

Les cohortes au niveau international : histoire et perspective

La plupart des grandes cohortes ont vu le jour après la Deuxième Guerre mondiale. La cohorte américaine Framingham, créée en 1949, a révolutionné la compréhension des maladies cardio-vasculaires. Récemment mise en place, la cohorte française Constances réunira 200 000 participants de 18 à 69 ans, fréquentant les centres d'examens de santé de la Sécurité sociale.

Les cohortes épidémiologiques dans la recherche internationale

Archana Singh-Manoux
Inserm, Unité 1018,
Centre de recherche
en épidémiologie
et santé des
populations

Le terme « cohorte » a été utilisé pour la première fois dans les années 1930 par l'Américain Wade Hampton Frost et le Norvégien Kristian Andvord pour décrire leurs travaux sur les taux de décès liés à la tuberculose [21]. Ces analyses de taux de mortalité par tuberculose ont été menées en fonction de l'âge et de la cohorte de naissance, séparément chez les hommes et les femmes. L'étude de cohorte est un terme qui a été largement utilisé à partir de la fin des années 1950 pour désigner des études portant sur des groupes d'individus « exposés » ou « non exposés » à des facteurs de risque, suivis, pour une période plus ou moins longue, de la mortalité ou de l'incidence de maladies. La plupart des grandes cohortes ont été développées après la Deuxième Guerre mondiale et sont devenues l'un des éléments majeurs de développement de l'épidémiologie moderne. Ces premières cohortes emblématiques, qui ont contribué à définir le rôle de l'épidémiologie dans le domaine des sciences médicales, ont été lancées de façon indépendante et plus ou moins simultanément au Royaume-Uni et aux États-Unis.

Les cohortes après-guerre

Des cohortes emblématiques, qui sont maintenant très connues, ont été mises en place dans la période d'après guerre, au moment où la nécessité de collecter les données de façon prospective s'est fait sentir face à la méfiance vis-à-vis des résultats d'études cas témoins rétrospectives, par exemple pour établir le lien entre tabac et cancer du poumon. Ces premières cohortes se sont souvent limitées soit à l'étude d'une maladie spécifique (maladie cardio-vasculaire pour la cohorte Framingham) ou d'une exposition spécifique (tabac pour la British Doctors' Study). On donne ci-dessous une brève description de trois de ces cohortes.

British Doctors' Study

En 1950, Bradford Hill et Richard Doll ont mis en évidence, à partir d'une étude cas témoin [15], que le tabagisme était un facteur de risque de cancer du poumon. Toutefois, ce résultat venant d'une étude cas témoin, il n'a pas été largement accepté. En 1951, Hill & Doll ont alors mis en place une cohorte prospective auprès

*Les références entre
crochets renvoient à la
Bibliographie générale
p. 51.*



des médecins britanniques, the British Doctors' Study, afin d'obtenir l'information sur le tabagisme d'un grand nombre d'individus, et d'évaluer si les fumeurs avaient un risque de mortalité par le cancer du poumon plus élevé que les non-fumeurs au cours du suivi. Une population de médecins a été choisie pour deux raisons : 1) leur suivi est simple – tous les médecins en Grande-Bretagne font partie de la «British Medical Association»; 2) étant médecins, leurs réponses aux questions sur le tabagisme sont plus précises en raison de leur formation médicale. Un questionnaire a alors été envoyé à tous les médecins britanniques, et 34 440 hommes et 6 194 femmes ont participé. Les premiers résultats ont été publiés en 1954 et les résultats ultérieurs avec un suivi plus long ont confirmé que le tabac était un facteur de risque pour le cancer du poumon.

American Cancer Society Study est une cohorte très similaire, commencée au même moment, mais de manière complètement indépendante de la British Doctors' Study. Avec l'aide d'adhérents bénévoles de l'American Cancer Society, une étude sur 190 000 hommes entre 50 et 69 ans a été lancée [14]. Les résultats de cette étude de grande envergure ont largement contribué à établir les risques associés au tabagisme.

Framingham Heart Study

En 1949, Dawber et son équipe ont mis en place une étude de cohorte basée dans l'État du Massachusetts qui a révolutionné notre compréhension des maladies cardio-vasculaires [12]. La caractéristique particulière de cette cohorte était que le suivi des participants n'était pas passif. À l'inclusion, comme pendant le suivi, les participants ont été invités régulièrement à fournir des informations via un auto-questionnaire, mais ont également subi un examen clinique approfondi qui incluait un électrocardiogramme et des prélèvements de sang. La cohorte initiale se composait de plus de 5 000 participants, dont la moitié était des femmes. En 1971, 5 100 nouveaux participants — les enfants des participants de la cohorte originale et leurs conjoints — ont été recrutés pour mener cette étude dans la deuxième génération. L'étude ne se concentre plus seulement sur les maladies cardio-vasculaires, elle inclut désormais le cancer, la démence, l'arthrite, l'ostéoporose et les troubles de l'audition et de la vision. Encore plus tard, en 2002, 3 900 petits-enfants des participants de la cohorte originale ont été inscrits dans la «Third Generation Study». Plus de 2 000 articles scientifiques ont été publiés à partir des données de cette cohorte. La poursuite de la cohorte sur une durée de plus de soixante ans et les examens cliniques répétés ont permis en particulier de contribuer de façon unique à la connaissance de la maladie cardio-vasculaire.

1946 Birth Cohort

L'autre cohorte majeure de cette période d'après-guerre a été mise en place en Grande-Bretagne par James Douglas en 1946. Cette cohorte a débuté tout d'abord

simplement pour comprendre les raisons de la baisse des taux de fécondité en Grande-Bretagne. Les objectifs étaient d'évaluer la disponibilité, l'utilisation et l'efficacité des maternités. Il s'agissait d'inclure toutes les femmes ayant accouché pendant une semaine en mars 1946. 13 687 femmes ont ainsi été recrutées. Le baby-boom d'après-guerre a rendu l'objectif initial de l'étude moins crucial, mais l'étude a été jugée importante pour évaluer l'impact des conditions familiales et sociales sur le développement des enfants. Finalement, un échantillon aléatoire de plus de 5 000 enfants, nés en mars 1946, a été mis en place. Les participants de cette cohorte, maintenant âgés de plus de 65 ans, ont été suivis depuis lors, avec l'analyse de bulletins scolaires pendant l'enfance et l'adolescence, la réalisation périodique d'interviews, de tests neuropsychologiques et d'examens médicaux. Cette cohorte a démontré l'utilité et la faisabilité des cohortes de naissance, ce qui a encouragé la réalisation des cohortes de naissance de 1958, 1970, et de 2000 en Grande-Bretagne. Les dispositifs longitudinaux d'observation n'ont cessé depuis lors de se développer à travers le monde. Ce type d'étude permet en effet d'adopter une approche dynamique du développement de l'enfant et de mieux comprendre ses trajectoires à long terme.

Les cohortes modernes

Les grandes études de cohorte actuelles ont commencé à se développer dans les années 1970, quand le caractère multifactoriel de l'étiologie des maladies chroniques a été largement reconnu. Parallèlement, les modèles causaux des maladies et les méthodes statistiques pour traiter les données longitudinales dans les cohortes sont devenus plus complexes. Alors que les premières études de cohorte utilisaient principalement des tables de survie avec des variables binaires (présent/absent) comme principale mesure de santé, les années récentes ont vu se développer des méthodes multivariées et l'utilisation de marqueurs de progression de la maladie mesurés au cours du temps, impliquant l'analyse de données longitudinales. La possibilité d'apparier les données des sujets à celles des registres nationaux de mortalité et de maladie a aussi contribué à la richesse de ces cohortes. Il est évidemment impossible aujourd'hui de recenser toutes les cohortes existant à l'échelle internationale. On ne décrira donc ici que quelques cohortes particulièrement illustratives des apports de cet outil épidémiologique.

Nurses' Health Study

La Nurses' Health Study a été mise en place aux États-Unis en 1976 chez des femmes âgées de 30 à 55 ans dans le but initial d'étudier les effets à long terme de la contraception orale sur la santé. Cette étude assure le suivi prospectif de 122 000 infirmières par auto-questionnaires. Dès 1989, 33 000 échantillons de sang ont été collectés et conservés en vue d'études ultérieures; une deuxième vague de collectes de sang s'est effectuée en 2000. La Nurses' Health Study II

a été mise en place en 1989 dans le but de répondre à des questions soulevées par la première étude et que la méthodologie d'origine n'avait pas permis d'aborder : effets de la contraception orale en tenant compte de l'âge de début, des habitudes alimentaires et d'autres facteurs de risque. Cette deuxième vague a inclus 116 000 femmes âgées de 25 à 42 ans. Les deux phases de cette étude continueront à explorer les facteurs de risque pour les principales maladies chroniques chez les femmes.

Physicians' Health Studies

La Physicians' Health Study I a débuté en 1982 aux États-Unis. Elle a inclus 33 223 médecins hommes, âgés de 40 à 84 ans à l'inclusion. L'objectif de cette étude était d'examiner si l'aspirine et le bêta-carotène sont efficaces respectivement pour la prévention des maladies coronaires et du cancer. De façon inhabituelle, il s'agit d'un essai randomisé qui a évolué en une étude observationnelle et est devenu une des plus grandes cohortes aux États-Unis. L'intervention a été arrêtée en 1995, et les données ont montré qu'une faible dose d'aspirine diminuait le risque d'un premier infarctus du myocarde de 44 %, mais que la consommation de bêta-carotène ne protégeait pas du cancer. Depuis l'arrêt de l'intervention, les médecins de cette cohorte continuent à remplir des questionnaires annuels pour faire avancer les connaissances sur la prévention des maladies cardiovasculaires, du cancer et d'autres maladies chroniques.

Actuellement, un deuxième essai randomisé, la Physicians' Health Study II, commencé en 1997, a pour but de tester les bénéfices et les risques associés à la consommation de vitamine E, de vitamine C, et de suppléments multivitaminés. Ces suppléments sont largement utilisés, mais leurs effets sur la santé restent méconnus. 14 642 médecins hommes, âgés de 50 ans et plus, participent à cette étude qui repose sur un suivi annuel par questionnaire portant sur l'observation de l'essai, l'utilisation de médicaments, la survenue de maladies graves ou d'effets indésirables, et qui permet le recueil d'autres informations sur les facteurs de risque.

European Prospective Investigation into Cancer and Nutrition (EPIC)

L'étude EPIC a été conçue pour étudier la relation entre l'alimentation, l'état nutritionnel, le style de vie, les facteurs environnementaux et l'incidence du cancer. Dans les années 1990, EPIC a recruté plus d'un demi-million (520 000) de personnes dans dix pays d'Europe occidentale : Danemark, France, Allemagne, Grèce, Italie, Pays-Bas, Norvège, Espagne, Suède et Royaume-Uni. Tous les participants ont répondu à plusieurs questionnaires ; des échantillons sanguins ont été prélevés sur 387 256 participants, atteints ou non de cancers, et stockés pour de futures analyses. Un tronc commun de données dans les dix pays comprend l'éducation, le statut professionnel et les expositions professionnelles, le statut vital des parents et des frères et sœurs, les

antécédents de maladies chroniques, la consommation sur la vie entière de tabac, d'alcool, l'activité physique, et chez les femmes les indicateurs de la maturation, de la contraception et de la reproduction sexuelle.

Les cohortes Whitehall

La première cohorte Whitehall, mise en place en 1967-1969, avait pour objectif d'évaluer dans une population britannique le risque de maladie cardio-vasculaire, qui jusque-là n'avait été examiné que dans l'étude américaine de Framingham. Cette première cohorte Whitehall comprend un examen clinique initial et un suivi limité à la mortalité enregistrée. Le nom Whitehall a été donné à l'étude parce qu'elle était constituée de 19 018 hommes fonctionnaires travaillant dans des bureaux situés dans la rue du centre de Londres appelée Whitehall, siège de plusieurs ministères. Un des résultats majeurs de Whitehall I concerne l'existence d'un gradient social de santé, c'est-à-dire l'existence d'une relation dose-réponse entre la situation sur l'échelle socio-économique et la santé. Il est en effet apparu que l'association entre la mortalité coronarienne et la position socio-économique était plus importante que celle avec d'autres facteurs connus de risque de décès d'origine coronarienne.

C'est ainsi qu'en 1985 une nouvelle cohorte, appelée Whitehall II, a été mise en place, toujours chez des fonctionnaires britanniques, avec l'intention explicite d'enquêter sur les raisons des différences de santé entre les groupes socio-économiques. Bien que tous les participants de l'étude soient des « cols blancs », un large éventail de professions est représenté, avec une échelle des salaires allant de 1 à plus de 10 entre le bas et le haut des catégories professionnelles. La première phase de l'étude a été initiée au cours de la période 1985-1988 ; elle comportait un examen clinique et un questionnaire auto-administré, qui sont depuis répétés régulièrement. La cohorte Whitehall II contribue de façon majeure à la recherche en épidémiologie sociale, et est internationalement considérée comme l'une des principales sources de connaissance scientifique sur les déterminants sociaux de la santé.

Les consortiums et les méga-cohortes

Les cohortes sont désormais un outil épidémiologique fondamental. Toutefois, de nombreux changements sont en cours dans la période actuelle. L'identification de plusieurs variantes génétiques associées aux maladies et leur interaction potentielle avec les facteurs environnementaux ont nécessité la mise en place de consortiums de cohortes existantes pour avoir suffisamment de puissance pour les analyses. Ces cohortes doivent aussi tirer profit de l'évolution rapide des connaissances et des techniques de biologie moléculaire (génomique, protéomique, etc.), nécessitant la mise en place de vastes biobanques. Ces nouvelles « méga-cohortes » sont décrites dans l'article « Les nouvelles « méga-cohortes » en population en Europe » (lire *Les nouvelles « méga-cohortes » en population en Europe*, p. 34). ▮



Les nouvelles « méga-cohortes » en population en Europe

Marie Zins
Marcel Goldberg
Inserm U1018,
plate-forme de
recherche Cohortes
épidémiologiques
en population –
Centre de recherche
en épidémiologie
et santé des
populations,
université
de Versailles-
Saint-Quentin

Les premières grandes cohortes épidémiologiques en population ont été mises en place après la Seconde Guerre mondiale aux États-Unis et en Grande-Bretagne. Elles ont été fondatrices de l'épidémiologie moderne, et beaucoup sont toujours actives et continuent d'apporter des données précieuses dans des domaines variés. Aux États-Unis, on peut citer la cohorte Framingham, initiée en 1949 [35], la Nurses' Health Study mise en place en 1976 chez des infirmières [19], ou en Grande-Bretagne la British Doctors' Study suivie depuis 1951 [16], les cohortes Whitehall I et Whitehall II mises en place respectivement en 1967-1969 et en 1985 [34], ou encore les cohortes de naissance inaugurées en 1946 avec la Birth Cohort 1946. En France, l'Étude prospective parisienne consacrée à l'étude des risques cardio-vasculaires a été une cohorte pionnière, initiée dès 1967 [17]. Quelques-unes de ces cohortes « historiques » sont décrites dans les articles *Les cohortes épidémiologiques dans la recherche internationale*, p. 31 et *Les cohortes « historiques » en France*, p. 37.

La nécessité de cohortes prospectives de très grande taille

Aujourd'hui, les cohortes en population sont devenues un outil scientifique relativement courant, et il est impossible de recenser toutes les cohortes existant à l'échelle internationale. Ainsi, on comptait en 2004 pas moins de 47 cohortes épidémiologiques ayant contribué de façon significative à l'établissement des risques de divers cancers dus au tabac, c'est-à-dire ayant un recul et une puissance suffisants pour mettre en évidence de tels risques, y compris pour des cancers peu fréquents.

Cependant, la nature des questions scientifiques auxquelles on demande aux cohortes d'apporter des réponses évolue. Actuellement, la recherche sur les causes des maladies de nature environnementale, professionnelle, sociale, nutritionnelle, biologique et génétique, ou en pharmacoépidémiologie, concerne de plus en plus des risques de faible ampleur, donc difficiles à mettre en évidence : effets potentiellement cancérigènes des téléphones portables, des faibles doses de rayonnements ionisants, rôle de polymorphismes génétiques vis-à-vis de maladies multifactorielles, etc.

Dans ce contexte scientifique, des cohortes de très grande envergure, avec un suivi à long terme et un phénotypage (caractérisation des maladies) de haute qualité, sont nécessaires pour assurer une puissance statistique suffisante permettant de mieux comprendre

le rôle des divers facteurs personnels et environnementaux et leur interaction avec des caractères génétiques complexes. Par exemple, des associations établies entre des polymorphismes génétiques et des maladies chroniques montrent des risques relatifs typiquement compris entre 1,1 et 1,4, et la mise en évidence de façon fiable de tels effets exige de très vastes ensembles de données. Ainsi, les études cas-témoins doivent réunir des milliers de cas, même pour les situations les plus simples ; lorsque la question de recherche concerne l'étude des interactions gène-environnement et gène-gène et l'analyse approfondie de relations de cause à effet, des dizaines de milliers de cas sont souvent nécessaires. Des dizaines de milliers de sujets peuvent aussi être nécessaires pour étudier un phénotype quantitatif (pression artérielle par exemple), parce que les effets d'origine génétique peuvent être très faibles [3].

Les « méga-cohortes » en Europe

C'est dans ce contexte qu'on voit se mettre en place, depuis les années 2000, une nouvelle génération de « méga-cohortes » en population, et on estime qu'actuellement plus de 100 cohortes prospectives de population de grande dimension sont à divers stades de développement et de réalisation dans le monde, dont certaines dans des pays européens dont on présente les principales.

Certaines cohortes sont déjà en place en Europe

On peut ainsi citer en Grande-Bretagne la Million Women Study (<http://www.millionwomenstudy.org/introduction/>), qui a inclus entre 1996 et 2001 plus d'un million de femmes âgées de 50 ans et plus, ou le projet UK Biobank, qui a inclus, de 2006 à 2010, 500 000 personnes âgées de 40 à 69 ans (<http://www.ukbiobank.ac.uk/>). En Norvège (pays de 4,5 millions d'habitants, soit 13 fois moins peuplé que la France), la cohorte CONOR (Cohort of Norway) suit 200 000 adultes, et la cohorte MoBa (Norwegian Mother and Child Cohort Study) a inclus 270 000 mères, pères et leurs enfants (<http://www.fhi.no/eway/?pid=238>). La cohorte européenne EPIC (European Prospective Investigation into Cancer and Nutrition) réunit 520 000 participants âgés de 20 ans et plus dans 10 pays (Danemark, France, Allemagne, Grèce, Italie, Pays-Bas, Norvège, Espagne, Suède et Royaume-Uni) ; elle a été mise en place entre 1993 et 1999 avec 7 pays participants, et le Danemark, la Norvège et la Suède ont rejoint ultérieurement la cohorte (<http://epic.iarc.fr/>).

Les références entre
crochets renvoient à la
Bibliographie générale
p. 51.

D'autres cohortes de très grande dimension sont actuellement à un stade de mise en place ou de préparation avancée

Suède : la cohorte LifeGene prévoit d'inclure 500 000 sujets âgés de 0 à 45 ans, grâce à un échantillonnage à deux niveaux : échantillon aléatoire de personnes de 18 à 45 ans, puis inclusion de leurs enfants et des membres adultes de leur famille ; les inclusions ont commencé en 2010 (<http://www.lifegene.se/In-english/>).

Pays-Bas : la cohorte LifeLines repose également sur un échantillonnage à deux niveaux : échantillon aléatoire de personnes âgées de 25 à 50 ans, puis inclusion des membres de leur famille (partenaires, parents et enfants). Plus de 60 000 sujets sont déjà inclus, et au total ce sont 165 000 participants qui sont attendus (<http://lifelines.nl/>).

France : la cohorte Constances vise à réunir un échantillon représentatif de 200 000 participants âgés de 18 à 69 ans au moment de l'inclusion qui a lieu dans des centres d'examen de santé de la Sécurité sociale ; les premières inclusions ont commencé début 2012 (<http://www.constances.fr/>).

Allemagne : la cohorte GeNatCo (German National Cohort) prévoit d'inclure un échantillon de 200 000 participants âgés de 20 à 70 ans ; le début des inclusions est prévu en 2013 (http://www.nationale-kohorte.de/informationen_en.html).

Malgré certaines différences, ces grandes cohortes en population présentent beaucoup de caractéristiques communes.

Elles sont centrées sur des thèmes d'intérêt général : les cohortes d'adultes s'intéressent particulièrement aux maladies chroniques et dégénératives fréquentes : cancer, maladies cardio-vasculaires et métaboliques, psychiatrie, démences, etc. Les « cohortes de naissance », qui suivent des enfants depuis la vie *in utero*, sont centrées notamment sur les facteurs liés aux différents stades du développement. Dans tous les cas, l'étude de la susceptibilité génétique à développer des maladies (ou en être protégé) est très présente (voire essentielle pour certaines cohortes), et le développement de biomarqueurs de détection précoce de pathologies est privilégié. De nombreux facteurs sont pris en compte, qu'ils soient de nature personnelle et familiale, environnementale et professionnelle, sociale, biologique et physiologique, psychologique, comportementale (alimentation, exercice physique, tabac, alcool...). L'analyse des inégalités sociales et territoriales de santé et de leurs déterminants est également présente dans plusieurs cohortes, de même que celle des consommations de soins et leur coût.

Elles incluent des recueils de données multiples reposant sur des techniques diversifiées : questionnaires, entretiens, examen médical, appariement à des bases de données nationales, et plusieurs cohortes recueillent des données de questionnaire par Internet. Certaines prévoient des investigations complémentaires spécia-

lisées sur des sous-ensembles de participants : ainsi la cohorte allemande GeNatCo prévoit la réalisation d'IRM corps entier pour 40 000 sujets. Enfin, toutes les cohortes mettent en place des biobanques associées, destinées à stocker des échantillons biologiques divers (ADN, sérum, cellules, selles...) pendant une très longue durée pour permettre ultérieurement des analyses biologiques de marqueurs nouveaux.

Du fait de la disponibilité de données nombreuses et diversifiées sur de très importants échantillons, la plupart de ces cohortes sont gérées comme des infrastructures pratiquant une large ouverture vers la communauté de recherche, notamment sous forme d'appels publics à projets permettant ainsi à des chercheurs extérieurs de bénéficier d'un accès aux données collectées.

Enfin, certaines cohortes, comme la cohorte française Constances, sont constituées d'échantillons représentatifs de la population générale, permettant ainsi la production d'indicateurs de santé destinés aux autorités de santé publique.

La nécessité de la mise en commun de données de différentes cohortes

Lorsqu'il s'agit d'analyser les maladies complexes les plus fréquentes pour étudier des relations étiologiques complexes ou des caractères quantitatifs liés à la maladie, même les plus grandes études ne génèrent pas suffisamment de cas. Il devient alors souvent nécessaire de mettre en commun les données de plusieurs grandes cohortes, comme c'est devenu la règle pour les études de génomique dans le cadre de consortiums de recherche. La mise en commun de données à grande échelle ne concerne évidemment pas uniquement les études de génétique, mais toute l'épidémiologie est concernée pour des raisons de puissance statistique. Les études internationales comparatives sur les services de santé, les déterminants sociaux de la santé ou les habitudes alimentaires, réunissant des données de cohortes de plusieurs pays, sont également indispensables pour réduire les biais potentiels découlant de l'accès à des ensembles de données restreints et spécifiques d'une population.

Ces dernières années se sont mises en place des collaborations internationales destinées à faciliter les mises en commun de données de cohortes en population, notamment grâce à une harmonisation aussi étroite que possible des données recueillies. En effet, pour que la mutualisation des données soit possible et éviter de fastidieux et coûteux recodages des données et une perte d'information, il faut que celles-ci aient une même définition, que les variables enregistrées soient identiques, ou du moins qu'on puisse établir aisément des correspondances. Ainsi à l'échelle internationale, le consortium P³G (Public Population Project in Genomics) promeut l'utilisation d'outils (questionnaires, échelles diverses, etc.) qu'il met à disposition de la communauté scientifique. À l'échelle européenne, le consortium BBMRI (Biobanking and Biomolecular



Resources Research Infrastructure) vise à harmoniser les méthodes de collecte et de conservation de matériel biologique, cette initiative étant relayée en France par l'Infrastructure Biobanques (lire *Coordination et partage de données de cohortes*, p. 29). Récemment s'est constitué au sein du consortium BBMRI, le projet LPC (Large Prospective Cohorts) qui associe une vingtaine de grandes cohortes en population provenant de 13 pays européens (la France y est présente par les cohortes Gazel et Constances), réunissant plus de 2,5 millions de sujets.

Ouverture à la communauté de recherche

Une autre forte tendance accompagne le développement des méga-cohortes : l'ouverture systématique des bases de données des cohortes à la communauté de recherche. En effet, les investissements nécessaires pour la construction et la maintenance des très grandes

cohortes sont tels qu'il n'est plus envisageable d'en limiter l'accès aux équipes qui les mettent en œuvre. Ces méga-cohortes sont maintenant considérées comme des infrastructures de recherche qui, à l'instar d'autres très grands instruments de recherche, comme les télescopes ou les accélérateurs de particules, sont conçues pour répondre à de multiples questions provenant d'équipes diversifiées. Ainsi, en France, la cohorte Constances a été récemment reconnue comme une « *Infrastructure nationale biologie santé* » par le ministère de la Recherche et de l'Enseignement supérieur.

De nombreuses institutions de financement de la recherche en santé ont récemment exprimé leur volonté de promouvoir le partage de données, tout en prenant en compte les aspects scientifiques, éthiques et juridiques, ainsi que la nécessaire valorisation scientifique des équipes et institutions qui ont mis en place et qui gèrent les cohortes [45].