche par-là de préoccupations de recherche clinique, n'a pas les mêmes enjeux de réalimentation et de pérennisation dans la durée, les mêmes partenaires institutionnels et financiers ni la même inscription dans un système d'informations global qu'une cohorte généraliste, visant à appréhender, en population générale ou auprès d'une large fraction de la population, une multiplicité de facteurs de risque.

²³ Site internet de MEMENTO, http://www.memento-cohort.org/memento_web/L%C3%A9tude/Lesobjectifs.aspx

À cet égard, la mission considère que, à la différence des appels à projets du PIA ou des premiers contours du projet France Cohortes (*cf. infra* point 2.5.4), les réflexions et actions conduites devraient partir d'une typologie distinguant au minimum trois grandes catégories de dispositifs :

- les cohortes en population, dont la plus importante est en France l'étude CONSTANCES : cohorte généraliste suivant 200 000 sujets âgés de 18 à 69 ans, elle se veut une infrastructure de recherche réunissant des informations sur de multiples aspects de leur état et de leurs problèmes de santé, ainsi que sur leur situation sociale et professionnelle, leurs expositions environnementales et leurs modes de vie. (*Cf.* annexe 4) ; la cohorte ELFE, qui suit 18 000 enfants depuis la naissance pour étudier de façon globale et multidisciplinaire les questions de santé, de développement et de socialisation chez l'enfant, en est un second exemple, tandis que la cohorte E3N/E4N suit quant à elle, de façon spécifique, environ 80 000 femmes qui étaient adhérentes à la MGEN en 1990 ainsi que leur famille (conjointes et descendants) ;
- des cohortes ciblées sur des patients, soit diagnostiqués et suivis pour certaines pathologies, soit sujets à des prises en charge ou des traitements spécifiques : c'est par exemple le cas des cohortes PSY-COH (troubles bipolaires et schizophrénie), CANTO (cancer du sein localisé), CKD Rein (insuffisance rénale chronique non terminale), COBLANCE (cancer de la vessie) ou MEMENTO (troubles cognitifs identifiés à un stade précoce) ;
- en recoupement partiel avec la catégorie précédente, des « plateformes » organisant des suivis longitudinaux à partir de sources de données (registres de patients) et/ou de méthodes de partagées entre scientifiques et cliniciens : c'est le cas du projet RADICO qui organise en amont et fédère treize cohortes de patients atteints de maladies rares ou, avec une autre optique, des projets OFSEP et CRYOSTEM, qui combinent fonctionnement harmonisé de registres de patients dans des centres de traitement spécialisés et suivi longitudinal enrichi pour une fraction d'entre eux.

Cette typologie, qui ressort de l'examen des quatorze projets soutenus dans le cadre du PIA, paraît pouvoir s'appliquer à la plupart des grandes cohortes financées sur fonds publics. L'étude NUTRINET, qui suit 160 000 participants âgés de plus de 18 ans sur les questions de nutrition, ressortit par exemple de la première catégorie, tandis que les cohortes de patients vivant avec le VIH ou une hépatite virale promues par l'ANRS se réfèrent clairement à la deuxième, et que le programme ESMÉ mis en œuvre par Unicancer pour suivre les prises en charge des cancers du sein métastatiques, des ovaires et du poumon s'apparente à la troisième (*cf.* annexe 5).

Il serait donc logique que la réflexion conduite sur les cohortes du PIA s'élargisse, à partir d'une catégorisation du même type, à l'ensemble des grandes cohortes ou dispositifs de suivi soutenus par les pouvoirs publics.

Recommandation n°1 Resituer les cohortes dans un cadre d'analyse adapté

- Distinguer, dans les politiques mises en œuvre, les grandes cohortes en population, les études longitudinales de patients concernés par certaines pathologies ou traitements, et les « plateformes » organisant des suivis longitudinaux à partir de méthodes et/ou de sources partagées.
- Englober, dans la définition de ces politiques, les cohortes en population financées par des voies autres que le PIA, et les dispositifs de suivi longitudinal placés sous l'égide des organismes de protection sociale ou des agences sanitaires.

1.2.3 Des éléments de bilan appréciables en termes de potentiel scientifique, de positionnement international et de structuration des milieux de recherche

Si l'on en revient aux seules cohortes soutenues par le PIA, le bilan qui peut en être tiré presque dix ans après le lancement du programme fait apparaître des acquis appréciables.

L'appel à projets Cohortes retient six critères d'évaluation des projets, appelés à être appréciés par des jurys internationaux : la pertinence de la proposition au regard des orientations de l'appel à projets, et en particulier, le caractère innovant et la faisabilité du projet de cohorte ; la qualité scientifique et méthodologique ; la qualité de la gouvernance ; l'impact médico-économique et socioéconomique ; l'adéquation moyens / faisabilité et la capacité à rapprocher des équipes et/ou projets de recherche publique, privée ou industrielle.

Ces jurys se sont réunis à deux reprises, à mi-parcours au début 2017 et à l'été 2019. La mission a pu bénéficier de leurs remarques et conclusions, et elle a elle-même pu avoir, à l'occasion de visites de travail d'une journée, des discussions avec les responsables des cohortes sur l'avancement et les apports scientifiques de leur projet ; elle en a tiré des appréciations d'ensemble qui figurent, pour chacune des quatorze cohortes analysées, dans les différentes fiches de l'annexe 4.

De façon plus transversale, la mission a pu constater que la qualité de la production scientifique, la structuration des communautés de recherche autour d'un champ et d'un instrument fédérateurs, ainsi que le développement des partenariats internationaux constituent, dix ans après, trois acquis appréciables, communs à la plupart des projets.

Les cohortes contribuent ainsi à conforter la place de la France dans la recherche médicale, dans des domaines où sa spécialisation est reconnue²⁴. Plusieurs projets ont développé des partenariats internationaux importants, dans un cadre européen (par exemple OFSEP avec les registres danois, suédois et italiens dans le cadre du consortium *BigMSdata*) ou international (par exemple PSY -COH, avec *Psychiatric Genetics Consortium* PGC et Enigma, ou RADICO dans le cadre des réseaux européens de référence maladies rares – ERN – que le projet a lui-même contribué à structurer). Il en est de même pour MEMENTO (engagé dans les consortia *European Alzheimer DNA Biobank* et *Cohorts for Aging Research in Genomics*), CONSTANCES (le *100K Cohort Consortium*) ou encore COBLANCE, dont le jury souligne le potentiel de collaboration, au-delà des premiers liens tissés avec l'Espagne, les Pays Bas et le Royaume-Uni). D'autres projets permettent à la France d'occuper une place de premier rang : c'est le cas avec CKD REIN, dont la cohorte dépasse par la taille le projet similaire du Japon (plus de 3 000 patients suivis contre environ 2 200), certaines cohortes hébergées par la plateforme RADICO, ou CRYOSTEM, qui offre l'une des plus grandes bases de données mondiales sur les transplantations et CANTO, qui est la seule cohorte de ce type au plan international.

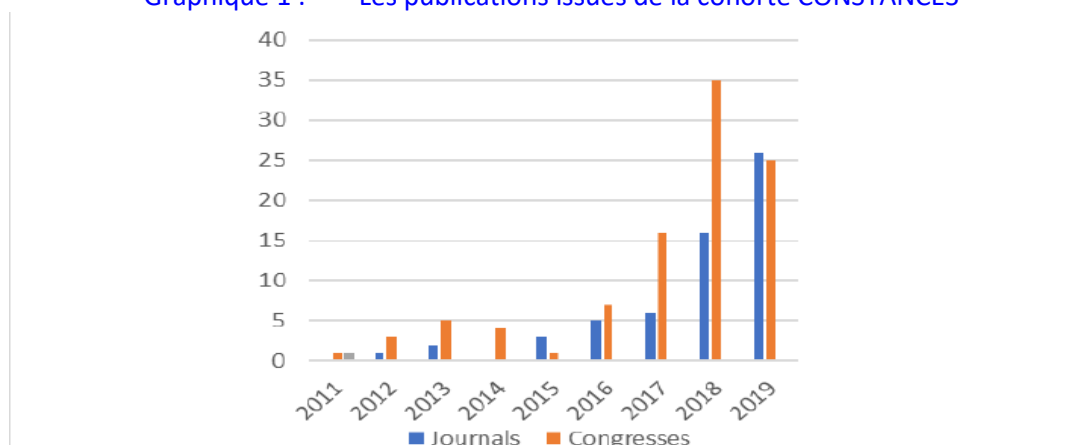
Par ailleurs, les cohortes contribuent à la structuration de communautés de recherche : l'ensemble des projets revendique une politique d'ouverture des données aux chercheurs (publics ou privés), qui se traduit par un nombre important de projets de recherche utilisant les données collectées. Pour autant, il s'agit souvent de projets déposés par des équipes proches de celle qui porte la cohorte. Ainsi, sur 108 projets de recherche déposés à l'ANR (recensement effectué par le département Biologie-santé) entre 2016 et 2018, seuls neuf projets utilisent des données collectées par des cohortes du PIA, soit moins de 10%. Ce volume relativement faible peut refléter une visibilité

²⁴ L'indice de spécialisation, calculé à partir de la part des publications, place la France juste au-dessus de la moyenne mondiale. cf. HCERES-OST, *La position scientifique de la France dans le monde 2000-2015*, mars 2018. https://www.hceres.fr/sites/default/files/media/downloads/Hc%C3%A9res_OST_Position_Scientifique_France.pdf

insuffisante des cohortes dans le monde de la recherche. L'effet structurant revêt par ailleurs plusieurs formes, qui peuvent aller de l'animation d'un réseau de centres de soins autour d'une démarche qualité (CRYOSTEM, OFSEP, PSY-COH) à l'animation d'une communauté de chercheurs jusqu'alors dispersés (CANTO, qui relie environ 120 professionnels, RADICO dans le cadre des filières de santé maladies rares).

Les cohortes présentent enfin, quelle que soit leur nature, un potentiel scientifique important²⁵: ainsi, RE-CO-NAI a permis la production de 112 articles, dont 91 en anglais et 39 dans des revues de rang A ; neuf articles dans des revues de haut niveau sont en cours dans le cadre du projet CANTO ; bénéficiant d'une grande ancienneté, E3N / E4N affiche plus de 500 publications entre 2011 et 2019. Toutefois, la plupart des cohortes n'ont pas encore débouché sur une production scientifique significative. En effet, l'appréciation finale de l'impact scientifique des cohortes semble prématurée, compte tenu du temps nécessaire pour tirer pleinement partie des données collectées, ainsi que le montre l'exemple de CONSTANCES, dont les publications se sont accélérées ces deux dernières années (cf. graphique 1 et annexe 4).

Graphique 1 : Les publications issues de la cohorte CONSTANCES



Source : Dossier d'évaluation 2019.

Il est donc préférable de s'en tenir, ainsi que l'a fait le jury international dans son évaluation finale, au constat d'un potentiel important de publications de haut niveau, qui plaide pour une inscription des cohortes dans la durée.

1.3 Une maturité encore incomplète et une utilisation très variable par les acteurs du système de santé

Compte tenu du temps de mise en place et de l'incomplétude de la plupart des projets, les acquis potentiels des cohortes financées par le PIA sont cependant pour une large part encore en devenir, montrant qu'elles doivent encore gagner en maturité (point 1.3.1). Quant à la notoriété et l'utilisation de ces cohortes en dehors du seul monde de la recherche, notamment par les directions des ministères sociaux et par les agences sanitaires, elles sont pour le moins variables, l'intérêt porté aux cohortes du PIA apparaissant loin d'être général et manifeste (point 1.3.2).

²⁵ La recherche médicale est en France la première discipline pour le nombre de publications, avec un total de 13 987 en 2016 (OST, *Dynamics of scientific production in the world, in Europe and in France, 2000-2016*, Paris 2019).

1.3.1 Des délais de construction et des horizons temporels longs impliquant une maturité et des acquis encore en devenir

Le besoin d'inscription des cohortes dans un temps long (à l'instar des exemples anglo-saxons – cf. annexe 7) s'explique par des raisons qui tiennent à la fois à la nature de l'instrument cohorte, au temps nécessaire pour compléter l'infrastructure et à des raisons d'ordre contextuel.

La nature de l'instrument cohorte génère en premier lieu des rendements croissants : parce qu'elles offrent « la possibilité d'analyses épidémiologiques longitudinales permettant de tenir compte au mieux de phénomènes liés au temps, notamment de la séquence temporelle exposition (ou intervention)- effet »²⁶, les cohortes requièrent une période longue de constitution et d'exploitation des données qu'elles collectent. Le caractère « exceptionnel » des résultats du projet E3N/E4N, souligné par le jury international, tient largement à l'accumulation continue d'informations depuis 1990.

Plusieurs cohortes sont, en outre, encore en phase de construction ou sont des infrastructures incomplètes : certains projets ont ainsi démarré l'inclusion tardivement (cas des cohortes PSY-COH, COBLANCE et des cohortes hébergées par la plateforme RADICO), ont différé l'appariement de leurs données avec celles du SNDS (réalisé uniquement pour CONSTANCES et en test pour RE-CO-NAI), ou ont reporté dans le temps si ce n'est la constitution de biobanques, du moins leur localisation définitive et/ou la réalisation des analyses génomiques et autres « -omics » (métabolomique, protéomique et microbiomique²⁷) permettant d'utiliser toutes leurs potentialités.

Enfin, ainsi que l'ont exposé les porteurs de projet à la mission, un temps important a été consacré aux démarches administratives imposées par exemple par le règlement financier de l'ANR (notamment la signature de l'accord de consortium²⁸ : il faut ainsi attendre mai 2015 pour la signature de celui de CONSTANCES, immédiatement rendu obsolète avec la perspective d'entrée de l'Université Paris Descartes dans l'accord, nécessitant un nouvel avenant, non encore signé, et la fin 2016 pour l'accord de consortium d'HOPE-EPI) ; ces accords peuvent être un élément utile pour préciser les responsabilités, notamment financières, des partenaires institutionnels, mais la longueur et les péripéties de leur négociation ont pu constituer un frein pour l'avancement des projets. Les contraintes imposées par le code des marchés publics (pour la gestion des questionnaires et des relances ou pour le stockage des données et prélèvements) ont également pu être lourdes et chronophages.

1.3.2 Une visibilité et des utilisations très variables par les acteurs du système de santé

Un autre élément de bilan s'agissant des cohortes du PIA est l'intérêt que leur portent les principaux acteurs du système de santé, au-delà des seuls milieux de recherche. Or, d'après les éléments recueillis par la mission lors des nombreux entretiens qu'elle a réalisés avec ces acteurs (cf. liste des

²⁶ Marcel Golberg et Marie Zins, *Principe et intérêt des cohortes épidémiologiques*, ADSP n° 78, mars 2012.

²⁷ La protéomique étudie la composition des protéines dans un tissu, la métabolomique les produits métaboliques (acides aminés, oses, lipides, acides nucléiques, protéines, vitamines, hormones) qui caractérisent l'adaptation du métabolisme aux modifications d'environnement et la microbiomique les microorganismes qui colonisent le corps humain.

²⁸ L'accord de consortium est un contrat entre partenaires d'un projet de recherche à visée essentiellement financière : prise en charge des coûts du projet, partage des risques et des résultats, répartition des droits de propriétaire ; son utilisation comme outil de gouvernance, en lieu et place des instruments utilisés dans la recherche publique, traditionnellement structurée en unité (de recherche ou de service) peut être discutée.

personnes rencontrées), cet intérêt et ces utilisations sont loin d'être systématiques, et varient fortement selon les types d'instruments et d'interlocuteurs.

La situation est ainsi pour le moins disparate en ce qui concerne les directions d'administration centrale des ministères chargés de la santé et des affaires sociales (encadré 3).

Encadré 3

L'utilisation des cohortes du PIA par les directions des ministères chargés de la santé et des affaires sociales

La Direction générale de la santé (DGS) manifeste surtout de l'intérêt pour les grandes cohortes en population (ELFE, CONSTANCES), qu'elle soutient à travers des conventions prévoyant des éclairages sur des sujets particuliers (hypertension et risques cardio-vasculaires chez les adultes, nutrition et activité physique chez les enfants...).

La Direction générale de l'offre de soins (DGOS) s'intéresse exclusivement aux cohortes qui permettent d'éclairer les logiques d'organisation des soins et l'évaluation des pratiques ; son intérêt se concentre à cet égard sur trois cohortes du PIA : CRYOSTEM (greffe de moelle), OFSEP (sclérose en plaques) et RADICO (maladies rares).

La Direction générale du travail (DGT) focalise son attention sur les cohortes soutenues dans le cadre des plans Santé au travail (PST) : elle suit dans ce cadre une convention avec Santé publique France (SpF) qui appuie les cohortes CONSTANCES et COSET, qui en sont le complément pour les affiliés aux régimes agricole et indépendant (cf. annexe 5).

La Direction de la recherche des études (Drees) utilise quant à elle peu les cohortes épidémiologiques, mis à part de façon ponctuelle. Donnant la priorité aux considérations de représentativité statistique, elle privilégie l'utilisation longitudinale des bases de données administratives (SNDS et données socio-fiscales) appariées aux panels ou enquêtes statistiques transversales qu'elle met en place en tant que de besoin ou pour répondre aux obligations communautaires (*European Health Interview Survey-EHIS* co-pilotée avec l'Irdes).

La Direction de l'animation de la recherche, des études et de la statistique (Dares) a une optique voisine dans le champ de la santé au travail : elle privilégie l'usage de l'enquête Conditions de travail, pour partie panélisée et appariée avec le SNDS, tout en suivant les résultats de CONSTANCES, qu'elle considère comme un dispositif très complémentaire.

Source : *Entretiens réalisés et éléments d'information recueillis par la mission*

Du côté des organismes de protection sociale, l'intérêt manifesté aux cohortes est avéré, mais concentré sur certaines d'entre elles : l'Assurance maladie apporte un soutien majeur à la cohorte CONSTANCES, mais met principalement l'accent sur son apport à la structuration de l'activité des centres d'examen de santé et à l'identification des pathologies dans le SNIIRAM (projet ReDSiam) plutôt que sur ses utilisations directes, dans la mesure où elle a développé, à des fins d'étude interne, ses propres « cohortes » à partir des données du SNDS. La Mutualité sociale agricole (MSA) est, quant à elle, surtout investie dans des cohortes ne relevant pas du PIA (COSET et AGRICAN cf. annexe 5).

Du côté des agences sanitaires, le paysage est aussi fortement contrasté selon les missions et les problématiques propres à chacune d'entre elles (encadré 4).

Encadré 4

L'utilisation des cohortes du PIA par les agences sanitaires

L'agence Santé publique France (SpF) utilise directement les données de cohortes Constances et Elfe pour valider ses estimations de prévalence et d'incidence de certaines pathologies (diabète et bronchopneumopathie chronique obstructive- BPCO), apprécier les liens entre pesticides et problèmes de santé ou évaluer les effets chez les enfants de l'exposition aux polluants pendant la grossesse.

Une autre agence, chargée de la sécurité du médicament et des produits de santé (ANSM), est au contraire relativement circonspecte vis-à-vis des cohortes du PIA dans la mesure où elles peuvent être financées par des partenariats privés, et met aujourd'hui l'accent sur son partenariat avec la Cnam dans le cadre du GIS Epiphare pour réaliser des études en matière de pharmacovigilance.

D'autres agences sont fortement investies dans le soutien aux cohortes entrant dans leur champ d'intérêt, que ce soit sous la forme d'un appui direct à ces infrastructures ou du financement de recherches spécifiques dans le cadre d'appels à projets : c'est le cas de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses), qui est partie prenante, hors PIA, de la cohorte AGRICAN, mais aussi de Cosmos, étude prospective sur le téléphonie mobile et la santé « nichée » au sein de Constances ; c'est aussi le cas de l'Institut national du cancer (Inca), qui a historiquement soutenu les cohortes CONSTANCES, CANTO et HOPE-EPI.

D'autres agences, enfin, de par les missions qu'elles exercent en matière de recherche, assurent directement tout ou partie du fonctionnement de certaines cohortes du PIA, en leur apportant un appui logistique, informatique et/ou administratif : l'agence de la biomédecine (ABM) gère ainsi le fonctionnement informatique de la cohorte CKD REIN, tandis que l'Agence nationale de recherche sur le SIDA et les hépatites virales (ANRS), qui est une agence autonome rattachée à l'Inserm, abrite directement en son sein la cohorte HEPATHER, fleuron de la dizaine de cohortes dont elle porte la gestion et le financement (cf. annexe 5).

Source : Entretiens réalisés et éléments d'information recueillis par la mission

Du côté des Hauts conseils et autorités indépendantes, l'intérêt pour les cohortes est aussi affirmé. Par exemple, le Haut Conseil de la santé publique (HCSP) prend en compte régulièrement leurs résultats dans le cadre de ses travaux, même si, dans ceux qu'il vient de consacrer au dispositif de suivi sur la santé de l'enfant²⁹, il déplore le caractère dispersé, fragmenté et peu coordonné des sources de données qui y contribuent. La Haute autorité de santé (HAS) estime de son côté que les cohortes sont au cœur de l'évaluation des pratiques des professionnels et de l'effet des traitements en vie réelle dont elle a la responsabilité et que leur utilisation peut aussi contribuer à la connaissance *ex post* des conditions d'utilisation des produits de santé mis sur le marché (bon usage, effets indésirables...). Dans ce contexte, elle a tendance à demander aux laboratoires pharmaceutiques de se tourner vers les cohortes académiques françaises comme « sources de référence » pour leurs études ; parmi ces « références », figurent notamment pour elle les cohortes HEPATHER, OFSEP et RADICO, mais aussi d'autres sources comme l'enquête ENTRED sur les patients diabétiques.

Enfin, d'autres acteurs, comme les organismes faisant appel à la générosité publique, suivent de très près certaines cohortes s'intéressant à des pathologies particulières. La Ligue contre le cancer a ainsi été un partenaire majeur des cohortes E3N/E4N et CANTO, même si elle souhaite désormais soutenir des projets de recherche plutôt que contribuer à financer des infrastructures. Les associations de patients sont de leur côté de plus en plus souvent associées au recrutement des participants, puis à la diffusion des résultats, de cohortes du PIA comme PSY-COH ou HEPATHER.

²⁹ Haut Conseil de la santé publique, Le dispositif statistique de surveillance et de recherche sur la santé de l'enfant, octobre 2019

1.4 De réelles fragilités, spécifiques aux projets ou tenant à des déficits généraux d'accompagnement

L'ensemble des projets de cohortes présente un certain nombre de fragilités qui ont pesé sur le développement des projets. Certaines sont spécifiques aux projets eux-mêmes (point 1.4.1), quand d'autres présentent un caractère plus général, voire systémique, et tiennent largement à des déficits d'accompagnement (point 1.4.2).

1.4.1 Des fragilités spécifiques à certains projets, suite à des difficultés scientifiques ou à des problèmes de gouvernance

Le déploiement des projets sélectionnés s'est parfois accompagné de difficultés qui ont conduit, soit à des recentrages ou à des redimensionnements (OFSEP, HOPE-EPI), soit, dans un cas, à l'arrêt définitif du financement (I SHARE). Elles témoignent de fragilités inhérentes aux projets.

Certaines sont d'ordre technique, et renvoient par exemple à des mésententes scientifiques (HOPE-EPI) ou à des problèmes d'appréciation de la puissance statistique de la cohorte : difficultés de recrutement et attrition importante pour PSY-COH, biais de sélection possibles dans CANTO, petite taille de la population suivie pour OFSEP-HD, ...

D'autres fragilités tiennent au relatif isolement des cohortes dans le paysage de la statistique nationale ou dans celui de la recherche publique : ainsi, nombre de projets, et notamment les cohortes en population, gagneraient à utiliser les données fiscales et sociales de l'INSEE plutôt qu'à construire *ex nihilo* des données socio-économiques sans doute moins robustes ; de même, un projet comme CRYOSTEM, dont la production académique demeure encore « modeste », selon le jury, gagnerait à s'inscrire davantage dans le paysage académique.

Enfin, certaines cohortes ont rencontré des difficultés de gouvernance, soit liées aux rapports avec les tutelles (dans le cas de CONSTANCES), soit internes, jusqu'à en paralyser parfois le développement (dans le cas de HOPE-EPI). D'autres ont mis en place des structures de gouvernance soit complexes (E3N/E4N), soit fragiles (CRYOSTEM), soit rarement réunies (MEMENTO), ou encore faisant incomplètement le départ entre les rôles respectifs des responsables du projet, des partenaires institutionnels directement impliqués, et, parfois, des partenaires privés (HEPATHER et CKD REIN).

Ces difficultés ont été parfois surmontées ou sont en voie de l'être, en raison notamment d'un fort niveau d'engagement des responsables de projet, mais d'autres appellent des encore des mesures correctrices (*cf. infra* point 3.1).

1.4.2 Des difficultés plus largement partagées, traduisant des déficits généraux d'accompagnement en termes d'infrastructures techniques, de politique scientifique et de gestion des compétences

D'autres fragilités s'avèrent partagées par l'ensemble des cohortes et présentent ainsi un caractère systémique : elles se concentrent sur des questions d'infrastructures techniques, de politiques scientifiques et de manque de compétences critiques.

1.4.2.1 Des problèmes d'infrastructure technique concernant les biobanques et l'accès aux données médico-administratives

La première fragilité tient à l'absence de vision d'ensemble et au manque de coordination des solutions trouvées par les responsables des projets aux questions d'infrastructures. Celles-ci portent, d'une part, sur le stockage des prélèvements biologiques, et, d'autre part, sur l'appariement avec les données administratives et le stockage sécurisé des corpus de données de santé ainsi constitués.

Concernant les biobanques, les responsables de cohortes ont été confrontés à l'échec du projet Biobanque simultanément prévu par le PIA1, qui visait à « *intégrer à l'échelon national les capacités de recueil et de stockage des échantillons biologiques d'origine humaine et les collections microbiennes, d'assurer la qualité des collections et des annotations cliniques associées, et de faciliter l'accès à ces collections pour les projets de recherche. Il s'agit d'une infrastructure distribuée s'appuyant sur 64 biobanques et 6 collections microbiennes, avec une gouvernance intégrée, une coordination des services et un accès unique à l'échelon national, qui assure un couplage avec les réseaux Européens de biobanques* ». ³⁰.

Ils ont donc été amenés à trouver des solutions *ad hoc*, qui, d'après les observations de la mission sont de quatre types :

- recours par deux cohortes à l'Établissement français du sang (EFS) ;
- stockage dans des centres de ressources biologiques (CRB) publics situés sur le territoire français, et ce sur une base centralisée ou, plus rarement, décentralisée (CRYOSTEM) ;
- stockage hors du territoire français (CONSTANCES dans le cadre d'un marché avec un institut situé au Luxembourg) ;
- stockage par des entreprises privées, sur la base d'une contractualisation dans une perspective de plus ou moins long terme.

Certaines de ces solutions ont d'ailleurs fluctué au cours du temps, un exemple en étant la cohorte HEPATHER dont les prélèvements biologiques ont été initialement stockés par l'Établissement français du sang (EFS), puis par le CRB de Saint Antoine, avant qu'un marché soit passé avec l'entreprise *Cell&Co* située en Auvergne.

Concernant les données collectées par les cohortes et leurs appariements avec les données médico-administratives, les modifications successives de la législation sur les données de santé (création du SNDS en 2016, transposition du RGPD dans la législation française en 2018, constitution du *Health data Hub* en 2019) et la nécessité de trouver, depuis 2019, des lieux de stockage et d'accès respectant le référentiel de sécurité défini par la loi, ont conduit les cohortes à un attentisme en matière de projets d'appariements avec le SNDS et/ou à des solutions plus ou moins durables : l'Inserm a ainsi été amené à passer à l'été 2019 une convention avec le Centre d'accès sécurisé aux données (CASD) pour l'accueil et la mise à disposition des données de la cohorte CONSTANCES, en l'attente du développement de ses propres projets (*cf.* annexe 6).

³⁰ Fiche de présentation du projet Biobanque, doté de 17 M€, à l'issue de sa sélection dans le cadre de l'appel Infrastructures nationales de recherche en santé biologie.

Le tableau 2 ci-dessous montre la diversité des solutions actuellement mises en œuvre par les cohortes dans ce contexte marqué tout à la fois par l'augmentation des exigences de sécurité et par des changements rapides d'environnement³¹.

Tableau 2 : L'état des situations des cohortes du PIA concernant les bases de données et le stockage des prélèvements biologiques

	Appariement SNDS	Stockage sécurisé des données	Stockage des prélèvements biologiques
CANTO	Non	Centre Georges François Leclerc de Dijon	CRB IGR + CRB centre Léon Bérard
CKD-REIN	Non	Agence de la biomédecine	Centres participants sous supervision de l'EFS et de la biobanque de Picardie
COBLANCE	Non	Serveur IGR	CEPH + Institut Curie + CRB Henri Mondor
CRYOSTEM	Non	Serveur Hôpital Saint Louis	CRB des différents centres
E3N / E4N	Non	Serveur IGR	EFS + CRB IGR
HEPATHER	Oui, en cours	Serveur ANRS	Cell&Co Clermont Ferrand
HOPE-EPI	En préparation	Serveur sécurisé Inserm	CRB des différents centres
MEMENTO	Non, demande en cours pour appariement probabiliste	Centre investigation clinique de Bordeaux	Institut Pasteur Lille +CRB CHU Bordeaux
OFSEP	Non	Base de données OFSEP puis IRT B<COM>	Base TK®
PSY-COH	Non	Serveur UPMC	CRB Henri Mondor + CRB La Pitié-Salpêtrière (ICM)
RADICO	Non	Serveur dédié Hôpital Trousseau (APHP - Inserm)	Divers CRB selon les cohortes
RECONAI	Oui	Serveur sécurisé Inserm	EFS
CONSTANCES	Oui	CASD	IBBL Luxembourg

Source : Dossiers d'évaluation 2019.

Si la grande variabilité des solutions trouvées ne constitue pas en soi un problème, leur caractère parfois aléatoire et l'absence de vision d'ensemble rendent difficiles une exploitation optimale des échantillons par un large ensemble d'utilisateurs.

1.4.2.2 L'absence de lignes directrices dans les relations avec les milieux de recherche concernant l'ouverture et la mise à disposition des données

Une deuxième difficulté renvoie à la diversité des pratiques retenues en matière de mise à disposition des données et échantillons collectés. Le cahier des charges invitait en effet les porteurs de projet à construire des cohortes « *ouvertes à la communauté scientifique et permettant la constitution de bases de données et de collections d'échantillons biologiques qui pourront être mutualisées afin d'optimiser la gestion administrative, réglementaire et matérielle, ainsi que l'encadrement éthique et la gestion de la propriété intellectuelle* ». Il n'est donc pas surprenant de voir l'ensemble des porteurs de projets se réclamer des mêmes principes de qualité et d'ouverture des données (cf. fiches en annexe 4).

³¹ Arrêté du 22 mars 2017 relatif au référentiel de sécurité applicable au Système national des données de santé, disponible sur : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000034265125&categorieLien=id>

Pourtant, la mission relève que la notion de *FAIR data*³², mise en avant par les instances européennes et nationales³³ de la recherche, n'apparaît guère dans les dossiers d'évaluation : chaque cohorte décrit des règles et pratiques qui lui sont propres vis-à-vis des projets présentés par les chercheurs académiques, et ce sujet ne semble guère constituer pour leurs équipes une préoccupation ou un axe de travail.

En outre, les règles et circuits concrets d'accès aux données diffèrent d'une cohorte à l'autre : accès gratuit (RECONAI) ou contribution aux coûts variables (OFSEP), modulables selon la participation ou non à la vie du projet (ELFE) ; examen par un conseil scientifique (le plus souvent) ou par l'instance de pilotage opérationnel (CANTO) ; inscription dans des axes de recherche prédéfinis ou appels ouverts ; examen préalable de faisabilité ou même vérification que le projet présenté ne contredit ou ne concurrence pas ceux portés par l'équipe responsable de la cohorte.

Tout se passe comme si chaque cohorte se comportait en PME productrice de données et d'échantillons dont elle entend rester propriétaire et dont elle régule, selon des règles qui lui sont propres, l'accès. Si l'on peut comprendre que des équipes cherchent à valoriser le travail souvent mal reconnu de production de données, on ne peut que regretter l'absence de cadre global qui rendrait plus visible le potentiel des cohortes et en faciliterait l'exploitation.

1.4.2.3 Des problèmes de gestion des ressources humaines et d'acquisition de compétences

Une troisième fragilité réside dans le manque de certaines compétences critiques, notamment en matière de gestion et de traitement des données, notamment lorsqu'elles ont vocation à être appariées avec des données médico-administratives ou d'autres données massives. Même s'il n'est pas possible de dresser un tableau complet de la structure des emplois des cohortes, compte tenu des incertitudes de périmètre et de la diversité des unités de compte (effectifs physiques ou équivalents temps plein) utilisé dans les dossiers d'évaluation (*cf.* annexe 2), la mission relève toutefois que la plupart des cohortes se caractérise par une proportion variable mais souvent élevée d'emplois temporaires : la part de personnels en contrats à durée déterminée varie de 23% (COBLANCE) à 62% (CONSTANCES), pour atteindre jusqu'à 95% (RADICO).

Face à ces contraintes tenant à la structure des emplois, les responsables des cohortes ont fait part à la mission de difficultés fréquentes concernant la pérennisation et la transmission des compétences ou expériences nécessaires au bon fonctionnement des dispositifs. La mission a d'ailleurs noté que les emplois temporaires étaient particulièrement fréquents lorsque les cohortes et leurs ressources humaines, étaient administrativement portées par des structures publiques, qui n'ont pas toujours la capacité de proposer des emplois statutaires, dans ce domaine comme dans d'autres de la recherche en santé. À l'inverse, les fondations coordonnant des cohortes (cas de CANTO ou d'OFSEP) ont eu recours à des contrats à durée indéterminée pour stabiliser leurs personnels (*cf.* annexe 4).

Ces obstacles à la stabilisation des personnels sont assortis de difficultés de recrutement, notamment pour les métiers de gestion et d'analyse des données (*data scientists* et *data analysts*) nécessaires pour collecter, stocker, traiter, documenter et mettre à disposition les données. En effet, le salaire moyen sur le marché pour ces nouveaux métiers – autour de 45/50 000 € annuels pour un *data*

³² Findable, Accessible, Interoperable, Reusable datas

³³ Plan national pour la science ouverte de juillet 2018, p. 6 : « notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes. », disponible sur : https://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf

scientist – excède largement les possibilités offertes par les opérateurs de recherche, organismes ou universités³⁴.

Ces fragilités communes à la plupart des cohortes rendent manifeste le besoin d'un cadre commun renforcé pour encadrer le développement des cohortes et les mettre au service, tant des communautés scientifiques que du monde de la santé.

2 Des perspectives et un environnement brouillés et incertains

Si le passage en revue des cohortes soutenues par le PIA fait à la fois apparaître des acquis et des fragilités, leurs perspectives, et surtout, leur environnement restent en large part brouillés et incertains, et ce pour un ensemble de raisons : leur place est loin d'être clairement définie dans le paysage des outils statistiques et épidémiologiques portant sur les thèmes prioritaires en santé (point 2.1) ; leurs liens restent à concrétiser avec un système de données médico-administratives lui-même en devenir (point 2.2) ; leur place reste à établir dans la structuration en cours de la politique de recherche en santé (point 2.3) ; des doutes existent sur la viabilité d'un « modèle économique » tablant initialement sur une large participation des partenaires privés (point 2.4) ; enfin, et surtout, l'absence d'une politique ambitieuse et coordonnée de mobilisation des données longitudinales pour la recherche en santé ne fournit ni priorités ni perspectives de financement claires pour guider leur évolution (point 2.5).

2.1 Une place incertaine des cohortes du PIA dans le paysage des outils statistiques et épidémiologiques portant sur les thèmes prioritaires en santé

Une première source de flou et d'incertitude réside dans le fait que les cohortes soutenues par le PIA sont loin d'être les seules à être considérées comme prioritaires dans les soutiens apportés par les pouvoirs publics et que, par ailleurs, leur articulation n'est pas clairement définie ou même réalisée *ex post* avec les autres grands dispositifs d'observation centrés sur les priorités de santé publique.

L'annexe 5 du rapport passe à cet égard en revue, à la fois pour la population générale et pour un ensemble de thèmes d'intérêt prioritaire en santé (nutrition, santé au travail, vieillissement et dépendance, maladies chroniques, cancers, VIH), une série d'autres outils d'observation structurants comportant, directement ou indirectement une dimension longitudinale, et mis en place à l'initiative des organismes publics ou fortement soutenus par eux.

2.1.1 D'autres cohortes épidémiologiques soutenues de façon séparée par les organismes publics

S'agissant des seules cohortes épidémiologiques, il apparaît d'abord que d'autres cohortes importantes ont été développées et soutenues en dehors du PIA, sans que les articulations ou les complémentarités aient été forcément organisées entre ces différents instruments, pourtant tous financés à hauteur significative par des fonds publics.

³⁴ La plus haute rémunération proposée par le barème des contractuels de l'Inserm établi en 2017 s'élève à 3 286,92 € brut mensuel pour un ingénieur de recherche de plus de 20 ans d'ancienneté

En dehors même des outils de suivi clinique développés dans un cadre hospitalier, on peut par exemple citer :

- l'étude NUTRINET, qui, ouverte en permanence à de nouveaux participants, suit en 2019 plus de 160 000 personnes de plus de 18 ans sur des questions ayant trait à leur alimentation, leur mode de vie, leur activité physique et leur état de santé ; non retenue au titre des cohortes ou infrastructures financées par le PIA, elle a bénéficié de soutiens de Santé publique France (SpF), du ministère de la santé, de Fondation sur la recherche médicale, ainsi que de financements sur projets de la part de l'ANR, l'Inca ou l'Anses ; bien qu'un projet de cohorte intégrée ait été envisagé à l'origine avec les responsables de CONSTANCES, cette démarche n'est pas entrée dans les faits, et seule une coopération « légère » et indirecte a été établie entre les deux études, sous forme de confrontation des réponses ou de répartition des sujets d'investigation tenant à la nutrition (*cf.* annexes 4 et 5) ;
- l'étude AGRICAN, réalisée en partenariat avec la MSA et le réseau FRANCIM (registres du cancer), qui reçoit des financements de l'Anses et du ministère de l'agriculture au titre du plan ECOPHYTO ; cette cohorte ciblant les agriculteurs et salariés agricoles suit encore, avec une profondeur de plus de dix ans, plus de 90 000 répondants concernant l'histoire détaillée de leurs expositions professionnelles, mais aussi leur état de santé et leurs comportements ou modes de vie (alcool, tabac, habitudes alimentaires...) ; il s'agit là aussi d'un projet conduit de façon spécifique par rapport aux autres cohortes centrées sur la santé au travail ou aux cohortes en population se proposant, avec leurs méthodes propres, d'évaluer l'impact des expositions environnementales sur la santé (CONSTANCES, E3N/E4N).

C'est néanmoins dans le domaine de la santé au travail que les articulations sont aujourd'hui le plus clairement définies, avec le portage et le financement par l'agence Santé publique France de deux cohortes dites COSET (Cohortes pour la Surveillance Épidémiologique en lien avec le Travail), prévues pour venir compléter l'étude CONSTANCES (centrée sur les affiliés au régime général), par le suivi de respectivement 30 000 et 20 000 assurés des régimes agricole et indépendants (ex-RSI) ; ces cohortes sont toutefois conçues pour être ré-ouvertes afin d'inclure de nouvelles générations et de conserver leur représentativité, alors que cela n'a pas, pour l'instant, été envisagé dans le cas de CONSTANCES.

A contrario, le suivi longitudinal des cancers et des maladies chroniques donne lieu à des initiatives qui demeurent dispersées. Ainsi, les cohortes soutenues par le PIA ne correspondent pas toutes aux priorités de l'Inca en matière de pathologies cancéreuses (qui iraient surtout actuellement, selon ses indications, aux cancers du poumon et du pancréas). De même, le suivi des pathologies chroniques, qui occupe une place centrale dans les préoccupations des grandes cohortes en population, est aussi ciblé par un projet ComPaRe de grande ampleur développé par l'Assistance publique-Hôpitaux de Paris (AP-HP), visant à recruter et suivre une *e*-cohorte de 100 à 200 000 adultes atteints d'une (ou plusieurs) maladie(s) chronique(s), en sollicitant leur participation directe pour recueillir des données « au plus proche du quotidien réel des patients »³⁵ (*cf.* annexe 5).

³⁵ Ce projet est financièrement soutenu par l'AP-HP, l'Université de Paris et l'Université Sorbonne Paris Cité et bénéficie d'un ensemble assez large de partenariats de la part d'associations de patients.

2.1.2 Des articulations insuffisamment maîtrisées entre les dispositifs d'observation des grands domaines de la santé publique

Plus généralement, au-delà des seules cohortes épidémiologiques, les différents domaines prioritaires des politiques de santé publique sont couverts de façon fragmentée, dispersée et peu articulée par une diversité de dispositifs d'observation (enquêtes, registres, bases de données médico-administratives), qui, de façon directe ou *via* des appariements, revêtent désormais le plus souvent une dimension longitudinale (*cf.* point 1.1.3, annexe 1 et annexe 5).

Cette dispersion et cette fragmentation ont clairement été mis en évidence dans un champ comme le suivi des médicaments en vie réelle où les experts auteurs en 2017 d'un rapport à la ministre de la santé indiquaient : « *Il y a beaucoup d'initiatives pour créer des sources de données, mais elles sont souvent locales, conçues indépendamment les unes des autres sans plan d'ensemble, ou bien sont des recueils ad hoc ponctuels. Cette dispersion empêche d'avoir des outils suffisamment puissants pour répondre aux questions que l'on se pose, au moins dans les domaines jugés prioritaires* »³⁶.

De même, s'agissant des données de recherche et d'étude sur la santé et le développement de l'enfant, l'avis élaboré en commun par le Conseil de l'enfance et de l'adolescence et Haut Conseil à la santé publique (HCSP) en octobre 2019 à partir de leurs travaux respectifs pose comme premier constat : « *Les données sur l'enfant sont nombreuses mais dispersées, incomplètes et manquent d'exploitation, de synthèse et de visibilité* »³⁷.

La mission a fait de son côté le même type de constat s'agissant, par exemple, des données relatives au processus de vieillissement et à la dépendance, auxquelles participent plusieurs dispositifs qui ont, entre eux, des relations ou articulations très limitées (*cf.* encadré 5 et annexe 5).

Encadré 5

Les suivis longitudinaux du vieillissement et de la dépendance

Parmi les dispositifs qui participent au suivi longitudinal du vieillissement et de la dépendance, figurent par exemple, sans que cette énumération soit exhaustive :

- un dispositif européen comme *Share*, constitué sous forme de panel longitudinal de 3 000 personnes de plus de 50 ans suivies tous les deux ans sur la période 2020-2024 et qui, initialement financé par l'appel d'offres Cohortes de l'IRESP, n'a pas été retenu par le PIA, et est désormais géré, après avoir été pris en charge par l'Insee puis par l'Irdes, par une UMR constituée entre le CNRS et l'Université Paris Dauphine ;
- une base de données longitudinales comme celle mise en place par la Drees, sous le nom d'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 ;
- les enquêtes *Care Ménages* et *Care Institutions* réalisées par la Drees en 2015-2016 et appariées avec sept sources de données administratives ; elles seront suivies en 2020-2022 d'une enquête Autonomie « tous âges » permettant à la fois l'étude des personnes âgées et en âge de travailler, ainsi que l'étude des populations handicapées par grand type de limitation ;

³⁶ Bernard Bégau, Dominique Polton et Franck von Lenne, 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, page 23.

³⁷ Haut Conseil de la famille, de l'enfance et de l'âge, Haut Conseil de la santé publique, *Avis relatif aux données de recherche et études sur la santé et le développement de l'enfant*, octobre 2019, page 2.

- les diverses cohortes ayant pour objet d'appréhender le développement et la prise en charge des maladies neuro-dégénératives (parmi lesquelles la cohorte MEMENTO), dont le Plan 2014-2019 consacré à cette thématique indiquait la nécessité d'une optimisation des données recueillies et d'une meilleure coordination entre études de cohorte³⁸ ;
- le recours aux données des cohortes en population, principalement CONSTANCES, mais aussi désormais E3N qui, au fur et à mesure de l'avancée en âge de ses participants voit ses thématiques d'intérêt se déplacer vers les maladies liées au vieillissement (Parkinson par exemple), alors qu'elle était précédemment centrée sur les risques de cancer (cf. annexe 4).

Source : Conseil national de l'information statistique et entretiens réalisés par la mission.

De façon plus générale, la mission a été frappée par le fait que, concernant les déterminants sociaux de la santé (diplômes, catégorie socio-professionnelle et surtout revenus des ménages), auxquels sont pourtant conduites à s'intéresser la plupart des études, à commencer par celles visant à appréhender les facteurs de risque des maladies chroniques ou le développement de l'enfant, seuls les dispositifs mis en place sous l'égide du système statistique public³⁹ ont accès ou projettent de mobiliser les données socio-fiscales rassemblées et traitées par l'Insee en collaboration avec les autres administrations publiques (ministère chargé des finances, CNAF) : même les plus importantes cohortes du PIA comme ELFE et CONSTANCES s'en tiennent en effet à des recueils d'informations déclaratives moins fiables et plus approximatives sur la base de questionnaires, transitant parfois par des professionnels de santé.

2.2 Des incertitudes sur le contenu et les modalités des liens avec un système de données médico-administratives lui-même en devenir

Les plateformes de données médico-administratives (BDMA) ont pour les cohortes plusieurs intérêts qui les rendent incontournables : permettre la constitution de nouvelles cohortes donnant lieu à un suivi « passif » ; servir de base d'échantillonnage ou de « témoin » ; les dispenser de certains recueils d'informations, en fonction du contenu et de la validité des données médicales rendues accessibles ; enrichir les informations collectées à la fois en étendant leur champ et leurs périodes d'observation et en fournissant une meilleure appréciation de leur représentativité (cf. annexe 1).

Or, la visibilité que les financeurs et les responsables de cohortes ont sur les potentialités et les modes d'accès à ces bases de données est aujourd'hui loin d'être assurée, ce qui constitue pour eux une source d'attentisme et de perplexité.

2.2.1 Des évolutions du SNDS...

En effet, le système national des données de santé (SNDS) institué en France par la loi du 26 janvier 2016 de modernisation du système de santé est à la fois très riche, mais d'accès techniquement complexe et en évolution continue avant et après cette date⁴⁰. L'annexe 6 du rapport retrace le

³⁸ Ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche, Ministère des affaires sociales, de la santé et du droit des femmes, Plan maladies neuro-dégénératives 2014-2019.

³⁹ Par exemple, l'*European Health Interview Survey -EHIS* s'agissant de la santé en population générale et le projet « d'EDP Santé » visant à croiser informations démographiques et socio-économiques détaillées et données médico-administratives sur la santé et le recours aux soins- cf. annexe 5),

⁴⁰ La loi de 2016 modifié et élargi le dispositif d'entrepôt de données constitué précédemment à partir du système national inter-régimes de l'assurance maladie (SNIIRAM) mis en place en 1998 et du programme de médicalisation des systèmes d'information (PMSI), dont le chainage était en vigueur depuis 2006.

contenu et les étapes de ces évolutions, ainsi que leurs implications pour les cohortes susceptibles de les utiliser.

Entre 2016 et la mi-2019, ce sont surtout les changements intervenus dans les modalités d'accès et les conditions d'utilisation de ces bases qui ont eu des répercussions sur l'environnement des cohortes : dépôt des demandes d'accès auprès de l'Institut National des Données de Santé (INDS), avis du CEREES (Comité d'Expertise pour les Recherches, les Études et les Évaluations dans le domaine de la Santé) mis en place en mai 2017 et chargé de fournir une appréciation méthodologique sur la cohérence entre la finalité de l'étude proposée, procédure maintenue devant la CNIL mais parfois sous des formes simplifiées et sans nécessité désormais, pour les projets de recherche scientifique publique, d'un décret en Conseil d'État pour les appariements demandant le recours numéro d'inscription au répertoire national d'identification (NIR). Même si ces modifications sont allées dans le sens d'une facilitation, avec des délais *a priori* encadrés, ces changements de réglementation répétés (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) ont conduit nombre de responsables de cohortes à retarder le dépôt des dossiers d'appariement projetés ou à reprendre plusieurs fois la procédure de soumission de leurs projets.

Par ailleurs, le référentiel de sécurité prévu par la loi de 2016 et défini par un arrêté du 22 mars 2017 a exigé qu'à compter de mars 2019, tous les systèmes mettant à disposition des données potentiellement « ré-identifiantes » constituées à partir du SNDS soient stockées dans un environnement homologué, ayant donné lieu à un audit préalable, avec des exigences particulières de sécurité pour l'accès aux informations. Cette exigence a également conduit certaines cohortes à demeurer dans l'attente de solutions adaptées ou à adopter, à titre transitoire et sans visibilité sur la durée de ces arrangements, des voies temporaires telles que le recours au Centre d'accès sécurisé aux données (CASD)⁴¹ (cf. point 1.4.2.1 et annexe 6).

2.2.2 ... aux nouveaux environnements créés par le *Health data Hub* et le projet France Cohortes

Depuis la mi-2019, deux autres évolutions sont en passe d'introduire de nouvelles modifications de forte ampleur dans l'environnement technique des cohortes, à la fois pour l'accès au SNDS et pour le stockage et la mise à disposition de leurs propres données.

En premier lieu, afin d'élargir le contenu médical des informations contenues dans le SNDS⁴², l'article 41 de la loi du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé a créé une plateforme des données de santé (ou *Health data Hub*-HDH) qui se substitue à l'INDS et qui vise à réunir, organiser et mettre à disposition les informations issues d'un SNDS progressivement

⁴¹ Comme l'a fait la cohorte CONSTANCES, sur la base d'une convention signée en juillet 2019 entre l'Inserm et le CASD, pour héberger et mettre à disposition des chercheurs ses données appariées dans des conditions sécurisées.

⁴² Les données médico-administratives du SNDS sont jusqu'à aujourd'hui limitées quant au contenu médical des informations collectées : ce contenu existe dans le cas des ALD et des soins hospitaliers, et peut être reconstitué de façon indirecte lorsqu'existent des médicaments ou des actes « traceurs » caractérisant spécifiquement certaines maladies, mais le système ne recueille pas de données cliniques ou paracliniques permettant de caractériser finement l'état médical des patients traités, notamment dans le cadre de la médecine de ville (compte rendus d'examen cliniques, résultats de prélèvements biologiques ou données d'imagerie).

enrichi à l'ensemble des données cliniques collectées lors des actes pris en charge par l'assurance maladie, ainsi qu'à un ensemble d'autres informations⁴³, les dispositions du type « Informatique et libertés et référentiels de sécurité du SNDS continuant par ailleurs à s'appliquer.

À la suite une mission de préfiguration pilotée par trois experts⁴⁴, la plateforme a été constituée en novembre 2019 sous forme d'un groupement d'intérêt public, qui intègre le secrétariat par lequel transitent toutes les demandes d'accès des porteurs de projet aux bases de données de santé⁴⁵, et qui continue à assurer leur accompagnement préalablement aux procédures devant la CNIL. Mais le *Health data Hub* se voit également conférer la mission d'effectuer en propre la mise à disposition des données du SNDS élargi (auquel s'adjoindraient les données des cohortes), ainsi que la promotion de l'innovation et la réalisation de traitements pour le compte de tiers ; il peut à cet égard accompagner, y compris financièrement, des projets sélectionnés et les producteurs de données qui s'y associent.

La mise en place du HDH a eu comme premières étapes le recensement des bases de données potentiellement concernées par une intégration dans le futur système (avec l'engagement de discussions avec les différents promoteurs de cohortes ou de plateformes), la mise en place d'une plateforme technologique (en l'occurrence le service d'hébergement de données *Cloud* de Microsoft) pour héberger de façon sécurisée des données individuelles de santé pseudonymisées et offrir des espace de travail et des services aux projets de recherche, et enfin la sélection d'une vingtaine de projets pilotes (*cf.* annexe 6).

D'un autre côté, dans le cadre de son projet d'unité mixte de services « France Cohortes » (*cf. infra* point 2.5.4), l'Inserm entend développer une plateforme informatique commune aux cohortes auxquelles l'organisme participe dans le cadre de son *data center* situé au Centre Informatique national de l'enseignement supérieur de Montpellier (CINES). Le projet a pour ambition de fournir un système d'information mutualisé qui soit conforme au référentiel de sécurité du SNDS et au RGPD, décharge les chercheurs de la gestion informatique des cohortes (matériels, logiciels, réseaux, installations, maintenance), soit ouvert et interopérable avec d'autres systèmes d'information, et doté de bulles sécurisées permettant aux chercheurs de traiter les données en toute sécurité. La solution-cible envisage ainsi une architecture technico-fonctionnelle reposant sur huit briques, dont les quatre premières ont trait au stockage de données et les quatre suivantes à leur exploitation et leur appariement (*cf.* annexe 6).

La mise en œuvre de ces deux projets doit avancer de façon concomitante au cours des années 2020 et 2021. Mais, en l'absence d'approche d'ensemble, des questions clés comme le contenu précis des services offerts de part et d'autre aux responsables de cohortes et, plus largement, aux autres chercheurs souhaitant bénéficier d'un accès à ce type de données, leurs « modèles économiques » (tarification des accès et services proposés aux utilisateurs), et surtout leur articulation concrète, entre eux et avec la plateforme de données sur le cancer par ailleurs mise en place par l'Inca, n'ont, pour l'instant, pas été précisées, non plus que, s'agissant du *Health data Hub*, les prochaines étapes d'enrichissement du SNDS en informations à contenu médicalisé.

⁴³ Notamment 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,

⁴⁴ *Health Data Hub*, Rapport de la mission de préfiguration pilotée par Marc Cuggia, Dominique Polton et Gilles Wainrib, octobre 2018.

⁴⁵ Le CEREES devenant le comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé.

2.3 Des incertitudes sur l’insertion des divers types de cohortes dans la structuration du paysage particulièrement complexe de la recherche en santé

Le caractère hybride, déjà signalé (*cf.* conclusion de la partie 1.1.2) des cohortes est souvent mis en avant par les porteurs de projet qui revendiquent pour elles à la fois le statut de « *projet de recherche* » et celui d’« *infrastructure de recherche*⁴⁶». Du premier, elles empruntent l’idée d’un questionnement préalable, de nature scientifique, c’est-à-dire contextualisé et faisant apparaître un besoin de connaissance auquel répondre ; du second, elles retiennent l’idée d’outil de production de connaissances (ici des données) et de services aux chercheurs (leur documentation et leur mise à disposition). Cette double nature est liée au fait que, à des degrés variables selon les projets, les questionnements préalables déterminent le contenu et les modalités des données collectées, qui nourrissent à leur tour de nouveaux questionnements et projets de recherche, et ce sur une période dépassant souvent une dizaine d’années.

Par leur caractère hybride, les cohortes s’insèrent difficilement dans les structures traditionnelles de la recherche, qui distinguent le plus souvent (c’est en particulier le cas à l’Inserm), activités de recherche et de service⁴⁷, les infrastructures ayant alors, soit le statut d’unité de service (sans activités de recherche), soit celui de plateforme au sein d’une unité de recherche.

La dynamique du PIA, centrée sur la notion de « *projet* » porté par des établissements « *partenaires* », dont un « *coordonnateur* », réunis dans un accord de consortium négocié après la signature de la convention attributive d’aide, ajoute un élément de complexité à cette distinction en favorisant le développement d’*objets*⁴⁸ aux statuts divers, plus ou moins bien insérés dans le monde de la recherche académique.

Dans le même temps, celui-ci est animé par des dynamiques de structuration qui voient le domaine de l’épidémiologie s’organiser, à l’échelle des grands sites franciliens dont le paysage a été profondément modifié par les fusions d’universités, en trois grands laboratoires de recherche – le CESP⁴⁹ (Paris Saclay), le CRESS⁵⁰ (Université de Paris) et l’IPLesp⁵¹ (Sorbonne Université) – tous créés postérieurement au démarrage des cohortes⁵².

⁴⁶ Selon la définition de l’Inserm, « les infrastructures de recherche correspondent à des équipements, des ressources et des services utilisés par les scientifiques pour mener leurs recherches et favoriser l’innovation dans leur domaine ».

⁴⁷ Dans les deux cas, l’unité est une création de nature conventionnelle initiée, lorsque la structure est mixte, par les établissements alors qualifiés de « tutelles ». La distinction emporte des conséquences non négligeables en matière d’affectation de politique scientifique, d’affectation des personnels et d’évaluation :

	Politique scientifique	Affectation des personnels	Evaluation
Unité de recherche	Définie et mise en œuvre par le directeur	Chercheurs et IT	Par le HCERES
Unité de service	Pas de politique scientifique propre ⁴⁷	IT uniquement	Interne (par les tutelles)

⁴⁸ Le terme revient fréquemment pour désigner les projets du PIA (IHU, IRT, Equipex, Labex) qui échappent au cadre traditionnel d’organisation de la recherche publique et revendiquent leur autonomie, notamment par rapport aux tutelles.

⁴⁹ Le centre de recherche en épidémiologie et santé des populations (UMR 1018) est un laboratoire commun de l’Inserm, de l’Université Paris Saclay et de l’Université de Versailles Saint Quentin en Yvelines (appelée à rejoindre l’UPS).

⁵⁰ Le centre de recherche en épidémiologie et statistique (UMR 1153) est un laboratoire commun entre l’Inserm et les équipes de Paris 5 Descartes, Paris 7 Diderot (désormais fusionné dans université de Paris) et Paris 13.

⁵¹ L’institut Pierre Louis d’épidémiologie et de santé publique (UMR-S 1136) est un laboratoire commun Inserm / Sorbonne Université créé en 2014 et renouvelé en 2019.

⁵² L’Ile-de-France représente 39 % de la production scientifique nationale en matière de recherche médicale (STRATER 2016) et 1,4 % de la production mondiale.

Il résulte de ces dynamiques un paysage particulièrement complexe, avec des cohortes portées par des opérateurs privés (CANTO, PSY-COH) ou publics, aux statuts divers (unité mixte de service – UMS – pour CONSTANCES et RECONAI, équipes internes à un laboratoire de recherche pour CKD REIN et E3N/E4N, association pour CRYOSTEM, fondations pour OFSEP et PSY-COH), parfois très intégrées à des laboratoires de recherche (HEPATHER et l'IPLesp) ou, au contraire, sans liens autres que scientifiques avec les structures académiques (CRYOSTEM, CANTO).

Cette complexité entraîne un certain nombre de difficultés :

- d'une part, elle nuit à la visibilité des cohortes dont le potentiel demeure peu exploité par les équipes de recherche autres que celles très proches de leurs responsables (*cf.* 1.3.1) ;
- d'autre part, elle ne favorise pas les échanges entre les divers projets et la recherche de solutions communes à des problèmes pourtant partagés, ainsi que l'a déploré le jury d'évaluation de septembre 2019 dans son commentaire final ;
- enfin, elle prive les cohortes d'un appui non seulement sur le plan des fonctions de support et de soutien (ce qui est l'ambition de France Cohortes), mais aussi de l'animation scientifique qu'autoriserait l'insertion des cohortes dans de grands laboratoires de recherche et dans des réseaux méthodologiques et scientifiques partagés.

Poursuivre l'insertion des cohortes dans un paysage de la recherche désormais en voie de stabilisation, avec le renouvellement en cours des grands laboratoires d'épidémiologie, constitue donc un enjeu important pour apporter aux cohortes un soutien dans la durée qui a fait défaut jusqu'à présent. Pour autant, la mission estime que cette insertion devrait se faire de manière différenciée et pragmatique, prenant en compte :

- le type et la portée de chaque cohorte, les cohortes en population et à vocation nationale méritant un pilotage qui devrait être plus large qu'une gouvernance interne à un laboratoire de recherche ;
- l'articulation optimale des instances de gouvernance entre le niveau cohorte et le niveau laboratoire, de manière à éviter une « comitologie » déraisonnable ;
- la nécessité de remédier à la complexité de certaines situations (*cf. infra* point 3.1).

À cet égard, il semblerait important à la mission que les réflexions en cours sur la loi de programmation de la recherche, qui se sont concentrées sur les contours de la feuille de route⁵³ et le mode de gouvernance des infrastructures de recherche⁵⁴ prennent en compte et affichent plus clairement l'objectif de considérer certaines cohortes comme de telles infrastructures, et d'adapter en ce sens leur fonctionnement et leur pilotage.

⁵³ Groupe de travail Financement, p.49 : « le groupe de travail propose la mise en place d'une feuille de route nationale élargie des IR. Pour être inscrite sur la feuille de route, chaque IR devra être réellement ouverte aux chercheurs, faire l'objet d'une large mutualisation. Un taux d'ouverture de l'ordre de 70 % pourrait être exigé »

⁵⁴ Groupe de travail Financement, p. 51 : « un resserrement de la gouvernance de chaque infrastructure autour de ses principaux financeurs, tout en garantissant la représentation des partenaires minoritaires, permettrait de gagner en efficacité ».

2.4 Des doutes sur la viabilité d'un « modèle économique » tablant sur une large participation des partenaires privés

« La valorisation de la cohorte et des résultats par le développement de partenariats avec le secteur privé » constitue l'un des objectifs de l'appel à projet Cohortes de 2010. La démarche promue par le PIA, qui était de rechercher un « effet de levier » aux financements publics, pèse avec le recul sur l'inscription des cohortes dans la durée, dans la mesure où un « modèle économique » (comprendre *business model*⁵⁵) apparentant ces dispositifs à des entreprises proposant un produit (des données) à des clients potentiels (publics ou privés) n'est, dans la plupart des cas, pas adapté à leur logique et à leurs objectifs scientifiques, notamment s'agissant des cohortes en population centrées sur les déterminants de santé et les facteurs de risque.

Dans ce cadre, les responsables de projets ont néanmoins tenté de satisfaire à la démarche du PIA en initiant deux approches complémentaires.

La première a consisté à rendre payant l'accès aux données pour les autres équipes de chercheurs, avec des redevances dont les tarifs restent néanmoins d'un niveau inférieur pour les chercheurs académiques que pour ceux du secteur privé, et peuvent varier selon le type de relations engagées avec l'équipe responsable de la cohorte (collaboration ou prestation de service pour l'accès aux données). Ce principe de tarification, comme d'ailleurs les pratiques en vigueur dans le monde académique⁵⁶, ne fait pas consensus : ainsi, le conseil d'administration de l'INED, établissement coordinateur de RECONAI, s'est prononcé contre le principe même de tarifier les accès aux données. D'autres cohortes ont adopté des pratiques diverses : certaines facturent les travaux statistiques préalables (OFSEP), ou des supports à la mise à disposition des données (E3N), d'autres demandent d'entrée une contribution financière (CANTO, CONSTANCES, où elle est fixée à 15 % des subventions publiques reçues) ou le reversement d'une partie des sommes contractualisées dans le cadre d'un projet partenarial (COBLANCE). Quelle que soit la diversité des solutions retenues, cette tarification n'apporte qu'une contribution modeste au financement des cohortes.

La seconde approche a eu recours à la mobilisation de financements privés pouvant eux-mêmes revêtir plusieurs formes. Si l'on excepte le cas de PSY-COH, où une entreprise s'est engagée dans une action de mécénat au profit de la cohorte de patients atteints de troubles bipolaires, l'apport de financements privés s'effectue le plus souvent dans le cadre de partenariats, parfois pluriannuels, qui permettent de bénéficier de livrables tels des tableaux de bord et des informations régulières sur l'activité de la cohorte ainsi que d'avoir accès à ses bases de données et parfois, à ses biobanques, en vue de la réalisation de projets de recherche. Deux modèles de partenariat ont, à cet égard, été mis en place par les cohortes : des partenariats public/multiprivés avec un ensemble d'industriels et des partenariats spécifiques au projet, sachant que les finalités et les modalités de ces derniers donnent en principe lieu, dans l'un et l'autre cas, à approbation par les instances scientifiques et décisionnelles de la cohorte.

55 Expression que la délégation générale à la langue française et aux langues de France recommande de traduire par « modèle d'entreprise ».

56 Le GIS Ibisa (infrastructures en biologie, santé et agronomie) promeut l'idée que « la plate-forme affiche de manière formalisée l'offre de service, les conditions d'accès et les tarifs ». Mais GENCI, société civile dédiée au calcul intensif, met à disposition des chercheurs gratuitement les heures de calcul de ces trois centres, CINES, IDRIS et TGCI.

Les premiers éléments de bilan montrent toutefois que :

- le niveau des financements obtenus, qui proviennent d'un nombre très réduit de partenaires, reste globalement modeste, bien que non négligeable (cf. annexe 2) : la part des financements privés représente, sur la période 2011-2018, 19,5% de l'aide attribuée (cf. annexe 2). Toutefois, ce financement s'avère élevé dans le cas d'HEPATHER où le montant des financements privés (12 M€) représente les deux tiers du total des financements publics (10 M€ d'aide du PIA et 8 M€ de financements de l'ANRS) ;
- leur volume et leur temporalité dépendent de la nature des cohortes :
 - les partenariats industriels des cohortes en population restent souvent très limités en nombre et en volume : par exemple, seules deux grandes entreprises ont un partenariat pluriannuel avec l'infrastructure CONSTANCES, principalement pour l'accès à des tableaux de bord et la possibilité de conduire des projets de recherche sur des populations d'intérêt spécifique ;
 - certaines cohortes de patients sont au contraire privilégiées par les industriels lorsque leur champ s'accorde avec des opportunités de développement de produits ou de thérapies (HEPATHER, CKD REIN, OFSEP) : la plupart de ces accès et projets visent à obtenir des résultats agrégés concernant de potentiels « publics cibles » des traitements en cours de préparation, mais ils contribuent parfois plus directement à la réalisation d'études post-AMM concernant l'impact et l'efficacité en vie réelle de certains traitements ; dans le cas où les enjeux sont importants, comme cela a été le cas avec la cohorte HEPATHER, les concours privés peuvent alors prendre une place prédominante pendant quelques années, avant de refluer par la suite, ce qui, dans les deux cas de figure, pose question pour le pilotage des travaux dans la durée ;
- la nature – privée ou publique – de l'établissement coordinateur (qui souvent pilote la structure de valorisation) n'a guère d'incidence sur la part de financement privés, sachant que pour les cohortes ayant pour partenaire l'Inserm, sa filiale Inserm Transfert (IT) a joué un rôle important dans la définition et la mise au point de ces partenariats (cf. *infra* encadré 6).

Au vu des résultats qui précèdent, la mission émet des doutes quant à la viabilité d'un modèle qui tablerait sur un financement important et durable des grandes cohortes nationales par des partenaires privés : leur motivation, comme leur horizon temporel, diffèrent nécessairement de celles d'instruments pensés pour se déployer sur le long terme, et dont la finalité première reste la production de connaissances.

Par définition complémentaires, ces financements doivent en outre être regardés à l'aune des contraintes légales (interdiction de vente et d'exploitation commerciale des données de santé⁵⁷) et des exigences croissantes d'intégrité scientifique pour la crédibilité même des projets engagés.

C'est pourquoi la mission considère que la question du modèle économique, sensible sur le plan sociétal, ne devrait pas être laissée à la seule initiative des porteurs de projet, mais être abordée dans un cadre politique global.

⁵⁷ Article L 1111-8 et article L 4113 du code de santé public.

Encadré 6

La politique d'Inserm Transfert en matière de cohortes

Filiale privée de l'Inserm en charge de sa politique de valorisation, IT accompagne depuis 2012 les dix cohortes soutenues par l'organisme dans leur politique partenariale. L'accès aux données produites par les cohortes, permettant de conduire des études cliniques, pharmaco-épidémiologiques ou d'identification de biomarqueurs, a conduit IT à cibler quasi-exclusivement les industries pharmaceutiques. Des partenariats mono ou multipartenaires, encadrés par des principes rappelant notamment le respect de l'indépendance scientifique et la maîtrise de sa stratégie par la cohorte, ainsi que des règles de transparence et d'éthique dans l'utilisation des données, formalisent la relation contractuelle.

Le niveau de financement demandé dans le cadre des partenariats peut varier selon le niveau et l'ancienneté du partenariat, les partenariats du type « Gold » permettant, par exemple, l'accès aux prélèvements biologiques et, parfois, la participation à certaines instances de la cohorte.

En contrepartie de livrables (tableaux de bord, liste des projets conduits, accès à des résultats obtenus dans le cadre d'un projet de recherche), qui excluent par principe tout accès direct aux données individuelles, les industriels s'engagent à contribuer financièrement à la construction et la maintenance de l'infrastructure et à prendre en charge les coûts associés à son projet de recherche.

Inserm Transfert incite par ailleurs les cohortes à mettre en place un « comité des industriels » pour conforter la transparence et la fluidité des relations entre la cohorte et ses partenaires privés. Ces partenariats ont concerné une petite vingtaine d'industriels et ont permis de mettre en œuvre 32 projets.

Source : Note interne d'Inserm Transfert.

2.5 L'absence d'une politique ambitieuse et coordonnée de mobilisation des données longitudinales pour la recherche en santé

Le besoin d'une politique ambitieuse de mobilisation des données longitudinales pour la recherche en santé a été souligné il y a plusieurs années dans le cadre des stratégies nationales de santé et de recherche⁵⁸.

Comme cela a été montré sans le point 1.2.1 de ce rapport, les financements dégagés par le PIA pour les cohortes à partir de 2010-2011 ont été substantiels et ont constitué une forme de rattrapage, que l'on peut considérer comme indispensable, tant au regard du retard pris par la France en la matière que de l'importance désormais acquise par ce type d'études longitudinales, devenues l'un des instruments de référence de la recherche internationale en santé publique. Bien que conséquent, cet effort ne s'est toutefois pas inscrit dans une véritable stratégie de mobilisation coordonnée de ce type de données.

Tout d'abord, la sélection des projets soutenus par le PIA a eu lieu sur la base de critères d'excellence appréciés par des jurys d'experts internationaux, à partir de projets émergeant du terrain (approche dite « *bottom/up* »), sans que des priorités méthodologiques (concernant par exemple la place des grandes cohortes en population) ou thématiques sur les besoins, sanitaires mais aussi, par exemple, sociaux, à considérer comme prioritaires d'un point de vue collectif, aient préalablement été

⁵⁸ « En épidémiologie et en santé publique, les données proviennent d'enquêtes horizontales ou de cohortes longitudinales en population générale. Le couplage de telles études avec les données extraites des grandes bases administratives représente un enjeu majeur pour l'essor de la recherche en santé publique, en pharmaco-épidémiologie et en sécurité sanitaire », SNR, Proposition des groupes de travail, disponible sur : <https://www.enseignementsup-recherche.gouv.fr/cid86688/strategie-nationale-de-recherche-france-europe-2020.html>

déterminées. Le mode de pilotage des projets sélectionnés n'a, en outre, qu'assez peu permis d'associer les ministères concernés à leur développement et à leur utilisation au bénéfice des politiques publiques.

Avant d'esquisser, comme le tente la quatrième partie de ce rapport, les voies d'une nouvelle politique en la matière, il est utile de confronter ce constat à des expériences étrangères beaucoup plus ambitieuses et conduites sur longue période (2.5.1). Le bilan n'a guère été conforté par les tentatives de structuration de la recherche en santé publique (2.5.2) et de coordination des instruments (2.5.3) intervenues au cours du temps, tandis que lancement récent du projet France Cohortes comporte encore nombre d'incertitudes (2.5.4).

2.5.1 Un manque d'ambition global au regard de certaines expériences étrangères

2.5.1.1 Des expériences étrangères fécondes, conduites sur longue période

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 (cf. annexe 7).

Or, sémantiquement et conceptuellement, les premières cohortes épidémiologiques sont nées au Royaume-Uni et aux États-Unis, témoignant de l'enracinement dans le monde anglo-saxon d'une puissante culture de santé publique et ouvrant la voie aux autres pays développés⁵⁹

Pionnières et passées à la postérité depuis, ces grandes cohortes fondatrices ont marqué les esprits par leurs résultats scientifiques importants et notoires, et ont contribué, de manière décisive, à l'essor de l'épidémiologie et à l'acquisition par cette discipline d'un véritable statut scientifique.

Elles ont permis de dépasser les seules enquêtes « cas/témoins », 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.

À partir des années soixante-dix, avec la reconnaissance du caractère multifactoriel de l'étiologie des maladies chroniques, les grandes cohortes se sont considérablement développées.

Cet accroissement quantitatif s'est doublé d'un perfectionnement qualitatif de l'instrument « cohorte » : établissement de modèles causaux des maladies et raffinement progressif des méthodes statistiques utilisées pour traiter les données longitudinales. S'est combinée 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.

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 de « méga-cohortes ». Leur nombre est estimé au début des années 2010 à plus de 100 de par le monde⁶⁰.

⁵⁹ 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.

⁶⁰ Marie Zins et Marcel Goldberg, « Les nouvelles méga-cohortes en population en Europe », *ADSP*, n°78, mars 2012.

De nouveaux développements scientifiques ont en effet rendu 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 leurs interactions avec les caractères génétiques.

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. 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, sachant que, dans ce mouvement général de rattrapage européen, le Royaume-Uni a néanmoins conservé sa place de locomotive, offrant un environnement propice à l'établissement et au développement de grandes cohortes en population, appariées aux sources administratives (*cf. infra* point 2.5.2.1).

À cet égard, 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 : ainsi, les « cohortes » scandinaves sont plutôt, 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.

Reste que la richesse des expériences étrangères dans le domaine des cohortes épidémiologiques peut être une source d'inspiration importante pour les politiques françaises, qui pourrait être mieux mise à profit.

2.5.1.2 Une politique particulièrement ambitieuse au Royaume-Uni, centrée sur les cohortes en population générale

L'expérience du Royaume-Uni est particulièrement précieuse à analyser, dans la mesure où ce pays 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, que ses autorités estiment nécessaires (*cf. annexe 8*).

Le financement de la recherche épidémiologique au Royaume-Uni est caractérisé par la coexistence et l'action complémentaire du *Medical Research Council (MRC)* et de deux autres institutions. À cet égard, un fait 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 où le *Wellcome Trust*, organisation caritative dédiée à la promotion de la recherche biomédicale, héritière des laboratoires pharmaceutiques éponymes, joue un rôle central.

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 sept agences du *UK Research and Innovation (UKRI)*, organisme public central chargé de la stratégie britannique en matière de recherche, qui met traditionnellement un accent notable sur les questions d'inégalités sociales, notamment en santé.

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 ».

Indépendamment de ces grandes cohortes, le NHS développe actuellement 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*).

2.5.1.3 Quelques enseignements de l'expérience britannique

Un certain nombre d'enseignements tirés de l'expérience du Royaume-Uni, peuvent éclairer les choix à faire en France en la matière. Ils découlent des pratiques déjà adoptées Outre-Manche comme des orientations nouvelles qui y sont envisagées ou en cours de planification.

L'ampleur, parfois considérable, et le spectre, très large, des grandes cohortes britanniques en population générale peuvent pour partie servir de modèle de référence, même s'ils résultent 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*, qui suit 500.000 personnes âgées de 40 à 69 ans incluses entre 2006 et 2010, par une méga-cohorte de cinq 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 cinq 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 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 les domaines de la santé et 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 qui organise une complémentarité des acteurs et des possibilités d'accroissement notable des financements ;
- la nécessité reconnue par ces grands acteurs de prioriser et de coordonner les décisions concernant les grandes cohortes en population ;
- le caractère quasi systématique des appariements avec des bases de données administratives, très riches en informations sociales, comme médicales. À cet égard, le Royaume-Uni dispose, avec le *NHS* d'un système de soins détenant des informations sanitaires longitudinales

concernant toute la population, et ouvrant, avec le chaînage, des possibilités considérables d'études cliniques, de santé publique et sociales ;

- l'intégration systématique des considérations liées à la génomique, à l'épigénétique, au métabolisme ou à la microbiologie (- *omics*), mais aussi les données d'imagerie ou de technologies émergentes, 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 applications et l'usage de l'Internet offrant, à cet égard, des possibilités intéressantes pour l'inclusion des personnes et la collecte des données.

Enfin, la connaissance de l'existence même des diverses cohortes, ainsi que l'accès aux données et aux échantillons qu'elles détiennent, sont des sujets fortement mis en avant comme des enjeux décisifs : la pratique de l'accès libre aux données, ou *open data*, est ainsi très répandue au Royaume-Uni, et un effort a été entrepris, compte tenu de la dispersion qui continue à exister en ce domaine, pour harmoniser les pratiques des cohortes, l'objectif affiché étant celui d'une politique d'*open data* généralisée.

2.5.2 En France, des projets de développement d'une politique coordonnée de recherche en santé publique restés peu opérants

L'idée de donner davantage d'ambition, de définir une stratégie et de mieux coordonner la recherche en santé publique n'est pas récente puisqu'elle présidait déjà à la création de l'institut virtuel de recherche en santé publique en 2004, puis du groupement d'intérêt scientifique Institut de recherche en santé publique (Iresp) mis en place en 2007, avec pour objectif de fédérer l'ensemble des acteurs conduisant des travaux de recherche en santé publique.

Cet effort se heurte d'une part à un pilotage incertain, partagé entre l'Institut thématique multi organismes (ITMO) de l'alliance Aviesan (conçu initialement pour peser sur la programmation de l'ANR⁶¹), dont le rôle est de « *faciliter et de coordonner l'activité des équipes de recherche qui développent des travaux consacrés à la santé publique ou à la recherche clinique* »⁶² et l'Institut thématique Santé publique de l'Inserm qui « *vise à faciliter et coordonner l'activité des équipes de recherche qui développent des travaux consacrés à la santé publique ou à la recherche clinique* »⁶³ - les deux fonctions étant d'ailleurs occupée par la même titulaire.

D'autre part, les bilans successifs des responsables de l'Iresp montrent, qu'au-delà de l'animation scientifique (avec plus d'une vingtaine de projets de recherche financés) et d'une analyse partagée des enjeux, peu d'avancées concrètes ont été enregistrées, que ce soit sur la question des usages des données de santé ou encore de la programmation de la recherche. Si un plan d'action pour l'organisation de la programmation de la recherche en santé publique lui a bien été demandé en 2014-2015, la présidente de l'Iresp déplorait, au moment de son départ deux ans plus tard, de ne pas en avoir vu la mise en place⁶⁴. Ce plan d'action préconisait l'élaboration d'une stratégie coordonnée

⁶¹ L'objet de la convention constitutive de l'Alliance précise qu'il s'agit « élargir la fonction de coordination nationale et de faciliter la mise en œuvre de la programmation scientifique concertée par les différentes institutions partenaires ».

⁶² Présentation de l'ITMO Santé publique sur le site d'Aviesan.

⁶³ Site de l'Inserm.

⁶⁴ Aviesan, *Plan d'action pour l'organisation de la programmation de la recherche en santé publique*, juillet 2015, et Geneviève Chêne, note de bilan, avril 2017, disponible sur le site de l'IRESP.

comprenant la définition de priorités thématiques s'articulant avec les stratégies nationales de santé et de recherche (SNS et SNR) ; ces priorités étaient appelées à se traduire dans un exercice de programmation ambitieux, doté de moyens suffisants et associant l'ensemble des acteurs de la recherche et de la santé.

Son absence de mise en œuvre, pointée par d'autres rapports récents (comme celui consacré par l'Igas et l'IGAENR en 2018 à la recherche clinique)⁶⁵, laisse pendante la question essentielle de l'articulation entre priorités de fond et outils d'observation, qui demeure traitée de façon cloisonnée par une variété d'acteurs ayant tendance à promouvoir leurs propres dispositifs.

La mission a, à cet égard, pu constater que, depuis quelques années, la dissociation s'est approfondie entre les instruments d'observation mis en place, en matière de santé, par le système statistique public et les cohortes épidémiologiques relevant du monde de la recherche, avec entre les deux peu d'échanges réciproques (cf. point 1.3.3 encadré 3). Les instances de coordination du système statistique que sont le Conseil national de l'information du système statistique (Cnis) et le comité du label) ne connaissent, en tout état de cause, parmi ces cohortes que les grands instruments à dimension sociale, à savoir ELFE et CONSTANCES.

2.5.3 Des tentatives embryonnaires de coordination et de mutualisation des cohortes

Les initiatives prises pour combattre le risque de dispersion restent, de l'avis de la mission, mais aussi des responsables de cohortes, très embryonnaires.

Une première initiative appréciable, prise par Aviesan, a consisté à mettre en place un catalogue en ligne des principales bases de données individuelles en santé de source française potentiellement utiles au développement de la recherche et de l'expertise en santé publique. Ce catalogue identifie de façon spécifique les cohortes, à côté des bases administratives et des registres ; et ce dans un spectre très large⁶⁶. Il semble toutefois que sa mise à jour ne soit plus complète ni régulière, ce qui est regrettable tant l'information partagée sur les outils d'observation existants est la base indispensable d'une politique d'appui et de coordination renforcée.

Les autres tentatives de coordination se réduisent à des initiatives intéressantes, mais d'écho relativement faible, telles « *l'animation d'un forum de mutualisation des réflexions entre chercheurs comme le club Cohortes (avec l'ANR et le CGI) et d'espaces de promotion des partenariats publics-privés autour des cohortes épidémiologiques comme le Cohort Innovation Day (avec l'Alliance pour la Recherche et l'Innovation des Industries de Santé, l'Ariis)* »⁶⁷.

Avis que semble partager le jury international réuni en septembre 2019, qui, dans ses commentaires généraux sur le bilan d'ensemble des cohortes du PIA, s'est étonné de cette quasi-absence de coopération conduisant chaque porteur de projet à inventer des solutions particulières à des problèmes communs.

⁶⁵ Christine d'Autume, Stéphane Elshoud, Emilie-Pauline Gallié, Laurent Gratieux, Armand Renucci, *Évaluation du programme hospitalier de recherche clinique (PHRC)*, rapport Igas-IGAENR, mai 2018.

⁶⁶ Le catalogue distingue les bases de données administratives, les registres de morbidité, les cohortes, les autres études longitudinales, les études de cas-témoins, et les études transversales <https://epidemiologie-france.aviesan.fr/epidemiologie-france/pages/le-portail-epidemiologie-france>

⁶⁷ Geneviève Chêne, *op.cit.*

2.5.4 Une relance récente, mais encore imprécise, au travers du projet Inserm France Cohortes, dont les contours et les priorités peuvent toutefois être discutés

Dans ce contexte, le projet France Cohortes porté par l'Inserm⁶⁸ apparaît comme une démarche pragmatique pour répondre à l'absence de mutualisation et au manque d'accompagnement des responsables de cohortes. Il s'agit, pour la direction de l'organisme, de répondre au « *cœur de besoins* » communs aux cohortes par une « *mise à l'échelle* » de l'expérience de la plateforme RADICO. Lancé depuis peu (les premiers groupes de travail se réunissent depuis la mi-2019 seulement), et bénéficiant de ressources du PIA à hauteur de 6,8 M€ (3,2 M€ programmés en 2020 pour France Cohortes *stricto sensu* et 3,6 M€ au titre des reliquats de crédits du projet Biobanque), ce projet présente encore, à la connaissance de la mission, de nombreuses incertitudes.

Elles portent notamment sur :

- son périmètre : portera-t-il sur les seules cohortes financées par le PIA et, au sein de ces dernières, uniquement sur les projets dont l'Inserm est partenaire (10) ? s'étendra-t-il à terme, hors PIA, à l'ensemble des cohortes impliquant des équipes de l'Inserm, et même au-delà pour la prise en compte de sujets partagés ?
- son objet : l'accent a été mis, dans un premier temps, sur la question des systèmes d'information (SI) (l'idée étant de construire « un seul SI, hautement paramétrable répondant à l'exhaustivité des usages et en adéquation avec les besoins métier »), et sur celle des ressources humaines, pour lesquelles quinze compétences clés ont été identifiées, dans une perspective d'appui et de mutualisation ; d'autres sujets, comme les biobanques, l'accompagnement juridique et administratif des projets ou les actions de communication, font aussi l'objet de fortes demandes et sont importants à prendre en compte à un horizon rapide pour apporter un soutien utile aux équipes des cohortes ;
- son calendrier de mise en œuvre, qui pourrait s'étaler, selon les chantiers, sur cinq à dix ans ;
- sa forme juridique et les ressources mobilisées : si le principe de constituer une UMS semble acquis, sa forme, ses tutelles (intégrant ou non certaines universités, notamment parisiennes) et les moyens qui y seront consacrés restent à ce stade encore en discussion.

Son articulation avec les structurations en cours des grands laboratoires de recherche (*cf.* point 2-4) reste enfin un sujet ouvert, les processus de gestion pouvant s'avérer singulièrement compliqués avec une multiplication d'acteurs et de niveaux (le siège, les services des universités cotutelles, la délégation régionale, l'UMS France Cohortes, les UMS de cohortes comme CONSTANCES et RECONAI et enfin les UMR concernées).

France Cohortes est donc en mesure d'apporter de premiers éléments utiles de réponses aux besoins de soutien exprimés par les responsables de cohortes, et ce dans le cadre d'une mutualisation, au moins partielle, qui peut être intéressante et utile. Toutefois, centré sur une démarche de mutualisation des fonctions de support et de soutien, le projet laisse entières deux questions :

- d'une part, celle de l'animation scientifique des cohortes, qui passe, pour la mission, à la fois par la mise en place de réseaux transversaux – concernant notamment l'identification des

⁶⁸ L'objectif 3 du contrat d'objectifs et de performance entre l'Etat et l'Inserm 2016-2020 vise à « *renforcer la visibilité et la gestion des infrastructures de recherche* » et annonce cette action de coordination (action 15) dans la perspective de mise en place du SNDS (action 16). https://www.inserm.fr/sites/default/files/2017-11/Inserm_ContratObjectifs_2016-2020.pdf

pathologies dans les bases de données médico-administratives ou la mise au point de modules de questionnement communs, par exemple sur la situation socio-économique, les modes de vie, ou le bien-être des participants – et par un renforcement des liens avec les laboratoires de recherche ; cela invite à intégrer dans la réflexion, de manière extrêmement précise, le partage des tâches et des ressources entre équipes responsables des cohortes, laboratoires d'accueil et France Cohortes, à la fois pour les fonctions scientifiques, de soutien et de support ;

- d'autre part, celle de la définition d'orientations et de principes communs s'imposant à l'ensemble des cohortes, en fonction de leur type et de leur objet, notamment en matière de relations partenariales avec les différents partenaires privés ou publics (notamment les acteurs de la santé), ainsi que de mise à disposition des données des communautés académiques.

À l'issue de ces réflexions, et en amont même de la définition d'une nouvelle politique concernant les cohortes et les données longitudinales en santé, la mission estime utile de mettre en œuvre quatre orientations.

Recommandation n°2 Actualiser, compléter et maintenir à jour le recensement du portail Epidémiologie France sur les bases de données individuelles en santé utiles pour l'expertise et la recherche en santé publique

Recommandation n°3 Organiser un partage d'informations et une réflexion autour des principales expériences étrangères en matière de cohortes épidémiologiques

Recommandation n°4 Considérer, dans la réflexion en cours sur la politique de structuration de la recherche, les cohortes en population comme des « infrastructures » à vocation nationale

Recommandation n°5 Remettre en question l'idée, s'agissant des cohortes soutenues par le PIA, d'un modèle économique visant un relais pérenne par des partenaires privés et d'un mode de pilotage différencié de celui des autres projets ou infrastructures du même type

3 À court-moyen terme, des besoins de consolidation et de clarification

Face aux fragilités, spécifiques ou plus largement partagées, recensées dans la première partie de ce rapport, et aux incertitudes sur leurs perspectives, leur positionnement et leur environnement énoncées dans la deuxième partie, la mission estime qu'une première priorité est, à court-moyen terme, de répondre aux besoins de consolidation et de clarification des cohortes soutenues depuis près de dix ans au travers du PIA. Celles-ci sont en effet loin d'avoir fourni tout le potentiel d'études et de résultats dont elles sont porteuses (*cf.* point 1.3.2) et, quelles que soient les limites de la politique de données longitudinales engagée dans ce cadre, il importe de tirer tout le parti possible des investissements d'ores et déjà réalisés, avant de relancer, comme le propose la quatrième partie de ce rapport, une politique plus ambitieuse à développer à l'horizon de dix ou quinze ans.

Les actions menées à court-moyen terme devraient ainsi procéder à une série de consolidations du dispositif existant (point 3.1), s'appuyant sur un « socle » garanti de ressources publiques pour les années à venir (point 3.2), ainsi qu'à un ensemble de clarifications dans l'environnement immédiat des cohortes (point 3.3).

3.1 Des besoins de consolidation dans plusieurs domaines

3.1.1 Des consolidations en termes de pilotage et de gouvernance

Un premier besoin de consolidation concerne la gouvernance et le pilotage de certaines cohortes, dont l'examen montre des défauts de fonctionnement ou des limites qui pèsent sur la poursuite du dispositif (*cf.* annexe 4). Les nouvelles conventions qui doivent être conclues avec l'ANR pour la période 2020-2024, et la révision des accords de consortium à réaliser à cette occasion, donnent l'opportunité de revoir ces modalités dans un sens à la fois plus opérant et plus efficace.

Parmi les sujets examinés par la mission figure au premier chef la gouvernance de la cohorte CONSTANCES, qui, eu égard aux difficultés de fonctionnement passées et à la restructuration en cours des universités franciliennes, a donné lieu à un débat sur le maintien de son statut actuel et de son rattachement à l'Université de Versailles Saint Quentin en Yvelines (UVSQ) ou, au contraire, son basculement vers l'un des grands laboratoires de recherche en santé publique en cours de constitution (CRESS ou CESP). La mission estime à cet égard que la solution la plus raisonnable pour les prochaines années consiste à conserver à l'infrastructure son statut actuel d'unité mixte de service (UMS), à conclure un nouvel accord de consortium associant à sa tutelle l'Inserm et les deux universités concernées (Université de Paris et Université Paris Saclay) dans le cadre de responsabilités financières mieux définies et partagées, à confier dans ce cadre une nouvelle lettre de mission à sa responsable et à établir des conventions de partenariat scientifique entre CONSTANCES et les différents laboratoires de recherche (CESP, CRESS, IPLesp, ...) avec lesquels elle est en relation et dont il serait malvenu qu'ils soient mis en rivalité à propos de la cohorte.

Parmi les consolidations à opérer de façon prioritaire, la mission a également pointé le blocage actuel du projet HOPE-EPI : les désaccords entre les partenaires d'origine du projet de 2010 continuent en effet à peser sur le déploiement du dispositif, y compris dans son format recentré après l'évaluation à mi-parcours (*cf.* annexe 4). Ce blocage, qui concerne, en amont même du projet de cohorte, le registre national des cancers de l'enfant, implique une action particulière des pouvoirs publics et un accompagnement fort de la gouvernance révisée du projet, notamment par l'Inca qui a un rôle majeur à jouer en ce domaine.

Concernant les autres projets, la mission a enfin relevé des améliorations à opérer concernant :

- la fragilité de la structure d'appui de CRYOSTEM, qui prend la forme d'une association n'impliquant directement ni les universités ni les centres hospitaliers à partir desquels le projet a été constitué ;
- le fonctionnement intermittent des instances de pilotage de MEMENTO, dont la relance peut participer d'une amélioration de l'accès et de l'appropriation des données par des équipes de recherche extérieures à son réseau de proximité ;
- une réflexion à engager sur les instances scientifiques et décisionnelles des cohortes CKD REIN et HEPATHER, dont certaines associent directement les partenaires industriels les plus importants de ces cohortes, ce qui a pu influencer sur la priorité donnée aux différents types de travaux et peut constituer une source d'instabilité dans la durée (cf. annexe 4).

3.1.2 Des consolidations en termes d'horizon et de potentiel scientifiques

Comme l'a montré la première partie de ce rapport, les cohortes lancées depuis 2010 représentent un potentiel scientifique important, dont il importe, en les inscrivant dans un temps long et en tentant d'en obtenir et d'en valoriser au mieux les résultats, de tirer tout le parti possible.

Il est clair à cet égard que la limitation initiale des projets à 2020 a souvent pour effet, compte tenu des délais de lancement, de limiter à quatre ou cinq années la durée de fonctionnement réel de cohortes financées par le PIA. De plus, la non prise en compte, au plan financier et/ou des compétences requises, d'aspects comme la constitution de biobanques, la réalisation d'analyses génomiques ou l'appariement avec les données médico-administratives, interrogent la réalisation de ces potentialités, dont il importe, aux yeux de la mission, qu'elle soit « ré-assurée » dans la période à venir.

Cela implique d'abord que la poursuite des suivis et de l'exploitation des informations recueillies soit garantie aux cohortes jusqu'au terme des périodes estimées scientifiquement pertinentes pour chacune d'entre elles : pour les cohortes ELFE et EPIPAGE, cela signifie par exemple que le suivi des enfants démarré à la naissance puisse se poursuivre pendant toute la période de leur adolescence et idéalement jusqu'à l'âge de 18 ou 20 ans.

Il importe également de s'assurer de la préservation et de l'utilisation dans les meilleures conditions possibles des échantillons biologiques collectés et de permettre la réalisation des analyses génomiques souhaitées, pour lesquelles certaines cohortes (dont CONSTANCES) ont pris du retard : la prise en compte des facteurs génétiques est en effet devenue une dimension incontournable de l'étude de la survenue et du cycle des maladies, en association et en interaction avec les autres facteurs de risques (sociaux, professionnels, environnementaux).

Ce raisonnement s'étend aux diverses recherches envisageables dans le domaine dit des « -omics », dans l'objectif de tirer le meilleur parti des prélèvements effectués « au long cours » auprès des participants suivis par les cohortes.

Il semble également indispensable à la mission que les appariements avec le SNDS ne soient pas de nouveau retardés dans l'attente de la mise en place des capacités informatiques et des services associés aux plateformes France Cohortes et/ou au *Health data Hub*. Ces projets mettront vraisemblablement plus de temps à apporter des solutions opérationnelles que leurs promoteurs ne l'ont annoncé au départ, et il ne faut pas exclure des dispositions intermédiaires qui, en termes de

techniques d'appariement ou de stockage sécurisé, permettraient aux cohortes d'accélérer la production de leurs résultats. Par ailleurs, l'étude, y compris juridique, de la possible mobilisation de données socio-économiques améliorées en provenance de la CNAF et/ou de l'Insee devrait être entreprise au plus vite, même si son aboutissement n'est pas immédiat.

Il faut toutefois noter que, pour la mission, le fait de permettre aux cohortes soutenues par le PIA d'aller jusqu'au bout de leurs potentialités ne signifie pas qu'un blanc-seing leur soit donné en termes de réouverture des inclusions ou d'élargissement de leur spectre d'intérêt, alors que les responsables de projets ont parfois été tentés, dans leurs derniers dossiers d'évaluation, de proposer ce type d'évolutions. Si des réouvertures peuvent sembler utiles pour suivre les nouvelles thérapies du cancer du sein ou les patients atteints d'une hépatite B, ce n'est pas forcément le cas pour toutes les cohortes de patients, et les pouvoirs publics doivent se conserver la liberté de donner la préférence à de nouveaux dispositifs de suivi.

L'évaluation régulière du pilotage et des avancées accomplies par les projets aidés demeure ainsi indispensable pour les années à venir. La mission propose à cet égard la mise en place d'un nouveau mode d'évaluation, ne se limitant pas à la réunion d'un jury scientifique sur la base de présentations rapides, mais comportant des visites de sites approfondies par des équipes d'évaluateurs mixtes, associant des compétences de type scientifique et administratif.

3.2 Des besoins récurrents d'un « socle » de ressources publiques dans les années à venir

Compte tenu du rappel qui vient d'être fait sur le caractère encore incomplet ou inachevé des résultats obtenus par les cohortes financées dans le cadre du PIA, et de l'intérêt de tirer le meilleur parti de leurs potentialités, la mission salue le fait que les aides publiques soient au moins pour partie poursuivies sur la période 2020-2024, ce qui était selon elle inévitable.

Les financements apportés par le PIA à l'issue de la phase d'évaluation résultent d'arbitrages⁶⁹ entre les demandes formées par les porteurs de projets et le respect d'une enveloppe préalablement déterminée, qui s'élève globalement à 103 M€ pour les cinq ans à venir à 85 M€ au titre de l'action Biologie-Santé, auxquels se sont ajoutés 18 M€ redéployés à partir de l'action « Instituts hospitalo-universitaires (IHU) »⁷⁰. 26,2 M€ ont été mobilisés au sein de cette enveloppe pour prolonger jusqu'en 2025 les dotations aux cohortes. La ventilation de ces dotations complémentaires, indiquée dans l'annexe 2 du présent rapport, fait apparaître des décisions de financement rapportées aux demandes d'aide très variables d'une cohorte à l'autre, de 29 % (HOPE-EPI) à 100 % (RECONAI), sans que ces variations soient aisément corrélables au contenu de ces demandes ou aux perspectives de cofinancement présentées par les porteurs de projets. Il ne semble pas non plus que les moyens ou services nouveaux, notamment en ressources humaines, susceptibles d'être apportés à compter de 2020 à travers la mise en place par l'Inserm de France Cohortes (*cf.* point 2.5.4) permettent de justifier ces variations. Les prolongations de financement annoncées au titre du PIA pour la période 2020-2024 sont en outre complétées au titre de 2020 par le ministère chargé de l'enseignement

⁶⁹ Sont pris en compte à la fois le classement des cohortes par le jury international réuni en septembre 2019 (avec quatre cohortes classées A : RECONAI, E4N, CKD Rein et CANTO) et de règles d'encadrement financier fixées ex ante (plafond de 30 % de l'aide initiale, plancher de 10 %). L'évaluation de CONSTANCES avait été réalisée préalablement, dans le cadre distinct.

⁷⁰ Suite à l'arrêt du financement de l'ICAN.

supérieur et de la recherche, à hauteur de 3,9 M€, en faveur des cohortes CONSTANCES, E4N et MEMENTO.

La diversité des modalités de présentation des demandes d'aide pour la période 2020-2024 (incluant dans certains cas l'année 2019, exprimant des besoins d'emplois en effectifs ou en équivalent temps plein) rend en tout état de cause difficile une analyse financière précise de la réalité du besoin à financer et, donc de la façon dont les décisions financières envisagées sont précisément en passe d'y répondre. Il est cependant possible de considérer les perspectives qui ont été mises en avant par les responsables des projets (cf. tableau 3).

Tableau 3 : Demandes d'aides présentées par les cohortes du PIA pour la période 2020-2024 (en M€)

Nom de la cohorte	Budget prévisionnel total	Principales perspectives mises en avant par les responsables de projet	Financement demandé
CANTO	14,96	Préparation de l'appariement avec le SNDS / Recrutements de statisticiens	4,76
CKD REIN	5,25	Maintien en conditions opérationnelles de la cohorte et recrutement d'un data manager supplémentaire pour l'analyse des données issues de l'appariement au SNDS	2,20
COBLANCE	3,51	Analyse des données appariées avec le SNDS Extractions et analyses moléculaires	1,65
CRYOSTEM	2,34	Maintien en conditions opérationnelles de la cohorte et recrutement de chefs de projet supplémentaires	1,32
E4N	24,9	Maintenance de la cohorte : 4 M€ Mise en place de la biobanque : 4,3 M€ Recrutements personnels : 12,3 M€	7,50
HOPE EPI	6,30	Préparation de l'appariement avec le SNDS / Consolidation, des composantes de l'infrastructure / Recrutement de statisticiens	3,10
HEPATHER	13,45	Coût de fonctionnement de la biobanque : 1,28 M€ Recrutements de data manager et biostatisticiens pour l'analyse des données issues de l'appariement au SNDS	2,90
MEMENTO	7,53	Maintien en conditions opérationnelles de la cohorte et coordination de l'étude : 2,1 M€ Suivi des patients inclus : 1 M€ Projet de réouverture de la cohorte : 1,2 M€	4,31
OFSEP	14,33	Consolidation du <i>data center</i> de la cohorte mère Mis en place de la nouvelle cohorte dite « haute définition »	5,45
PSY-COH	6,84	Achèvement des inclusions Mise en place de la biobanque	1,2
RADICO	6,90	Consolidation de l'hébergement des cohortes dans l'infrastructure / Poursuite des inclusions dans les cohortes Recrutement de personnels	3,5
RECONAI	11,89	Maintenance de la plateforme Financement de l'enquête auprès des 10,5 ans (2,4 M€) Recrutement de 3,5 ETP (<i>data managers</i>)	2,7
CONSTANCES	29	Renforts en compétences épidémiologie statistique et Ingénierie informatique et statistique (<i>data manager</i> , Exploitation statistique du SNDS). Développement des analyses génomiques	18

Source : Dossiers d'évaluation 2019.

Les motivations des demandes de financements présentées pour la période 2020-2024 confirment à cet égard que les cohortes s'inscrivent principalement dans des perspectives de consolidation correspondant, dans leurs grandes lignes, aux axes évoqués dans le paragraphe 3.1.2, et qu'elles recherchent surtout la stabilisation de leurs personnels, notamment pour des compétences clés qui reflètent leur position « intermédiaire » entre infrastructures et projets de recherche.

3.2.1 En termes de financement

Les dossiers font ainsi ressortir trois ensembles de besoins spécifiques, au-delà de la seule couverture des charges de fonctionnement récurrentes liées à la poursuite des inclusions ou des opérations de suivi :

- un besoin de consolidation ou de maintenance des infrastructures, le plus souvent sur le plan des systèmes d'information (RADICO, RECONAI) ;
- un besoin de complément visant, soit des appariements avec le SNDS (CANTO, COBLANCE, HOPE EPI), soit le développement des analyses génomiques (CONSTANCES, PSY-COH) ;
- enfin, dans certains cas, un besoin de financer des projets nouveaux (OFSEP-HD) ou des enquêtes prévues de longue date (RECONAI).

Certaines cohortes ont cependant également présenté des projets de réouverture de leurs inclusions (MEMENTO) ou d'élargissement de leur spectre (CANTO, CRYOSTEM), sur lesquels le jury d'évaluation a manifesté, à l'instar de la mission, des réticences ou de la circonspection.

Sous réserve d'un pilotage et de mutualisation renforcés, également évoqués par le jury et sauf à renoncer *de facto* aux fruits des investissements déjà réalisés, la mission considère qu'il importe de garantir pour l'ensemble des projets, sur les cinq ans à venir, un socle de ressources permettant de compléter les infrastructures créées. L'attribution de ces enveloppes devrait toutefois être conditionnée au fait que les projets avancent effectivement, en dépassant d'éventuels blocages constatés aujourd'hui (*cf.* point 3.1.1). À l'opposé, la perspective d'arrêter le financement de certains projets peut paraître problématique au vu des efforts déjà consentis et dans la mesure où, d'après les constatations de la mission, aucun d'entre eux n'a réellement failli.

3.2.2 En termes de stabilisation et de consolidation de certains personnels

Le besoin de stabilisation des personnels recrutés sur CDD et l'apport de compétences, notamment dans les métiers ayant trait à la gestion et l'exploitation des données, font par ailleurs l'objet de demandes de l'ensemble des responsables de projet.

Un des facteurs clés de la pérennisation des cohortes concerne en effet la stabilisation des personnels et le maintien de compétences liées aux politiques de données. Dans ce cadre, l'effort annoncé par l'Inserm, comme celui des universités qui, à l'instar de l'UVSQ (avec quatre recrutements en contrat à durée indéterminée pour CONSTANCES) ont accepté de mobiliser des emplois permanents, mériterait d'être accompagné, au-delà des seuls recrutements, d'une politique de formation commune, et, ainsi que l'a suggéré le jury de 2019, d'une large diffusion des savoir-faire : les expériences acquises par CONSTANCES en matière d'appariements avec le SNDS, par CRYOSTEM en matière d'accès aux données et échantillons, ou encore par RADICO en matière de gestion des systèmes d'information mériteraient incontestablement d'être davantage partagées, et cela de

manière moins informelle et plus organisée qu'actuellement, dans le seul cadre du club Cohortes d'AVIESAN (cf. point 2.5.3).

Par ailleurs, nombre des besoins en la matière sont mutualisables, ainsi que le montrent les réflexions du groupe de travail piloté par l'Inserm sur « *les métiers de cohortes* », qui constituent un préalable utile pour identifier, à partir d'un recensement des activités réalisées par les cohortes, leurs différentes fonctions clés (cf. encadré 7).

Encadré 7

Les principaux métiers des cohortes recensés par le groupe de travail de l'Inserm

- attachés de recherche clinique (ARC) pour la collecte et la saisie des données ;
- *data managers* ;
- qualitiens ;
- biostatisticiens ;
- chargés de communication (notamment à destination des participants) ;
- juristes ;
- coordinateurs techniques (spécialisés par thématique dans les grandes cohortes en population ;
- gestionnaires RH et budgétaires (en lien avec les délégations régionales, les universités et/ou les laboratoires de recherche de rattachement).

Source : Groupe de travail Inserm

La liste des métiers ainsi recensés montre que plusieurs d'entre eux (ARC, gestionnaires RH et budgétaires, juristes, communicants, qualitiens notamment) pourraient donner lieu à mutualisation, d'autant que ces fonctions mobilisent souvent moins d'un équivalent temps plein par cohorte, quand d'autres nécessitent en revanche une relation de proximité plus forte avec l'objet et la conduite scientifiques de la cohorte (coordination technique, métiers des données).

Associer, ainsi que l'a initié le projet France Cohortes, une stabilisation - *via* des emplois permanents - de compétences clés en termes de ressources humaines et une mutualisation des fonctions entre cohortes s'avère donc une démarche intéressante et porteuse. Pour la mission, elle gagnerait, d'une part, à être élargie à l'ensemble des parties prenantes (Inserm, universités concernées, agences sanitaires), de manière à pouvoir disposer d'un tableau consolidé des moyens humains qui fait aujourd'hui défaut, et, d'autre part, à associer pleinement les responsables des cohortes, de manière à bien circonscrire les activités qui doivent rester en proximité du projet et les tâches qui peuvent être pour tout ou partie mutualisées.

Il est donc nécessaire, aux yeux de la mission, que les parties prenantes aux cohortes s'engagent au plus vite sur un volume de financement et de personnels permanents permettant de tirer pleinement profit de l'effort financier engagé depuis une dizaine d'années.

3.3 Des besoins de clarification concomitants de l'environnement immédiat des cohortes

Un dernier volet indispensable d'une politique visant à tirer le meilleur parti des cohortes existantes consiste à procéder à des clarifications dans leur environnement immédiat, dans la mesure où responsables, chercheurs et financeurs ont besoin d'être au clair sur plusieurs points pour programmer et organiser au mieux la conduite concrète de leurs projets dans les prochaines années.

Le premier besoin d'éclaircissement porte sur les possibilités qui seront réellement offertes au fil du temps par les données médico-administratives, notamment en termes d'informations médicales enrichies. Il paraît d'abord important que, quels que soient les élargissements prévus du SNDS, les travaux conduits dans le cadre du réseau ReDSiam (réseau pour mieux utiliser les données du SNDS)⁷¹ se poursuivent pour progresser dans le repérage, la validité et la fiabilité de l'identification des pathologies à partir des recueils médico-administratifs actuels (*cf.* annexe 6). Concernant l'élargissement du SNDS lui-même, il serait utile, au-delà des projets pilotes qui viennent d'être lancés, de disposer rapidement de perspectives tangibles sur le contenu et le calendrier des incorporations de données cliniques qui seront effectivement opérées dans les quatre ou cinq ans à venir. Une étape majeure, évoquée depuis plusieurs années, est l'intégration des résultats des examens biologiques réalisés en ville : si elle continue à sembler possible aux responsables de l'assurance maladie à un horizon assez rapproché, elle nécessite une coopération et des investissements de la part des laboratoires d'analyse, s'agissant de l'homogénéisation du codage des données et la métrologie des équipements, dont il importe de s'assurer.

De même, la politique des registres de morbidité ou de patients qu'entend conduire le ministère chargé de la santé est un élément important de l'univers des cohortes, qui peuvent directement s'appuyer ou être couplées à ce type de registres (*cf.* annexe 4) ; elle mérite d'être redéfinie de façon claire et rapide, en aval des réflexions engagées au sein des agences sanitaires et par le Haut conseil de la santé publique (HCSP).

Enfin, les responsables de cohortes ont besoin, pour prendre leurs décisions, d'être éclairés au plus vite sur le rôle respectif et la répartition des fonctions qui s'opérera entre le projet France Cohortes, le *Health Data Hub* et la plateforme de l'Inca concernant un certain nombre de points clés :

- l'articulation technique et fonctionnelle de ces nouveaux systèmes d'information ;
- les lieux et structures destinés à réaliser les appariements entre les données de chaque cohorte et celles du SNDS, et les modalités de stockage sécurisé qui en résultent ;
- la nature et le coût des services qui leur seront proposés à cette occasion ;
- les conditions pratiques de transition entre les systèmes d'information déjà utilisés par les cohortes, qui présentent des caractéristiques fonctionnelles et techniques et des niveaux de maturité variables, et les nouveaux systèmes mutualisés qui doivent leur succéder ;
- l'alimentation en retour de ces différentes plateformes par les données des cohortes elles-mêmes, et les conséquences susceptibles d'en découler concernant l'ouverture et la mise à disposition de ces informations aux chercheurs extérieurs aux équipes productrices.

Au total, les actions de consolidation et de clarification à entreprendre pour tirer tout le parti possible des cohortes financées depuis plusieurs années au titre du PIA devraient selon la mission emprunter six directions.

⁷¹ Le réseau ReDSiam développe et met à disposition des algorithmes permettant d'identifier des pathologies à partir des 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. Il repose sur des groupes de travail transversaux ou par grands domaines pathologiques.

Recommandation n°6 S'assurer d'un mode de pilotage et de fonctionnement consolidé pour l'ensemble des cohortes :

- en initiant une redéfinition et/ou une clarification des modes de pilotage propres à certaines cohortes ;
- en demandant à plusieurs de ces cohortes d'améliorer le mode de fonctionnement de leurs instances de gouvernance, sans forcément en alourdir le pilotage.

cf. annexe 4

Recommandation n°7 Revoir les modalités d'évaluation des cohortes en organisant des visites de sites par des équipes d'évaluateurs d'origine à la fois scientifique et administrative.

Recommandation n°8 Tirer parti au plus vite des possibilités déjà offertes par le SNDS :

- en favorisant la réalisation rapide des appariements prévus avec le SNDS, au besoin en recourant à des conditions temporaires de stockage sécurisé des données évitant d'attendre la mise en place opérationnelle effective de France Cohortes ou du *Health data Hub* ;
- en poursuivant et amplifiant, dans le cadre d'une coordination *ad hoc*, les travaux initiés dans le cadre des projets ReDSiam afin de développer au plus vite l'identification robuste des pathologies dans le SNDS.

Recommandation n°9 Stabiliser un socle de financement public pluriannuel permettant aux cohortes ayant bénéficié de financements publics de tirer le meilleur parti des investissements réalisés :

- en mettant en place les analyses génomiques (ou autres *-omics*) permises par les prélèvements collectés ;
- en poursuivant les suivis et l'exploitation des informations recueillies jusqu'au terme de la période scientifiquement pertinente ;
- en employant un noyau significatif de personnels permanents.

Recommandation n°10 Étudier les conditions d'une possible mobilisation de données socio-économiques détenues par les services statistiques publics pour fiabiliser et enrichir les données des cohortes en population.

Recommandation n°11 Préciser rapidement à l'intention des responsables de cohortes les évolutions prochaines des bases de données médico-administratives et le rôle des plateformes d'appui qui leur seront proposées :

- en présentant aux responsables de cohortes un schéma clair des données actuellement et prochainement accessibles *via* les données médico-administratives, et précisant l'horizon réaliste de leur mise à disposition ;
- en clarifiant les fonctions et rôles précis qui seront exercés en matière d'appariements avec le SNDS, de stockage sécurisé des données, et d'ouverture aux chercheurs académiques respectivement par France Cohortes, l'Inca et le *Health Data Hub* ;
- en définissant et priorisant à l'intention des responsables de cohortes l'offre de services du futur dispositif France Cohortes, incluant le renforcement des fonctions juridiques, administratives et budgétaires et l'animation à visée scientifique ou de communication.

4 À moyen-long terme, la nécessité d'une politique plus ambitieuse et coordonnée de mobilisation des données longitudinales pour la recherche en santé

Au-delà des consolidations et clarifications nécessaires pour tirer le meilleur parti possible des cohortes existantes, la mission s'est interrogée sur les actions qui mériteraient d'être conduites, à plus long terme, pour développer une politique de données longitudinales à hauteur des besoins dans le champ sanitaire et social, où les cohortes viendraient s'inscrire de façon cohérente en considération de leurs apports propres.

L'horizon envisagé ici est d'environ dix à quinze ans (2030 et au-delà) dans la mesure où, à cette échéance, beaucoup de cohortes « patients » du PIA auront fini leur vie naturelle, et où, s'agissant des grandes cohortes en population, les enfants suivis par ELFE auront plus de vingt ans tandis que CONSTANCES, si elle n'est pas ré-ouverte, n'aura dans son échantillon que peu d'individus de moins de 40 ans.

Pour définir cette politique dont la mission pense, au vu des expériences internationales, qu'elle devrait avoir une ambition affirmée, une réflexion s'avère nécessaire très en amont, dans la mesure où elle devrait nécessairement reposer sur trois piliers : la nécessaire construction de priorités programmatiques et d'un cadre de coordination (point 4.1), la réponse spécifique à trois grands types de besoins (point 4.2) et la conduite parallèle d'actions d'accompagnement dans un cadre essentiellement mutualisé (point 4.3).

4.1 La nécessaire construction de priorités programmatiques et d'une coordination d'ensemble des outils d'observation

La définition d'une politique de données longitudinales donnant leur juste place aux cohortes épidémiologiques repose sur deux soubassements préalables : la définition en amont de priorités programmatiques reconnues et financées pour la recherche en santé (point 4.1.1) et la mise en place d'une coordination spécifique en matière d'outils d'observation (point 4.2.2).

Pour ces deux aspects, la mission estime que les pouvoirs publics devraient assumer la fonction d'impulsion stratégique qui leur revient en propre.

4.1.1 Le besoin en amont de priorités programmatiques pour la recherche en santé

Il n'entrait pas dans le champ de la mission de réfléchir à la programmation et à la coordination d'ensemble de la recherche en santé publique qui a donné lieu dans les années récentes à plusieurs rapports portant soit sur l'organisation globale d'un plan d'action⁷² soit sur des volets spécifiques comme la recherche clinique⁷³. Dans l'un et l'autre cas, ces rapports préconisaient l'élaboration d'une stratégie explicite et coordonnée comprenant la définition de priorités thématiques s'articulant avec les stratégies nationales de santé et de recherche (SNS et SNR) ; ces priorités étaient appelées à se traduire dans un exercice de programmation ambitieux, doté de moyens suffisants et associant

⁷² Aviesan, Plan d'action pour l'organisation de la programmation de la recherche en santé, mai-décembre 2014.

⁷³ Christine d'Autume, Stéphane Elshoud, Emilie-Pauline Gallié, Laurent Gratioux, Armand Renucci, *op. cit.*

l'ensemble des acteurs de la recherche et de la santé, de façon à dépasser le cloisonnement actuel des politiques de soutien à la recherche menées par chacun d'entre eux.

Sans ré-ouvrir cette question, la mission a considéré que l'élaboration d'un tel programme était un préalable indispensable à la définition d'une politique spécifique concernant les cohortes. En effet, l'une des limites importantes du PIA est d'être parti de projets proposés par des équipes de recherche sans référence à des priorités politiques reconnues par les acteurs, ce qui a favorisé une multiplicité non coordonnée d'autres dispositifs ou infrastructures portés par chacun d'eux en toute « légitimité ». Pour jouer le rôle attendu, ce programme de recherche en santé publique devrait donner lieu à une impulsion à très haut niveau et être doté d'un budget suffisant, établi pour plusieurs années, et comportant une fraction conséquente consacrée au développement des outils d'observation⁷⁴.

Il devrait, à cet égard, être assorti d'une cartographie en amont des outils d'observation déjà disponibles pour les principales priorités thématiques et de ceux qui seraient nécessaires pour les éclairer dans la durée, sur la base d'une réflexion prospective. Une telle cartographie, assortie de propositions pour améliorer l'articulation et la complétude des sources, a déjà été réalisée pour certains domaines d'observation comme le suivi des médicaments en vie réelle⁷⁵ et, sous l'égide du Conseil de l'enfance et de l'adolescence et du Haut Conseil à la santé publique (HCSP), la connaissance statistique de la santé et du bien-être des enfants⁷⁶.

Les dispositifs de suivi longitudinal sont appelés à prendre une place essentielle dans cette démarche, sachant que, en leur sein, les besoins spécifiques en cohortes devraient être établis en fonction des informations déjà disponibles dans les bases de données administratives (*Health data Hub* en fonction de son évolution à l'horizon de dix ans, registres dont l'utilité a été reconnue suite aux évaluations et confirmée, données socio-fiscales) et des manques repérés en la matière, s'agissant notamment de pathologies difficiles à identifier ou de données cliniques, biologiques et génomiques.

La consultation de la société civile (Conférence nationale de santé, Conseil national de l'information statistique) ainsi que du Parlement serait dans ce cadre, et au vu des enjeux éthiques soulevés par les évolutions récentes (monétisation des données de santé, médecine prédictive), une étape indispensable, non seulement pour l'élaboration du programme d'ensemble de recherche en santé, mais aussi de son volet concernant spécifiquement les outils d'observation.

4.1.2 La nécessité en aval d'une coordination spécifique en matière d'outils d'observation, assise sur une gouvernance adaptée et des mécanismes d'évaluation

Une fois les choix de principe effectués quant au développement et au financement de nouveaux outils d'observation, un pilotage spécifique devrait être institué concernant la mise en œuvre de ces décisions, et la construction concrète de dispositifs partagés, au premier rang desquels les cohortes.

⁷⁴ Certains interlocuteurs entendus par la mission ont envisagé la possibilité que ce budget soit porté par les crédits de l'assurance maladie, comme celui des grandes agences sanitaires comme l'ABM, l'ANSM, Santé publique France, et, à terme, l'Inca.

⁷⁵ 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.

⁷⁶ 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

Ce pilotage devrait associer l'ensemble des parties prenantes (ministères concernés, système statistique public, organismes et milieux de recherche, caisses de sécurité sociale, agences sanitaires, fédérations d'établissements, grandes fondations...), de façon à ce que leurs besoins puissent précisément être pris en compte, mais aussi qu'ils puissent contribuer concrètement à la mise en place des projets par la voie de soutiens en compétences statistiques, informatiques ou logistiques.

Un mode d'interaction apparaît à cet égard à trouver entre les besoins et priorités formulés par les acteurs et les propositions émanant des équipes de recherche qui exerceront la responsabilité scientifique des projets. L'exigence de dispositifs d'observation puissants et partagés implique en effet un processus de co-construction d'outils qui auront vocation à irriguer de façon large le monde de l'expertise et de la recherche. L'expérience de la cohorte Autisme en cours de définition, et qui, décidée sur le fondement d'une priorité politique affichée, a donné lieu à une note de cadrage rédigée par des chercheurs servant de base à un appel à manifestation d'intérêt, est un exemple dont il faudra tirer les leçons, quant à la pertinence et la possibilité d'étendre une telle méthode, qui paraît très intéressante quant à son principe.

Dans le cadre de ce pilotage partagé, une évaluation régulière et systématique des projets devrait en outre être prévue, sur la base de séances de travail approfondies entre leurs responsables et des experts de plusieurs origines, scientifiques, statistiques et/ou administratives.

4.2 La nécessaire prise en compte de trois grands types de besoins

Il reviendra au programme national de recherche en santé publique et aux instances de coordination mentionnées au paragraphe précédent de définir précisément les projets et les financements à consacrer au développement des outils d'observation jugés nécessaires, en réservant parmi eux une juste place aux cohortes épidémiologiques, considérées dans leur apport spécifique. La mission souhaite néanmoins d'ores et déjà indiquer que, s'agissant de ces cohortes, trois grands besoins devront être pris en compte, sur lesquels un groupe de travail pourrait sans attendre ouvrir une réflexion commune.

4.2.1 L'installation à l'horizon de dix ans d'une à deux grandes cohortes en population, dotées de plusieurs caractéristiques

Un premier besoin concerne les cohortes généralistes en population qui, si l'on se réfère à l'exemple britannique, ont un intérêt avéré pour les recherches sur le rôle respectif et les interactions entre facteurs biologiques, génétiques, sociaux et environnementaux dans la survenue des problèmes de santé. Au plan des politiques publiques, de telles infrastructures sont utiles pour éclairer d'une part l'impact spécifique des inégalités sociales de santé et d'autre part les politiques de prévention à envisager, au regard de l'importance relative des différents facteurs de risques et/ou expositions.

La mission pense qu'il serait important que la France dispose en permanence de deux grandes infrastructures de ce type :

- une cohorte d'enfants à la naissance, renouvelée régulièrement, et dont les contours seraient à redéfinir en fonction de l'évaluation des apports et des limites de l'étude ELFE sur longue période (cf. annexe 4) ;
- une cohorte d'adultes à large spectre, visant de l'ordre de 500 000 participants, généralisant et améliorant l'expérience réalisée avec le dispositif CONSTANCES.

Il reviendra aux instances appropriées de décider, après expertise approfondie, si une telle cohorte peut ou non prendre pour point de départ et/ou englober les personnes actuellement suivies dans le cadre des études CONSTANCES et COSET et, le cas échéant ELFE et E3N. Il leur faudra également déterminer si elle doit rechercher une représentativité statistique permanente grâce à la ré-ouverture régulière de ses inclusions, ou s'il vaudrait mieux procéder, à l'instar des Britanniques, au lancement périodique d'études renouvelées.

En tout état de cause, le dispositif envisagé à moyen-long terme devrait dépasser les limites actuelles de la cohorte CONSTANCES qui couvre exclusivement les assurés du régime général acceptant d'être suivis par des centres d'examen de santé, n'a pas à ce stade intégré d'analyses génomiques et recueille en propre des informations sur les revenus des ménages participants sans les rapprocher des données socio-fiscales rassemblées et traitées par l'Insee (*cf.* annexe 4).

Le dispositif à concevoir à long terme devrait à cet égard :

- viser une représentativité statistique satisfaisante et régulièrement attestée, en prévoyant les ré-ouvertures, suivis « passifs » ou renouvellements nécessaires ;
- pouvoir jouer un rôle potentiel de « cohorte témoin » ou d'accueil de « cohortes nichées » pour un large ensemble d'études d'ampleur plus réduite ;
- développer le recours, après évaluation des expériences conduites par Nutrinet et E4N, aux technologies digitales et à Internet pour le recrutement et le suivi des participants, afin de limiter les coûts et de réduire les déperditions liées aux phénomènes d'attrition ;
- être conçu, quant au contenu des informations collectées, en articulation étroite avec les systèmes d'information médico-administratifs (en tenant compte des informations accessibles à cette date *via* le *Health data Hub*) et, après une procédure spécifique vis-à-vis de la CNIL, ceux concernant les trajectoires professionnelles et les revenus sociaux et fiscaux ;
- posséder une variété de ressources en termes de résultats d'examens cliniques, prélèvements biologiques, d'imagerie, de génétique et de *-omics* ; l'intégration des pistes de recherche relatives à la génomique, à l'épigénétique, au métabolisme ou à la microbiologie, ainsi qu'aux données d'imagerie ou aux technologies émergentes apparaît en effet, au regard des expériences internationales, incontournable pour un tel équipement, sur des sous-échantillons de taille conséquente (*cf.* annexes 7 et 8) ;
- permettre une identification la plus fine possible des expositions environnementales des participants à la cohorte, en combinant des mécanismes de géolocalisation de leurs lieux de scolarité, de travail et de résidence et des questionnaires spécifiques sur leurs modes de vie ; les effets des expositions environnementales (pesticides, perturbateurs endocriniens, qualité de l'air...) sur la santé, en association ou non avec d'autres facteurs, sont à cet égard considérées, tant par l'Anses que par plusieurs rapports de l'Igas, comme donnant lieu à des besoins majeurs de suivi longitudinal dans la durée, ce qui pourrait inviter, comme dans le projet Cosmos⁷⁷, à coupler cohorte généraliste et cohortes spécifiques.

Les deux cohortes ainsi constituées ont vocation à être considérées comme des infrastructures de recherche à portée nationale, et gérées comme telles. Cette gestion ne devrait donc pas dépendre d'une équipe ou d'un centre de recherche particuliers, mais d'un groupement scientifico-

⁷⁷ Étude prospective sur la téléphonie mobile et la santé « nichée » dans la cohorte CONSTANCES *cf.* supra, point 1.3. encadré2.

administratif créé à cet effet, doté de la personnalité morale et associant à son pilotage l'ensemble des parties prenantes et des utilisateurs potentiels.

Ce pilotage devrait en outre pouvoir s'appuyer sur un conseil scientifique pluridisciplinaire, un comité de déontologie et une structure consultative des représentants des patients.

Dans ce cadre, une offre de services devrait être prévue en direction des utilisateurs de façon à faciliter le développement de projets de recherche ou d'études longitudinales pouvant être « adossées » à ces grandes cohortes, elles-mêmes appariées aux bases administratives. Le coût de ces services et accès devrait être pour l'essentiel pris en compte dans le budget de fonctionnement de ces équipements, sans refacturation systématique aux différents projets, tout du moins s'agissant des chercheurs académiques et des partenaires publics.

4.2.2 La définition de priorités et de modes d'appui adaptés aux dispositifs de suivi longitudinaux centrés sur des pathologies ou des patients traités

Il est loisible de s'interroger dans ce nouveau contexte sur la place que devraient à l'avenir occuper les cohortes de patients exposés, atteints ou traités pour des pathologies spécifiques, en lien avec les registres dont le maintien aura été décidé.

Sans vouloir une nouvelle fois préempter les travaux des instances de programmation mentionnées au point 4.1, la mission estime que le maintien d'un portage et d'un financement dédiés à ce type de projets restera indispensable à l'avenir, même s'il convient de les prioriser et d'en examiner soigneusement le contenu au regard de l'ensemble des informations disponibles, et notamment de l'enrichissement progressif des bases de données médico-administratives (repérage plus sûr et validé des pathologies, informations issues des dossiers hospitaliers et des réunions de concertation pluridisciplinaires, résultats des analyses biologiques et des examens radiologiques de prévention...).

Des besoins de dispositifs longitudinaux complémentaires, et notamment de cohortes, semblent à cet égard devoir persister, pour répondre à plusieurs objectifs :

- assurer un suivi régulier et systématique de la prise en charge des patients atteints de maladies chroniques, assorti de questionnaires sur leurs modes d'accès à la prévention et aux soins, leur satisfaction vis-à-vis de ces prises en charge et leurs problèmes de réinsertion sociale et /ou professionnelle dans la durée ;
- asseoir la présence française à l'international à travers une politique scientifique de « niches » sur des pathologies pour lesquelles les chercheurs français ont une place particulièrement reconnue ;
- suivre sur plusieurs années les répercussions sanitaires d'éventuelles catastrophes environnementales ou industrielles, telles celles qui ont concerné l'usine AZF à Toulouse ou l'entreprise Lubrizol à Rouen ;
- mettre en place, au-delà de la programmation générale du suivi des médicaments en vie réelle⁷⁸, un suivi particulier des traitements innovants susceptibles d'apparaître dans des horizons très courts, comme, dans la période actuelle, les thérapies géniques développées en onco-hématologie (CAR-T), dont le coût est par ailleurs très élevé ; la HAS et l'Inserm ont ainsi demandé la mise en œuvre d'une telle approche coordonnée pour les médicaments de

⁷⁸ Bernard Bégau, Dominique Polton et Franck von Lennep, op.cit.

thérapies innovantes (MTI), alors que plusieurs initiatives ont été lancées de façon dispersée ; d'autres types de traitements peuvent aussi nécessiter le déclenchement rapide de ce type de suivi, par exemple quant à la pertinence, la qualité et l'efficacité de la chirurgie en matière d'obésité⁷⁹.

Il importe donc que la programmation des dispositifs d'études longitudinales envisagée ci-dessus laisse une place à ce type de cohortes plus spécifiques et ponctuelles, et ce avec une certaine souplesse, même si une priorisation et une vigilance quant aux risques de redondance des informations collectées doivent sans conteste être diligentées.

La définition d'une enveloppe et d'un circuit financiers spécifiques à ces cohortes est d'autant plus nécessaire que les orientations actuelles du programme hospitalier de recherche clinique (PHRC) n'en permettent plus le financement⁸⁰. Il en est de même, hors PIA, pour les appels à projets de l'Agence nationale de la recherche qui, sauf exception, sont limités dans leur montant, consacrés à des projets de recherche de l'ordre de deux à trois ans, et donc le plus souvent réalisés à partir d'outils d'observation déjà existants.

Alors que ces cohortes sont, sauf cas particulier, plutôt destinées à rester pilotées par des équipes et laboratoires de recherche existants, une solution de portage administratif, financier et technique de certaines d'entre elles par les agences sanitaires mériterait en outre d'être considérée. Elle permettrait d'en conforter l'assise, comme cela a été le cas pour les cohortes développées par Santé publique France dans le domaine de la santé au travail, des cohortes de l'ANRS sur le SIDA et les hépatites, ou de la cohorte CKD Rein, fortement épaulée par l'Agence de la biomédecine (ABM).

4.2.3 Des besoins de suivis longitudinaux à long terme, pouvant ou non prendre la forme de cohortes, dans d'autres domaines que la seule épidémiologie

Une politique d'ensemble concernant les dispositifs de suivi longitudinal ne devrait pas se limiter au champ de l'épidémiologie et de la santé publique, mais s'étendre à des domaines connexes comme l'organisation et la régulation du système de santé ou les politiques sociales et médico-sociales.

4.2.3.1 Dans le domaine de la santé, des besoins concernant l'organisation des parcours de soins et la couverture sociale en matière de santé

En matière de régulation du système de santé, les besoins de suivi longitudinal portent d'abord sur l'organisation concrète des parcours de soins et leur suivi « en vie réelle » : ils concernent notamment l'évaluation des « parcours » ciblés par la stratégie « Ma santé 2022 », la recherche sur le rôle et l'efficacité des soins primaires et la mise en lumière de secteurs particuliers, parfois en situation critique, comme la psychiatrie et la pédo-psychiatrie.

Nombre de ces besoins peuvent recevoir des réponses à travers l'usage des bases de données médico-administratives, mais des enquêtes ou suivis longitudinaux complémentaires peuvent s'avérer utiles, notamment pour recueillir le point de vue des personnes (et, le cas échéant, de leur médecin traitant) sur leur orientation au sein du système de santé et leurs conditions de prise en charge.

⁷⁹ Julien Emmanuelli, Vincent Maymil et Pierre Naves, *Situation de la chirurgie de l'obésité*, Rapport Igas, janvier 2018.

⁸⁰ Christine d'Autume, Stéphane Elshoud, Emilie-Pauline Gallié, Laurent Gratieux, Armand Renucci, *Évaluation du programme hospitalier de recherche clinique (PHRC)*, op.cit.

Par ailleurs, le SNDS ne comporte pas de données sur les couvertures maladie complémentaires des assurés et les restes à charge qu'ils ont à supporter, dont on sait qu'ils sont un déterminant important des consommations de soins, et ce alors même que l'enquête *European Health Interview Survey (EHIS)* est beaucoup moins précise sur ce point que l'ancienne Santé et protection sociale de l'Irdes. La prise en compte de ce sujet est donc un aspect à ne pas négliger dans les suivis longitudinaux à mettre en place, d'autant que des réformes importantes, dont l'impact devra être suivi et évalué, sont récemment intervenues en la matière (complémentaire santé solidaire, 100% Santé).

4.2.3.2 Dans les domaines sociaux et médico-sociaux, des besoins concernant les enfants et certaines populations en difficulté

Les investigations de la mission ont principalement porté sur les sujets sanitaires et elle n'a effectué que des incursions limitées dans les domaines proprement sociaux, à l'exception de ceux qui sont le plus en interaction avec les problèmes de santé, comme les conséquences du chômage et de la pauvreté, la prise en charge du handicap et du vieillissement ou les conditions de développement des enfants et des adolescents.

La mission n'a pas, dans ces domaines, pu prendre en compte l'ensemble des travaux de recherche disponibles ni les enseignements des expériences internationales, mais elle s'est appuyée sur des rapports et avis comme ceux du Conseil de l'enfance et de l'adolescence et du Haut Conseil à la santé publique (HCSP) sur le suivi statistique des enfants⁸¹, qui est une thématique essentielle justifiant des études longitudinales sur longue période pour appréhender les processus de formation des inégalités et les interventions clés permettant d'y remédier. Elle a également interrogé les membres de l'Igas sur les préconisations faites en ce sens dans leurs rapports récents ou des personnalités qualifiées.

Sans viser à l'exhaustivité, ces analyses pointent notamment le besoin d'instruments d'observation longitudinale concernant :

- les enfants et adolescents pris en charge au titre d'un handicap ou de troubles psychiques, en tenant compte des données déjà collectées par le système d'information des maisons départementales des personnes handicapées (MDPH) ;
- les enfants et adolescents concernés par des actions de protection de l'enfance, qui devraient à la fois donner lieu à la réalisation d'études sur la maltraitance et les violences subies à partir des cohortes existantes, à l'ouverture aux chercheurs de données départementales améliorées et appariées aux autres bases, et à un suivi rétrospectif et prospectif à long terme concernant un ou plusieurs sous-échantillons d'entre eux ;
- l'impact des expérimentations conduites en matière de lutte contre le chômage de longue durée sur le bien-être et la santé psychique des bénéficiaires ;
- les migrants confrontés à des problématiques simultanées d'accès aux droits, d'insertion sociale et de santé.

⁸¹Conseil 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.

4.3 Des besoins d'accompagnement à prendre en compte de façon parallèle et essentiellement mutualisée

Le troisième pilier indispensable à moyen-long terme d'une politique de données longitudinales en santé concerne l'accompagnement des responsables scientifiques de ces dispositifs et des chercheurs désireux d'y accéder pour des projets de recherche ou des études venant s'y adosser.

Les préconisations de fond de la mission diffèrent peu sur ce point de celles émises dans la deuxième partie du rapport à propos de la consolidation de l'environnement des cohortes actuelles, mais elles tendent à les généraliser et à les systématiser de manière à inscrire les projets dans un cadre prédéterminé et connu de tous et à assurer *ex ante* l'accompagnement nécessaire à leur déroulement.

Cet accompagnement devrait porter sur trois axes : les infrastructures « support » que constituent les biobanques et la structuration des bases de données issues d'appariements avec les données administratives ; les besoins en appuis juridique, budgétaire, scientifique et de communication ; la définition des modes de partenariat avec le monde économique.

4.3.1 Des infrastructures « support » à mettre en place pour l'ensemble des dispositifs

Le premier type d'appui à mettre en place de façon généralisée à l'intention de l'ensemble des responsables de dispositifs de suivi longitudinal concerne les appariements avec les bases de données administratives et/ou médico-administratives et le stockage sécurisé des données ainsi constituées. Des initiatives sont, comme on l'a vu, en cours en ce sens dans le cadre du *Health data Hub* et du projet France Cohortes, mais elles n'en sont qu'à leurs prémises, et surtout, sont pour l'heure loin d'être à même d'offrir un cadre organisationnel, juridique et technique homogène susceptible de répondre aux besoins des cohortes, relevant ou non de l'Inserm, dès le stade de leur lancement.

Un tel cadre « mutualisé » devrait à cet égard comporter :

- la définition d'une offre de services assurant notamment un appariement facilité avec les bases de données administratives en santé comme fiscal-social (Insee, CNAF,...) ;
- l'accès à des modalités de stockage des données offrant toutes les garanties nécessaires en termes de sécurisation, mais aussi d'accessibilité aux chercheurs habilités, sur la base de principes dits *de FAIR Data access*⁸²;

Un tel accompagnement devrait s'appuyer sur des compétences adéquates dans les domaines du *data management* et des biostatistiques et suppose un effort particulier en matière de formation, tant initiale que continue.

Ce cadre mutualisé devrait être assorti d'une mise à disposition des données des communautés scientifiques dans le cadre d'une politique d'accès ouvert (*open access*) transparente, passant par une simplification et une harmonisation des règles d'accès et de partage des données et par une facilitation des procédures administratives y afférant : cela impliquera pour chaque cohorte de

⁸² Ces principes consistent à rendre les données repérables, accessibles, interopérables et réutilisables (*Fair : findable, accessible, interoperable and reusable*)

renoncer à définir comme aujourd'hui sa propre procédure en la matière, s'ajoutant aux demandes et circuits spécifiques concernant l'accès aux données appariées.

Les conditions financières d'accès aux bases de données ainsi constituées devraient dans ce cadre également être précisées, à la fois pour les partenaires privés (*cf.* infra point 4.3.3) et pour les chercheurs académiques. Pour ces derniers, la mission préconise un accès gratuit à ces services, ou une participation financière limitée au coût spécifique de la prestation réalisée. Cela implique que l'essentiel des coûts de fonctionnement des infrastructures d'appariement et de stockage soit pris en compte dans leur propre budget de fonctionnement, sans être, comme c'est parfois le cas aujourd'hui, refacturé à chaque projet de recherche, lui-même financé sur fonds publics.

Un deuxième cadre d'infrastructure indispensable à construire concerne les biobanques destinées à la recherche en santé. Il s'agit d'un élément très important pour les dispositifs longitudinaux, qui, par nature, impliquent la conservation de prélèvements biologiques nombreux et divers, et ce sur des périodes longues. Contrairement à la question des données, ce besoin n'a semble-t-il pas été pris à bras le corps ni dans le cadre du PIA, en dépit des intentions initialement affichées (*cf.* point 1.4.2.1), ni dans la définition en cours du projet France Cohortes (*cf.* point 2.5.4), à l'exception d'une réflexion sur la « qualification » des collections biologiques aux normes européennes et sur la « certification » des centres qui les accueillent.

Or, il importe à moyen-long terme de définir une politique de stockage des collections massives ainsi constituées : cette politique devrait viser à offrir des infrastructures d'accueil situées sur le territoire français et relevant si possible d'un cadre de gestion public, afin d'éviter les risques liés aux possibles changements de stratégie économique de partenaires privés n'étant pas forcément à même de s'engager sur le très long terme.

4.3.2 Le renforcement de l'appui aux porteurs de cohortes en matière juridique, budgétaire, de communication et d'animation scientifique

Par ailleurs, les fonctions de support et d'appui, souvent négligées, constituent, un deuxième volet indispensable d'une politique d'accompagnement à destination des responsables scientifiques des dispositifs de suivi longitudinal, qui suppose :

- un renforcement des fonctions de support pour la gestion des ressources – humaines, financières et informatiques – nécessaires au bon fonctionnement de chacun des dispositifs ;
- le développement des fonctions d'appui aux porteurs de cohortes en termes de communication, à la fois envers les participants recrutés et suivis sur la base du volontariat, les associations de patients et, le cas échéant, le grand public ;
- l'entretien d'un réseau d'animation scientifique permettant un dialogue entre les différentes cohortes, par exemple concernant l'usage des bases de données administratives (données de santé et données sociales), le repérage affiné et validé des diverses pathologies, les techniques de géolocalisation et d'acquisition des données environnementales ou même, dans certains cas, les algorithmes de modélisation complexe.

4.3.3 Un pilotage plus cohérent et harmonisé des relations avec les industriels

La définition d'une politique partenariale en direction du monde économique constitue le troisième axe d'un accompagnement à harmoniser et professionnaliser. Compte tenu des enjeux éthiques et de crédibilité scientifique, celle-ci ne peut dépendre exclusivement des porteurs du projet, ni même d'une structure de valorisation, mais incombe pleinement aux pouvoirs publics.

À cet égard, la mission s'est étonnée de voir Inserm Transfert, l'ANSM et la HAS adopter des positions différentes concernant les partenariats développés entre les différentes cohortes, notamment de patients, et les laboratoires pharmaceutiques, alors même que ces cohortes, largement financées sur fonds publics, ont vocation à être utilisées par l'ensemble des acteurs publics du monde de la recherche et de l'expertise en santé.

Une doctrine commune, en vue d'une maîtrise collective permettant une utilisation transparente et partagée semble donc devoir être discutée et élaborée par ces organismes et autorités, sous l'égide des deux ministères concernés.

Par ailleurs, au-delà des seules industries pharmaceutiques, des discussions pourraient être initiées avec d'autres secteurs économiques (par exemple les mutuelles ou des nouveaux acteurs du marché des *datas*), ce qui permettrait de diversifier ces partenariats, toujours dans un cadre dont la maîtrise devrait être étroitement encadrée par les pouvoirs publics.

Au total, la mission pense que la définition d'une politique plus ambitieuse de mobilisation des données longitudinales pour la recherche en santé passe par neuf orientations complémentaires.

Recommandation n°12 Définir à l'horizon de cinq ans des priorités s'inscrivant dans les stratégies nationales de santé et de recherche et un programme à moyen-long terme insérant les dispositifs de suivi longitudinaux dans une politique ambitieuse de recherche en santé et recensant, par grande thématique, les outils d'observation disponibles et nécessaires.

Recommandation n°13 Associer les parlementaires et les représentants de la société civile à l'élaboration et au choix de ces priorités de programmation.

Recommandation n°14 Assurer un pilotage partenarial des outils d'observation relevant des priorités ainsi définies :

- en mettant en place une instance opérationnelle de coordination des dispositifs d'observation et de suivi longitudinal en santé associant les ministères concernés, les organismes de recherche, le système statistique public, les agences sanitaires et les caisses de sécurité sociale ;
- en expérimentant des modes d'appels à projets articulant des priorités thématiques fixées par les pouvoirs publics et des recommandations de méthode issues du monde de la recherche ;
- en évaluant régulièrement les dispositifs ou infrastructures ayant fait l'objet d'un soutien ;
- en veillant à l'adaptation des projets en fonction de l'évolution des techniques ou des données disponibles.

Recommandation n°15 Remettre en place à l’horizon de dix ans une à deux grandes cohortes en population générale :

- en programmant sur le long terme l’existence, de façon complémentaire aux bases de données médico-administratives, d’une grande cohorte d’adultes et d’une cohorte d’enfants à la naissance ;
- en donnant à ces grandes cohortes une ambition affirmée, en visant une représentativité en population générale, une articulation avec les systèmes médico administratifs et une variété de ressources en termes de prélèvements biologiques, d’imagerie, de génétique et de *-omics* ;
- en prévoyant pour ces cohortes un mode de pilotage spécifique reconnaissant leur fonction d’infrastructure de recherche à portée nationale.

Recommandation n°16 Définir des priorités et des modes opératoires concernant les dispositifs de suivi longitudinal à lancer concernant des pathologies ou des traitements spécifiques :

- en incluant dans ces priorités le suivi de certaines maladies chroniques, des traitements innovants et coûteux ainsi que des domaines de « niche » où la recherche française est en position d’excellence ;
- en prévoyant pour les cohortes relevant de ces priorités des circuits de financement spécifiques.

Recommandation n°17 Intégrer aux priorités nationales concernant les besoins d’observation et de suivi longitudinal à long terme :

- dans le domaine de la santé, les questions d’organisation des parcours de soins et de couverture des frais de santé ;
- dans les domaines social et médico-social, le devenir des enfants handicapés et de ceux pris en charge par l’aide sociale à l’enfance.

Recommandation n°18 Assurer aux dispositifs longitudinaux prioritaires un environnement et des mécanismes d’accompagnement adaptés :

- en entreprenant une structuration au niveau national des biobanques destinées à la recherche sur le territoire français ;
- en consolidant et développant l’appui aux responsables de cohortes en matière juridique, budgétaire, et de communication ;
- en soutenant, dans les domaines de l’épidémiologie et de la santé publique, l’attractivité des métiers dans les domaines *du data management* et des biostatistiques.
- en mettant en place, *via* le *Health data Hub* ou le CASD, des processus d’appariement systématique entre les données des cohortes épidémiologiques, notamment en population, et un vaste champ de données administratives, non seulement en santé, mais aussi médico-et fiscalosociales (Insee, CNAF...).

Recommandation n°19 Clarifier et faciliter les processus d'accès des chercheurs académiques aux données des études épidémiologiques financées sur fonds publics :

- en définissant des règles homogènes d'accès aux données valables pour l'ensemble des sources et des dispositifs ;
- en prévoyant soit un accès gratuit à ces infrastructures, soit une participation financière limitée au coût de la prestation de service réalisée pour chaque projet.

Recommandation n°20 Établir une doctrine claire, publique et partagée entre organismes de recherche et autorités sanitaires (HAS, ANSM) concernant les partenariats entre firmes industrielles et dispositifs de suivi longitudinal financés sur fonds publics, et s'appliquant quel qu'en soit l'organisme porteur.

LETTRE DE MISSION



MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION

Les Directeurs de Cabinet

Paris, le **30 AVR. 2019**

Madame la cheffe de service, Monsieur le chef de service,

Les cohortes constituent un des instruments de référence pour la recherche épidémiologique et ont une utilité majeure pour la recherche en santé publique, la recherche biomédicale et la décision publique.

Certaines des 267 cohortes répertoriées dans le portail « Epidémiologie France » occupent une place particulière au niveau national, voire international : il s'agit des 13 cohortes sélectionnées par le PIA en 2012, auxquelles s'ajoute la cohorte « Memento », de même ampleur et ambition mais financée par le plan Alzheimer et la direction générale de la recherche et de l'innovation (DGRI).

L'évaluation des cohortes « Constances », « CKDRein », « E4N », « ReCoNai » et « Hépathér » réalisée début 2017 a conclu que les objectifs avaient été atteints, et a reconnu leur grande qualité scientifique. Le succès de ces cohortes ayant été constaté, une réflexion est nécessaire pour garantir leur fonctionnement après la fin du financement initial du PIA, en 2019.

Ceci est d'autant plus important que, du fait de leur nature longitudinale, la valeur scientifique de ces cohortes s'accroît avec la durée de suivi des personnes. Les possibilités d'études complémentaires, de participation à des programmes européens et internationaux, de valorisation des données et de partenariats industriels augmentent de façon exponentielle au cours du temps après constitution de chaque cohorte.

Pour pouvoir être pérennisées, ces cohortes doivent toutefois bénéficier d'un certain nombre d'aménagements :

- un nouveau projet scientifique et mode de fonctionnement objectivant les nouveaux besoins scientifiques et s'adaptant à un suivi dans la durée ;

.../...

Madame Nathalie DESTAIS

Cheffe de l'Inspection générale des affaires sociales
39-43, quai André Citroën
75739 Paris Cedex 15

Monsieur Jean-Richard CYTERMANN

Chef de service de l'IGAENR
31-35, rue de la Fédération
75015 Paris

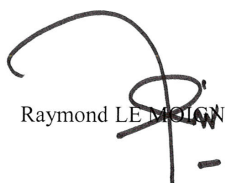
- un financement poursuivi après 2019 afin de maintenir leur activité dans le cadre de leur nouveau projet scientifique et de fonctionnement ;
- un modèle économique permettant de valoriser les données et de développer des partenariats industriels ;
- un statut et une gouvernance appropriés pour ces projets à très longue durée de vie : alors que les dysfonctionnements observés liés à la complexité du montage institutionnel fragilisent la réalisation des objectifs de ces cohortes, voire hypothèquent leur existence même, la définition d'un statut et de modalités de gestion garantissant un fonctionnement plus agile et facilitant les coopérations public-privé à long terme est nécessaire ; la réflexion sur ce sujet devra inclure la question de la place des différentes tutelles, de leur coordination, et de leur implication dans la gestion prospective des emplois et des compétences ;
- la mutualisation de certaines activités ou compétences au-delà des actions conduites par l'ITMO Santé publique d'AVIESAN : mise en conformité pour répondre aux exigences du RGPD et du référentiel de sécurité du système national des données de santé (SNDS), partage d'outils d'exploitation du SNDS, bio-banques, mobilisation de compétences particulières (en bio-informatique, en épidémiologie, en droit).

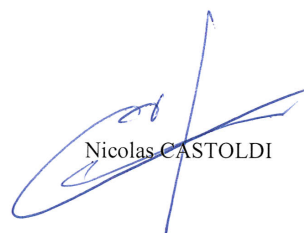
Nous souhaitons qu'une mission conjointement confiée à l'IGAS et à l'IGAENR puisse conduire une analyse du fonctionnement actuel de ces instruments de recherche et, dans un objectif de pérennisation des cohortes au-delà de 10 ans, formuler des propositions sur les quatre sujets prioritaires suivants :

- la stratégie nationale poursuivie en matière de cohortes (quelles cohortes, en réponse à quels besoins, avec quelles priorités) ;
- le pilotage et la gouvernance de ces outils de recherche (qui fait quoi, pour quels résultats attendus, dans quel cadre de suivi et d'évaluation, avec quelles perspectives de mutualisation, et quel positionnement des tutelles) ;
- leur cadre de gestion (quel statut juridique, quel cadre de mutualisation de moyens ou de compétences, quelle organisation des partenariats) ;
- leur modèle économique (quelles règles et conditions de financement, quels partenariats public-privé, quelles modalités de valorisation des données tant pour les acteurs publics que pour leurs partenaires privés).

La mission s'articulera avec les autres travaux en cours sur ces sujets et pourra élargir sa réflexion aux cohortes ne relevant pas de financements du PIA ou mises en place dans le domaine social.

Votre rapport est attendu pour le 1^{er} octobre 2019.


Raymond LE MOUËL


Nicolas CASTOLDI

LISTE DES ANNEXES

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

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

Annexe 3 : Panorama général des cohortes soutenues dans le cadre du PIA

Annexe 4 : Analyse des 14 cohortes soutenues au travers du PIA

- Les cohortes en population
 - La cohorte CONSTANCES
 - La cohorte E3N/E4N
 - Les cohortes RECONAI (ELFE et EPIPAGE2)
- Les cohortes et plateformes de patients
 - La cohorte CANTO
 - La cohorte CKD REIN
 - La cohorte COBLANCE
 - La cohorte CRYOSTEM
 - La cohorte HEPATHER
 - Le projet HOPE-EPI
 - La cohorte MEMENTO
 - La cohorte OFSEP
 - La cohorte PSY-COH
 - La plateforme RADICO

Annexe 5 : Quelques cohortes ou dispositifs de suivis longitudinaux soutenus par d'autres canaux que le PIA

Annexe 6 : Les potentialités liées au SNDS et aux plateformes des données de santé en cours de constitution

Annexe 7 : Les grandes cohortes épidémiologiques en population générale : historique et panorama international

Annexe 8 : L'expérience des grandes cohortes épidémiologiques au Royaume-Uni

LISTE DES PERSONNES RENCONTREES HORS VISITES DES COHORTES

Cabinet de la ministre des solidarités et de la santé

Raymond Le Moign, directeur

Valérie Mazeau-Woynar, conseillère

Cabinet de la ministre de l'enseignement supérieur, de la recherche et de l'innovation (Mesri)

Isabelle Richard, conseillère santé

Direction générale de la recherche et de l'innovation (DGRI)

Anne Paoletti, directrice scientifique du secteur Biologie et santé

Elisabeth Vergès, directrice du service Stratégie recherche, innovation recherche

Benoît Lavallart, chargé de mission au service Stratégie recherche, innovation recherche

Nicolas Dromel, chef du département des grandes infrastructures de recherche

Isabelle Diaz, chargée de mission au département des grandes infrastructures de recherche

Direction générale de l'enseignement supérieur et de l'insertion professionnelle (DGESIP)

Alain Abecassis, chef du service de la coordination des stratégies de l'enseignement supérieur et de la recherche

Direction générale de la santé (DGS)

Jérôme Salomon, directeur général

Benoit Marin, adjoint à la sous-directrice de la politique des produits de santé et de la qualité des pratiques et des soins ;

Olivier Brahic, adjoint au sous-directeur Veille et sécurité sanitaire

Arnaud de Guerra, chargé de mission à la Mission Stratégie-Recherche

Marie Seval, conseillère « information et santé » à la Mission Stratégie-Recherche

Patrice Dosquet, chargé de mission à la sous-direction Santé des populations et prévention des maladies chroniques

Secrétariat général pour l'investissement (SGPI)

Jean-Christophe Dantone, directeur du programme Santé et biotechnologies

Direction générale de l'offre de soins (DGOS)

Stéphanie Decoopman, cheffe de service, adjointe à la directrice générale

Laure Maillant, cheffe du bureau innovation et recherche clinique

Camille Ruiz, adjointe au chef de bureau évaluation, modèles et méthodes

Direction de la recherche, des études, de l'évaluation et des statistiques (Drees)

Jean-Marc Aubert, directeur

Muriel Barlet, sous directrice de l'observation de la santé et de l'assurance maladie

Stéphanie Combes, cheffe de la mission d'administration des données de santé

Philippe Raynaud, chef du bureau de l'état de santé de la population

Patrick Aubert, sous-directeur de l'observation de la solidarité

Direction de l'animation de la recherche, des études, de l'évaluation et des statistiques (Dares)

Malik Koubi, sous-directeur des salaires, du travail et des relations professionnelles

Thomas Coutrot, chef du département Conditions de travail et santé

Amélie Mauroux, adjointe au chef du département Conditions de travail et santé

Direction générale du travail (DGT)

Frédéric Tézé, sous-directeur des conditions de travail, de la santé et de la sécurité au travail

Corinne Piron, cheffe de l'inspection médicale du travail

Lucie Mediavilla, chargée de mission pour le suivi des opérateurs

Secrétariat général du ministère de l'agriculture et de l'alimentation

Philippe Quittat-Odelain, chef du bureau de la santé et de la sécurité au travail

Anne Marie Soubielle, chargée de mission maladies professionnelles agricoles

Laure Wurtz, chargée de mission risques chimiques

Institut national des statistiques et des études économiques (Insee)

Jean-Luc Tavernier, directeur général

Chantal Cases, directrice des statistiques démographiques et sociales

Institut national des données de santé (INDS)

Dominique Polton, présidente

Centre d'accès sécurisé aux données (CASD)

Kamel Gadouche, directeur

Philippe Donnay, directeur adjoint

Agence nationale de la recherche (ANR)

Thierry Damerval, président-directeur général

Arnaud Torres, directeur des grands programmes d'investissement de l'Etat

Dominique Dunon-Bluteau, responsable du département biologie et santé

Mathieu Lévi-Strauss, responsable scientifique au département biologie et santé

Loreline Robbe, chargée de projets scientifiques au département biologie et santé

Institut national de la santé et de la recherche médicale (Inserm)

Gilles Bloch, président-directeur général

Claire Giry, directrice générale déléguée

Corinne Alberti, co-directrice de l'Institut thématique multi-organismes (ITMO) Santé publique de l'Alliance nationale pour les sciences de la vie et de la santé (Aviesan), directrice de l'IRESP

Sammy Sahnoune, directeur du département Système d'information

Institut national d'études démographiques (Ined)

Magda Tomasini, directrice

Marie-Aline Charles et Bertrand Geay, responsables de l'Etude longitudinale française depuis l'enfance (UMS Ined-Inserm-EFS)

Institut national du cancer (Inca)

Thierry Breton, directeur général

Philippe-Jean Bousquet, directeur de l'observation, des sciences, des données et de l'évaluation

Tristan Roué, chargé de mission au pôle recherche et innovation

Agence nationale de recherches sur le sida et les hépatites virales (ANRS)

François Dabis, directeur

Ventzislava Petrov-Sanchez, responsable recherche clinique

Thibault Robert, secrétaire général

Thierry Menvielle, responsable affaires financières

Autres présidents ou directeurs d'établissements coordonnateurs

Alexander Eggermont, directeur général de l'Institut Gustave Roussy (IGR)

Eric Solary, directeur de la recherche de l'IGR

Alain Bui, président de l'Université Versailles Saint Quentin en Yvelines (UVSQ)

Alexis Constantin, vice-président de la commission recherche de l'UVSQ

Catherine Billard, vice-présidente de la FVU

Frédéric Dardel, ancien président de l'Université Paris Descartes

Sylvie Retailleau, présidente de l'Université Paris Saclay (UPS) et ex présidente de l'Université Paris-Sud

Etienne Augé, vice-président recherche de l'Université Paris Sud

Armelle Regnault, directrice de la recherche de l'Université Paris Saclay

Manuel Tunon de Lara, président de l'Université de Bordeaux

Guillaume du Chaffaut, directeur général adjoint des Hospices civils de Lyon

Stéphanie Fazi-Leblanc, directrice générale adjointe du CHU de Bordeaux

Centre de recherche épidémiologie et statistique (Cress)

Philippe Ravaud, directeur

Stéphanie Chardon, directrice générale

Centre de recherche en épidémiologie et santé des populations (Cresp)

Bruno Falissard, directeur

Sabine Barbut, secrétaire générale

Caisse nationale de l'assurance maladie (Cnam)

Nicolas Revel, directeur général

Olivier Lyon-Caen, médecin-conseil national

Claude Gissot, directeur de la stratégie, des études et des statistiques

Saïd Oumeddour, sous-directeur responsable de la prévention et de la promotion de la santé

Marie Christine Lacroix Schneider, responsable de la coordination des centres d'examen de santé à la direction déléguée à la gestion et à l'organisation des soins

Santé publique France (SpF)

Jean-Claude Desenclos, directeur scientifique

Anne Gallay, directrice des maladies non transmissibles et traumatismes

Yann Le Strat, directeur appui, traitements et analyses de données (DATA)

Didier Che, adjoint au directeur des maladies infectieuses

Florence Lot, responsable d'unité à la direction des maladies infectieuses

Jean-Baptiste Richard, direction appui, traitement et analyse des données

Béatrice Geoffroy, chef de projet COSET-MSA à la direction santé environnement travail

Jean-Luc Marchand, chef de projet COSET-Indépendants à la direction santé environnement travail

Agence nationale de surveillance du médicament et des produits de santé (ANSM) et Gis Epi-phare

Mahmoud Zureik, directeur scientifique, responsable du GIS Epi-phare épidémiologie des produits de santé

Rosemary Dray-Spira, directrice adjointe du Gis Epi-phare

Evelyne Duplessis, directrice déléguée à l'informationnel et au pilotage de l'ANSM

Patrick Maison, conseiller scientifique à l'ANSM

Agence de la biomédecine (ABM) (contact téléphonique)

Hugo Gilardi, directeur général par intérim

Yves Perel, directeur général adjoint chargé de la politique médicale et scientifique

Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses) (contact téléphonique)

Roger Genet, directeur général

Gérard Lasfargues, directeur général délégué en charge du pôle sciences pour l'expertise

Caroline Semaille, directrice générale déléguée en charge du pôle produits réglementés

Mathieu Schuller, directeur de l'évaluation des risques

Caisse nationale de solidarité pour l'autonomie (CNSA)

Virginie Magnant, directrice

Julie Micheau, directrice scientifique

Haute Autorité de Santé (HAS)

Dominique Le Guludec, présidente

Chantal Bélorgey, directrice de l'évaluation médicale, économique et de santé publique

Catherine Grenier, directrice de l'amélioration et de la qualité des soins

Michèle Morin-Surrocca, responsable du service médico-économique et de santé publique

Claire Brotons, adjointe à la responsable du service médico-économique et de santé publique

Marion Pinet, cheffe de projet pour les études post-inscription au service d'évaluation du médicament

Haut Conseil pour la santé publique (HCSP)

Franck Chauvin, président

Bernard Faliu, secrétaire général

Thomas Lefèvre, copilote du groupe de travail pour la saisine relative à la politique nationale des registres

Assistance publique-Hôpitaux de Paris (AP-HP)

Philippe-Gabriel Steg, vice-président du directoire chargé de la recherche

Tabassome Simon, vice-présidente de la commission recherche

Mathieu Resche-Rigon, co-responsable du service de biostatistique et information médicale de l'hôpital Saint Louis

Florence Favrel-Feuillade, directrice de la délégation de la recherche clinique et de l'innovation (DRCI)

Anne Gysembergh-Houal, responsable des partenariats R&D à la DCRI

Fédération Unicancer

Jean-Yves Blay, président (entretien téléphonique)

Pascale Flamant, déléguée générale

Claire Labreuveux, directrice de la R&D

Anne-Laure Martin, directrice de la recherche clinique et translationnelle

Institut de recherche et documentation en économie de la santé (Irdes)

Denis Raynaud, directeur

Catherine Sermet, directrice adjointe

Paul Dourgnon, directeur de recherche

Conférences des présidents d'université (CPU)

Manuel Tunon de Lara, président de la commission des questions de santé

Fondation Alzheimer

Philippe Amouyel, directeur général

Sylvie Ledoux, déléguée scientifique

Ligue contre le cancer

Axel Kahn, président

Iris Pauporté, déléguée à la recherche

LEEM et ARIIS

Thomas Borel, directeur recherche, innovation, santé publique et engagement sociétal

Gaëlle Nachbaur, directrice recherche de GSK, représentante de l'Ariis

Ophélie de Dreux-Brézé, responsable projets et partenariats scientifiques (LEEM)

Entreprises

Anny Tirel, directrice des services médicaux de MSD France (contact écrit)

Valérie Machuron, responsable acquisition données externes au centre de données médicales du Laboratoire Roche

Muriel Licour, directrice Europe épidémiologie, données de vraie vie, à AstraZeneca (contact écrit)

Personnalités qualifiées

Marcel Goldberg, professeur émérite d'épidémiologie à l'Université Paris Descartes

Alfred Spira, professeur honoraire de santé publique et d'épidémiologie à la Faculté de Médecine de Paris Saclay

François Bourdillon, ancien directeur général de Santé publique France

William Dab, ancien directeur général de la santé, professeur titulaire de la chaire d'Hygiène et sécurité du Conservatoire national des arts et métiers

Arnaud Fontanet, professeur à l'Institut Pasteur et au Conservatoire national des arts et métiers, chargé en 2018-2019 de la chaire Santé publique au Collège de France

Henri Leridon, directeur de recherche émérite à l'Ined

Florence Jusot, professeur en sciences économiques à l'Université de Paris Dauphine, *Country Team Leader* du projet *Share France*,

Thomas Renaud et Benjamin Lévy, gestionnaires du projet *Share France*

Serge Hercberg, Université Paris 13 et hôpital Avicenne, directeur de l'Équipe de recherche en épidémiologie nutritionnelle (EREN)

Joël Ankri, vice-président du comité de suivi du Plan Maladies neurodégénératives

Geneviève Chêne, professeur de santé publique, ISPED/Université de Bordeaux, ancienne directrice de l'Iresp

Pierre Yves Geoffard, ancien directeur de l'École d'économie de Paris, co-auteur du rapport du Cnis de 2017 sur l'accès des chercheurs aux données administratives

LISTE DES PERSONNES RENCONTREES LORS DES REUNIONS DE TRAVAIL AVEC LES COHORTES DU PIA

Cohorte CONSTANCES (9 juillet 2019)

Équipe de la cohorte

Marie Zins, directrice scientifique et technique

Marcel Goldberg, co-investigateur

Guillaume Airagnes

Julie Gourmelen

Nathalie Lopes

Anna Ozguler

Adeline Renuy

Céline Ribet

Emmanuel Wiernik

Autres chercheurs

Yves Roquelaure, Université et CHU d'Angers

Rachel Nadif, Inserm

Partenaires institutionnels

Saïd Oumeddour, Caisse nationale de l'assurance maladie,

Jean-Claude Desenclos, Santé publique France, membre du conseil scientifique

Mireille Caralp, Inserm Transfert

Partenaires industriels

Anny Tirel, MSD France

Cohorte E3N/E4N (20 septembre 2019)

Équipe de la cohorte

Gianluca Severi, responsable

Marie-Christine Boutron-Ruault, co-responsable

Pascale Gerbouin-Rérolle, responsable administrative

Iris Cervenka

Rafika Chaït

Douae El-Fatouhi

Élodie Faure

Agnès Fournier

Roselyn Gomes

Nasser Laouali

Camille Laplanche

Conor MacDonald

Yahia Mahamat-Saleh

Francesca Mancini

Autres chercheurs

Alexis Elbaz, Inserm

Pascal Guénel, Inserm

Florence Menegaux, Inserm

Raphaele Seror, AP-HP

Thérèse Truong, Inserm

Raphaëlle Varraso, Inserm

Partenaires institutionnels

Alexander Eggermont, Institut Gustave Roussy

Éric Solary, Institut Gustave Roussy

Sabine Barbut, CESP

Karim Ould-Kaci, Fondation MGEN

Iris Pauporté, Ligue nationale contre le cancer

Mireille Caralp, Inserm Transfert

Partenaires industriels

Thomas Czernichow, Epiconcept

Plateforme RECONAI (17 juillet 2019)

Équipe de la plateforme

Marie-Aline Charles, coordonnatrice du projet

Bertrand Geay, co-directeur de l'UMR ELFE

Pierre-Yves Ancel, coordinateur du projet EPIPAGE 2

Catherine Guével

Cécile Zaros

Jean-Louis Lanoé

Véronique Pierrat

Valérie Benhamou

Mélanie Durox

Adriana Candea

Elodie Lordmi

Sophie de Visme

Laurence Germany

Judith van der Waerden

Autres chercheurs

Henri Leridon, responsable du projet ELFE 2005-2010

Pierre-Henri Jarreau

François Goffin

Michelle Kelly-Irving

Barbara Heude

Franck Ramus

Sophie Nicklaus

Anne Solaz

Florence Zeman

Maria Melchior

Partenaires institutionnels

Magda Tomasini, directrice de l'Ined

Loup Wolff, ministère de la culture

Frédérique Chave, Caisse nationale des allocations familiales

Enguerrand du Roscoat, Santé publique France

Vanessa Besson-Dubourg, Inserm Transfert

Hélène Blanché, Fondation Jean Dausset

Claire Gerdil, Établissement français du sang

Cohorte CANTO (16 septembre 2019)

Équipe de la cohorte

Fabrice André, coordonnateur

David Aubry, Unicancer

Romain Boidin, Unicancer

Coralie Courtinard, Unicancer

Jérôme Lemonnier, Unicancer

Anne-Laure Martin, Unicancer

Patricia Rigou, Unicancer

Aude Sirven, Unicancer

Pierre Beuchet, Institut Gustave Roussy

Patrick Arveux, Centre Georges-François Leclerc

Aurélie Bertaux, Centre Georges-François Leclerc

Sandrine Boyaux, Centre Léon Bérard

Autres chercheurs

Marie Lange

Sarah Dauchy

Gwenn Menvielle

Agnès Dumas

Inès Vaz-Luis

Paul Cottu

Sofia Rivera

Partenaires institutionnels

Philippe-Jean Bousquet, Institut national du cancer

Iris Pauporté, Ligue nationale contre le cancer

Pierre Fumoleau, Institut Curie - Unicancer

Partenaires industriels

Valérie Machuron, ROCHE

Muriel Licour, AstraZeneca

Cohorte CKD REIN (21 octobre 2019)

Équipe de la cohorte

Bénédicte Stengel, coordinatrice de la cohorte

Christian Jacquelinet, Agence de la Biomédecine

Sophie Renault

Marie Metzger, Inserm

Elodie Speyer

Natalia Alencar de Pinho

Solène Laville

Autres chercheurs

Sophie Liabeuf, CHU d'Amiens

Karine Legrand, CHRU de Nancy (par téléphone)

Carole Ayav, CHRU de Nancy (par visio-conférence)

Luc Frimat, vice-président de la société francophone de néphrologie, CHRU Nancy (par visio-conférence)

Sandra Wagner, CHRU Nancy (par visio-conférence)

Maurice Laville, ancien président de la société francophone de néphrologie, CHU Lyon-Sud

Ziad Massy, président du conseil scientifique de CKD-REIN (par visio-conférence)

Joost Schanstra, Inserm, Toulouse (par téléphone)

Christian Combe, CHU Bordeaux, ancien président de la société francophone de néphrologie, dialyse et transplantation (par visio-conférence)

Partenaires institutionnels

Yves Edouard Herpe, Biobanque de Picardie

Anne Boland, Centre National de Recherche en Génomique Humaine (par téléphone)

Sabine Barbut, CESP

Mireille Caralp, Inserm Transfert

Partenaires industriels

Serge Maillet, Vifor Fresenius Medical Care Renal Pharma (par téléphone)

Mélanie Hueber, Otsuka Pharmaceutical France

Anne-Charlotte de Gouville, GSK

Cohorte COBLANCE (16 octobre 2019)

Équipe de la cohorte

Simone Benhamou, coordinatrice du projet

Yves Allory, comité de coordination de la cohorte

Julia Bonastre, comité de coordination de la cohorte

François Radvanyi, comité de coordination de la cohorte

Karine Groussard

Anthony Mangin

Tasnim Hedoui

Partenaires institutionnels

Sabine Barbut, CESP

Tatiana Malherbe, Institut Curie

Odette Mariani, Institut Curie

Pierre Beuchet, Institut Gustave Roussy

Hélène Blanché, Fondation Jean Dausset

Bertrand Denis, AFNOR

Cohorte CRYOSTEM (23 septembre 2019)

Équipe de la cohorte

Régis Peffault de Latour, coordinateur scientifique

Boris Calmels, coordinateur des CRB

Jean-Hugues Dalle, représentant des cliniciens pédiatriques

Claire Fontenille

Émilie Robert

Partenaires institutionnels et chercheurs

Jacques-Olivier Bay, président de la SFGM-T.

Christian Chabannon, Institut Paoli- Calmettes, Marseille

Juliette Berger, CHU Estaing, Clermont-Ferrand

Marie-Thérèse Rubio, présidente du comité scientifique de la SFGM-TC, CHRU Nancy

Gérard Michel, AP-HM, Marseille.

Partenaires privés

Jean-Mehdi Grangeon, KOM Agency.

Stéphane Couchoux, responsable secteur « Fondations, Mécénat & Entreprises », Fidal.

Laurent Jacotot, Modul-Bio.

Cohorte HEPATHER (5 novembre 2019)

Équipe de la cohorte et partenaires gestionnaires

Fabrice Carrat, Sorbonne Université et APHP, responsable scientifique

Hélène Fontaine, APHP Paris, co-investigatrice,

Carole Cagnot, ANRS

Céline Dorival, Sorbonne Université

Ventzislava Petrov-Sanchez, ANRS

Guillaume Le Meut, ANRS

Thierry Menvielle, ANRS

Isabelle Goderel, Inserm

Frédéric Chau, Sorbonne Université

Clovis Lusivika-Nzinga, Sorbonne Université

Grégory Pannetier, Sorbonne Université

Guillaume Dussailant, Sorbonne Université

Autres chercheurs

Pierre Nahon, co-investigateur LICAVIR (par téléphone)

Georges-Philippe Pageaux (par téléphone)

Victor de Ledhingen (par téléphone)

Dominique Larrey (par téléphone)

Slim Fouraty (par téléphone)

Partenaires institutionnels

Mireille Caralp, Inserm Transfert,

Mathias Bruyand, Santé publique France

Partenaires privés

Christophe Hezode, Gilead (par téléphone)

Rachel Blum, Abbvie, (par téléphone)

Anne-Marie Stelart, Abbvie (par téléphone)

Plateforme HOPE-EPI (8 octobre 2019)

Équipe de la plateforme

Jacqueline Clavel, coordinatrice du projet

Brigitte Lacour

Emmanuel Désandes

Lucie Vignon

Valérie Bernier

Claire Poulhalon

Sofiène Ben Salha

Stéphanie Goujon

Latifa Idbrik

Nicolas Simon

Soraya Aakki

Autres chercheurs

Florent de Vathaire, Inserm (cohorte FCCSS)

Pascal Auquier, AP-HM (cohorte LEA)

Gérard Michel, AP-HM (cohorte LEA)

Partenaires institutionnels

Philippe-Jean Bousquet, Institut national du cancer

Sébastien Denys, Santé publique France

Cohorte MEMENTO (11 octobre 2019)

Équipe de la cohorte

Geneviève Chêne, co-investigatrice principale

Carole Dufouil, co-investigatrice principale

Valérie Boilet, Cheffe de projet

Vincent Bouteloup

Delphine Jean

Laetitia Lacaze-Buzy

Emilie Laheranne

Lisa Le Scouarnec

Nathalie Thiery

Jonviea Chamberlain

Leslie Grasset

Jérémie Lespinasse

Autres chercheurs

Jean-François Dartigues, investigateur principal des cohortes PAQUID et 3C

Mathieu Ceccaldi, ancien président du comité scientifique

Partenaires institutionnels

Philippe Moretto, Université de Bordeaux

Stéphanie Debette, Université de Bordeaux

Fazi-Leblanc, CHU de Bordeaux

Partenaires industriels

Olivier Courrèges, société *Qynapse*

Cohorte OFSEP (25 septembre 2019)

Équipe de la cohorte

Sandra Vukusic, coordinatrice du projet

Claire Rigaud-Bully

Romain Casey

Irena Vukusic

Wladys Craz

Anne-Cécile de Giacomoni

Autres chercheurs

Gilles Edan, CHU de Rennes

François Cotton, CHU de Lyon

David Laplaud, CHU de Nantes

Eric Thouvenot, CHU de Nimes

Thibault Moreau, CHU de Dijon

Jean Pelletier, Assistance publique – Hôpitaux de Marseille

Francis Guillemin, CHRU de Nancy

Partenaires institutionnels

Guillaume du Chaffaut, Hospices civils de Lyon

Bruno Lina, Université Lyon 1

Michel Clanet, Fondation Edmus

Gilles Salles, Fondation Edmus

Emmanuelle Leray, Ecole des hautes études en santé publique (EHESP)

Xavier Kurz, Agence européenne des médicaments (EMA)

Jean-Frédéric de Leusse, Association pour l'aide à la recherche sur la sclérose en plaques (ARSEP)

Partenaires industriels

Thibault Lemoine, NOVARTIS Pharma

Isabelle Chouette, NOVARTIS Pharma

Paul-Henri Depoortere, ROCHE

Naimi Missoum, BIOGEN France

Agathe Maury Le Breton, BIOGEN France

Cohorte PSY-COH (13 septembre 2019)

Équipe de la cohorte et centres associés

Marion Leboyer, coordinatrice

Pierre-Michel Llorca, CHU et Université de Clermont-Ferrand, investigateur principal

Dominique Costagliola, iPLesp, Inserm

Isabelle Durand-Zaleski, UPEC, AP-HP

Jasmina Mallet, Paris

Paul Roux, Versailles

Emilie Olié, Montpellier

Raoul Belzeaux, Marseille

Stefan Albagli, Fondation FondaMental

Maria Chovet, Fondation FondaMental

Ophélie Godin, Fondation FondaMental

Autres chercheurs

Stéphane Jamain, Inserm

Ryad Tamouza, Inserm

Christophe Junot, CEA,

Josselin Houenou, Inserm et CEA,

Franck Schurhoff, Inserm

Partenaires industriels

Slah Aridhi, société *Sensoria analytics*

Valérie Autier, société *Metabrain Research*

Daniel Szeftel, société *Semeia*

Partenaires associatifs

Annie Labbé, Argos 2001

Marie-Jeanne Richard, Unafam

Plateforme RADICO (2 octobre 2019)

Équipe de la plateforme

Serge Amselem, coordinateur du projet

Paul Landais, comité de direction du programme

Annick Clément, comité de direction du programme

Sonia Guéguen, comité de direction du programme

Autres chercheurs

Nicolas Chassaing, cohorte RADICO-ACOEIL

Vincent Cottin, cohorte RADICO-PID

Estelle Escudier, cohorte RADICO-DCP

Bénédicte Héron, cohorte RADICO-MPS

Xavier Jeunmaître, cohorte RADICO-SEDVASC

Partenaires institutionnels

Mireille Caralp, Inserm Transfert

Catherine Nguyen, Inserm

Laurent Gouya, Université Paris Diderot

Guillaume Jondeau, Assistance publique – Hôpitaux de Paris

Partenaires industriels

Caroline Blanc-Crouzier, Shire-Takeda

Anne-Sophie Chalandon, LEEM

Marco Fiorini, Ariis

Olivier Delaitre, Boehringer

Katell Le Lay, Boehringer

LISTE DES PERSONNES RENCONTREES LORS DU DEPLACEMENT AU ROYAUME-UNI (28 AU 30 OCTOBRE 2019)

Jean Arlat, conseiller pour les affaires scientifiques et technologiques, Ambassade de France

Marguerite Moleux, conseillère pour les affaires sociales, Ambassade de France

Naomi Allen, Director, UK Biobank

Jonathan Sellors, Legal Counsel, UK Biobank

Sir John Bell, Regius Professor, Oxford University

Bruna Galobardes, Portfolio Manager, epidemiological studies, Wellcome Trust

Rebecca Hardy, Director, Cohort and Longitudinal Studies Enhancement Resources (CLOSER)

Lisa Calderwood, Centre for Longitudinal Studies (CLS)

Alice Sullivan, Centre for Longitudinal Studies (CLS)

Gavin Malloch, programme manager for Public Health Partnerships, Medical Research Council

Ming Tang, National Director Data and Analytics, NHS Improvement

SIGLES UTILISES

ABM : Agence de la biomédecine

ANR : Agence nationale de la recherche

ANRS : Agence nationale de recherches sur le sida et les hépatites virales

Anses : Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail

ANSM : Agence nationale de sécurité du médicament et des produits de santé

APA : Allocation personnalisée d'autonomie

AP-HP : Assistance publique – Hôpitaux de Paris

ARC : Attaché de recherche clinique

Ariis : Alliance pour la recherche et l'innovation des industries de santé

ASE : Aide sociale à l'enfance

Aviesan : Alliance pour les sciences de la vie et de la santé

BDMA : Base de données médico-administratives

CAR-T : *Chimeric antigen receptor T*

CASD : Centre d'accès sécurisé aux données

CDD : Contrat à durée déterminée

CDI : Contrat à durée indéterminée

CEREES : Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé

CESP : Centre de recherche en épidémiologie et santé des populations

CGI : Commissariat général à l'investissement

CINES : Centre informatique national de l'enseignement supérieur

Cnaf : Caisse nationale des allocations familiales

Cnam : Caisse nationale d'assurance maladie

Cnil : Commission nationale de l'informatique et des libertés

Cnis : Conseil national de l'information statistique

CNRS : Centre national de la recherche scientifique

CNSA : Caisse nationale de solidarité pour l'autonomie

COSET : Cohorte pour la surveillance épidémiologique en lien avec le travail

CRB : Centre de ressources biologiques

CRESS : Centre de recherche en épidémiologie et statistiques

CSP : Catégorie socio-professionnelle

Dares : Direction de l'animation de la recherche, des études et des statistiques

Depp : Direction de l'évaluation, de la prospective et de la performance

DGOS : Direction générale de l'offre de soins

DGS : Direction générale de la santé

DGT : Direction générale du travail

Drees : Direction de la recherche, des études, de l'évaluation et des statistiques

EFS : Établissement français du sang

EHIS : *European health interview survey*

ENEAS : Echantillon national d'études sur l'autonomie et l'hébergement des seniors

ERN : *European reference networks*

ESRC : *Economic and Social Research Council*

GIS : Groupement d'intérêt scientifique

GIP : Groupement d'intérêt public

HAS : Haute autorité de santé

HCERES : Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur

HCSP : Haut conseil de la santé publique

HDH : *Health data hub*

IGAENR : Inspection générale de l'administration, de l'éducation nationale et de la recherche

Igas : Inspection générale des affaires sociales

IGÉSR : Inspection générale de l'éducation, du sport et de la recherche

IHU : Institut hospitalo-universitaire

INBS : Infrastructures nationales en biologie santé

Inca : Institut national du cancer

INDS : Institut national des données de santé

Ined : Institut national des études démographiques

Insee : Institut national de la statistique et des études économiques

Inserm : Institut national de la santé et de la recherche médicale
IPLeSp : Institut Pierre Louis d'épidémiologie et de santé publique
Iresp : Institut de recherche en santé publique
ITMO : Institut thématique multi-organismes
LEEM : Les entreprises du médicament
MDPH : Maison départementale des personnes handicapées
MRC : *Medical Research Council*
MSA : Mutualité sociale agricole
MTI : Médicaments de thérapies innovantes
NHS : *National Health Service*
NIR : Numéro d'inscription au répertoire national d'identification
OMS : Organisation mondiale de la santé
OST : Observatoire des sciences et techniques
PIA : Programme des investissements d'avenir
PGC : *Psychiatric genetics consortium*
PHRC : Programme hospitalier de recherche clinique
PMSI : Programme de médicalisation des systèmes d'information
RGPD : Règlement général de protection des données
SGPI : Secrétariat général pour l'investissement
SNDS : Système national des données de santé
SNIIRAM : Système national inter-régimes de l'assurance maladie
SNR : Stratégie nationale de recherche
SNS : Stratégie nationale de santé
SpF : Santé publique France
UKRI : *United Kingdom Research and Innovation*
UMR : Unité mixte de recherche
UMS : Unité mixte de service
VIH : Virus de l'immunodéficience humaine