

AVIS 151⁺

Tenir compte des origines
des patients dans les
données de santé

ccne.

COMITÉ CONSULTATIF NATIONAL
D'ÉTHIQUE POUR LES SCIENCES
DE LA VIE ET DE LA SANTÉ

AVIS 151

Tenir compte des origines des patients dans les données de santé

AVIS ADOPTÉ À L'UNANIMITÉ DES VOTES EXPRIMÉS PAR LES MEMBRES DU
COMITÉ CONSULTATIVE NATIONAL D'ÉTHIQUE LORS DU COMITÉ PLÉNIER DU 16
AVRIL 2026

SOMMAIRE

RÉSUMÉ.....	7
1. INTRODUCTION : ENJEUX SCIENTIFIQUES, ÉTHIQUES ET JURIDIQUES AUTOUR DES DONNÉES DE SANTÉ RELATIVES À L'ORIGINE.....	10
1.1. CONTEXTE ET MOTIVATION DU CCNE.....	10
1.2. ANGLE CHOISI, CHOIX MÉTHODOLOGIQUE ET ÉTAPES DE LA RÉFLEXION.....	11
2. DONNÉES DE SANTÉ RELATIVES À L'ORIGINE : DE QUOI PARLE-T-ON ?.....	14
2.1. QUESTIONS DE VOCABULAIRE.....	14
2.1.1. Origine géographique, ethnique, raciale.....	14
2.1.2. Le terme de « race » comme construit social.....	15
2.2. TYPES DE DONNÉES POUR CARACTÉRISER L'ORIGINE.....	16
2.2.1. Données objectives sur les origines nationales et le parcours migratoire.....	16
2.2.2. Données subjectives ou de perception sur la position ethno-raciale.....	16
2.3. ORIGINE ET GÉNÉTIQUE.....	18
2.3.1. Inexistence de races biologiques humaines.....	18
2.3.2. Groupes ethniques et caractéristiques génétiques spécifiques.....	19
3. DILEMME ÉTHIQUE : NÉCESSITÉ ET RISQUES DE PRENDRE EN COMPTE LES ORIGINES DANS LE DOMAINE DE LA SANTÉ.....	21
3.1. DANS LA RECHERCHE EN SANTÉ.....	21
3.1.1. Nécessité de documenter les origines.....	21
3.1.2. Risques associés.....	25
3.2. DANS LA PRÉVENTION ET LE SOIN.....	28
3.2.1. Nécessité de prendre en compte les origines.....	28
3.2.2. Risques associés.....	30
4. RÉGLEMENTATION.....	32
4.1. DONNÉES SENSIBLES : CADRE LÉGAL EN FRANCE ET DANS L'UNION EUROPÉENNE.....	32
4.1.1. Le cadre général : le principe d'universalisme républicain.....	32
4.1.2. complété par un régime d'exception.....	33
4.1.3. Réutilisation des données sensibles et consentement : l'exemple des données génétiques.....	35
4.2. LIMITES ET DÉFAUTS DU SYSTÈME ACTUEL.....	36
4.3. ÉLÉMENTS DE COMPARAISON DANS LE MONDE.....	40
5. FAUT-IL FAIRE ÉVOLUER LE SYSTÈME FRANÇAIS EN MATIÈRE DE DONNÉES DE SANTÉ RELATIVES À L'ORIGINE ?.....	42
5.1. RECONNAITRE ET UTILISER PLEINEMENT LA FLEXIBILITÉ DU CADRE LÉGAL EXISTANT EN MATIÈRE D'USAGE DES DONNÉES SUR L'ORIGINE DANS LA SANTÉ.....	42
5.2. FACILITER L'ACCÈS AUX DONNÉES OBJECTIVES DISPONIBLES.....	43
5.3. INTÉGRER LE PAYS DE NAISSANCE DANS LE SNDS.....	43
5.4. RENFORCER LA FORMATION DES PROFESSIONNELS DE SANTÉ AUX ENJEUX LIÉS À L'ORIGINE : EXPOSITIONS, INÉGALITÉS, DISCRIMINATIONS.....	44

ANNEXES.....	45
ANNEXE 1 : COMPARAISONS INTERNATIONALES.....	45
1. <i>En Amérique : États-Unis et Canada.....</i>	<i>45</i>
2. <i>En Europe : Royaume-Uni, Allemagne, Belgique, Pays-Bas, Suède</i>	<i>47</i>
ANNEXE 2 : COMPOSITION DU GROUPE DE TRAVAIL ET PERSONNES	
AUDITIONNÉES	52

RÉSUMÉ

Il existe deux grands types de données de santé pour prendre en compte l'origine des personnes : des informations dites objectives, relatives à leur origine géographique, et des informations dites subjectives, liées à la position ethnique ou sa perception sociale. Si les premières renseignent sur les déterminants environnementaux ou biologiques, les secondes se concentrent sur les expériences sociales et les inégalités ou discriminations. Les questions que soulèvent ces données sur les origines présentent de nombreuses dimensions : éthiques, méthodologiques, épistémologiques, juridiques et politiques.

Dans le domaine de la recherche médicale et du soin, le recours inconsidéré à des catégories mal définies pour caractériser ces origines peut conduire à des interprétations erronées et renforcer les biais inhérents au caractère socialement construit des différences. À l'inverse, le souci de ne pas discriminer peut avoir pour effet d'affaiblir l'efficacité de la recherche, et dans certains cas du soin, en invisibilisant des différences qui existent.

Il convient donc de considérer avec rigueur les données et catégories utilisées. Des catégories telles que la « race » ou l'« ethnicité » ne constituent pas des outils scientifiquement valides pour expliquer les différences de santé sur une base biologique. Elles peuvent cependant parfois servir pour repérer des inégalités sociales de santé, ou constituer des repères opératoires pour quelques mutations génétiques (drépanocytose, gène BRCA). Mais pour des pathologies complexes telles que le cancer, le diabète, l'asthme ou les maladies cardiovasculaires, elles sont scientifiquement infondées et potentiellement trompeuses. Leur mobilisation non encadrée en recherche expose à des erreurs d'interprétation aux conséquences sérieuses. Par ailleurs, les groupes génétiques ne recourent pas nécessairement les groupes culturels, et les interactions entre génome et environnement s'avèrent plus déterminantes qu'un classement, nécessairement réducteur, qui serait basé sur l'appartenance ethnique.

C'est dans ce contexte que le Comité consultatif national d'éthique (CCNE) a entrepris d'analyser la place que peuvent - ou doivent - occuper les données relatives à l'origine

dans le système de santé. La question centrale est de déterminer dans quelle mesure ces données existent, sont mobilisables et contribuent à améliorer la qualité des soins et l'équité en santé, tout en évaluant les risques de mésusage ou d'atteinte aux droits fondamentaux. Le présent avis s'inscrit dans le prolongement des réflexions que le CCNE conduit depuis plusieurs décennies sur les usages des données de santé et sur les risques inhérents à toute réduction de la personne à certaines de ses caractéristiques : assignation, profilage et stigmatisation. Il s'agit de prendre la pleine mesure de ces risques pour en limiter et en maîtriser la portée.

Pour répondre à ces questions, l'avis procède en quatre temps.

Il s'attache d'abord à clarifier le vocabulaire et la nature des catégories mobilisées. Les termes « origine », « ethnie » et « race » ne recouvrent pas les mêmes réalités selon qu'ils sont employés en biologie, en sociologie ou dans le langage ordinaire. La génétique des populations a établi l'absence de fondement biologique à la notion de race humaine : la quasi-totalité de la variabilité génétique de l'espèce se situe entre individus, et non entre groupes. Ce que les discours courants désignent comme une différence de « race » relève en réalité d'un fait social (le processus par lequel certaines personnes se voient assignées à des catégories construites à partir de marqueurs physiques ou culturels perçus, avec des effets concrets sur leurs trajectoires de vie et de santé). Ce paradoxe - la race est biologiquement inexistante mais socialement opérante - constitue le point de départ nécessaire de toute réflexion sur le sujet.

L'avis examine ensuite le dilemme éthique au cœur du sujet : ces données peuvent être à la fois nécessaires et porteuses de risques. Nécessaires, parce que l'origine géographique informe sur les expositions à certains agents infectieux, sur les effets des parcours migratoires sur la santé, sur les inégalités d'accès aux soins, et parfois sur des particularités génétiques pertinentes pour l'adaptation des traitements. Porteuses de risques, parce que le recours non encadré à des catégories ethniques conduit à biologiser des déterminants qui sont en réalité sociaux, à stigmatiser des groupes et à occulter les véritables causes des inégalités de santé, lesquelles tiennent le plus souvent aux conditions de vie, de logement et de travail.

L'avis analyse ensuite le cadre juridique français. Contrairement à une idée répandue, ce cadre n'interdit pas par principe tout usage de ces données : il l'autorise sous conditions strictes, notamment à des fins de recherche scientifique et de santé publique. Ce régime dérogatoire demeure cependant mal connu, ce qui conduit chercheurs et soignants à pratiquer une autocensure, au détriment de la recherche des inégalités et des politiques publiques susceptibles d'y répondre. Ainsi une donnée essentielle pour repérer ces inégalités, comme le pays de naissance, pourtant autorisée par le cadre réglementaire, n'est pas retenue dans certaines grandes bases de données, à commencer par le Système National des Données de Santé.

L'avis formule enfin quatre recommandations. La première invite à mieux faire connaître la flexibilité du cadre légal existant, afin de lever les obstacles injustifiés à un usage raisonné de ces données. La deuxième recommande de faciliter l'accès aux données objectives sur l'origine géographique et le parcours migratoire. La troisième propose d'intégrer la variable « pays de naissance » dans le SNDS (techniquement réalisable et juridiquement compatible avec le droit en vigueur). La quatrième plaide pour renforcer la formation des professionnels de santé aux enjeux liés à l'origine de leurs patients afin de mieux appréhender les expositions, les déterminants sociaux et les biais implicites susceptibles d'altérer la qualité et l'équité des soins.

1. INTRODUCTION : ENJEUX SCIENTIFIQUES, ÉTHIQUES ET JURIDIQUES AUTOUR DES DONNÉES DE SANTÉ RELATIVES À L'ORIGINE

1.1. CONTEXTE ET MOTIVATION DU CCNE

Les débats contemporains autour de l'usage des données de santé mettent en évidence une tension éthique centrale : les informations relatives à l'origine géographique ou ethnique peuvent à la fois constituer des outils précieux pour améliorer la prise en charge de certaines populations et pour identifier, puis réduire, les inégalités en matière de santé et de soins, tout en étant susceptibles d'alimenter, directement ou indirectement, des logiques de discrimination. Ce dilemme prend une acuité particulière en France, où l'article premier de la Constitution pose un principe d'égalité devant la loi de tous les citoyens « sans distinction d'origine, de race ou de religion »¹.

La difficulté de concilier la nécessité de disposer de données fiables et d'indicateurs pertinents avec l'exigence d'un universalisme républicain revêt ainsi une portée singulière. La question devient alors celle de savoir si l'intérêt clinique ou de santé publique peut légitimement l'emporter sur le risque de stigmatisation et plus généralement si cette tension peut être surmontée.

Le Comité consultatif national d'éthique (CCNE) a souhaité analyser de manière précise la place que peuvent – ou doivent – occuper les données relatives à l'origine dans le système de santé. Il s'agit de déterminer dans quelle mesure ces données existent, sont mobilisables et contribuent à améliorer la qualité des soins et l'équité en santé, tout en évaluant les risques de mésusage ou d'atteinte aux droits fondamentaux.

¹ « La France est une République indivisible, laïque, démocratique et sociale. Elle assure l'égalité devant la loi de tous les citoyens sans distinction d'origine, de race ou de religion. », Constitution de la Ve République française, art. 1er, promulguée le 4 octobre 1958. Version en ligne consultée le 03/02/2026 ; https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000019240997/2023-10-18. On s'étonne parfois, aujourd'hui, que la Constitution évoque le terme de race. Mais il faut se rappeler que son article 1^{er} puise sa source dans l'alinéa 1^{er} du préambule de la constitution de 1946 écrite au lendemain d'une guerre marquée par le combat contre le nazisme et ses théories raciales.

L'émergence récente de débats publics et scientifiques autour de ces enjeux signale l'existence de zones d'ombre – juridiques et méthodologiques – qui appellent une réflexion éthique approfondie.

Le présent avis s'inscrit dans le prolongement des réflexions du Comité consultatif national d'éthique sur les usages des données de santé et sur les risques inhérents à toute réduction de la personne à certaines de ses caractéristiques. Dès l'avis n°4 du 6 mai 1985², sur les registres médicaux le CCNE soulignait déjà que le respect de la personne humaine peut être menacé lorsque celle-ci se trouve « enfermée dans son passé individuel et familial par les renseignements accumulés et traités ». Plus récemment, l'avis n°130 sur les données massives en santé du 29 mai 2019³ énonçait cet avertissement : « l'autonomie de chacun s'accroît mais la connaissance très précise des personnes et de leur état de santé induit le risque d'un profilage ; celui-ci met en cause la protection de la vie privée et peut aboutir à la stigmatisation de personnes ou de groupes ». Enfin, dans son avis n°143 du 16 février 2023⁴ consacré aux plateformes de données de santé, le Comité a souligné que les liens entre certaines maladies et les caractéristiques ethniques des patients soulèvent des questions dont l'importance justifie qu'elles soient traitées dans un avis ultérieur. La réflexion conduite par le CCNE dans le présent avis s'inspire de ces réflexions passées qui ont mentionné les risques associés à toute réduction d'un être humain à certaines de ses caractéristiques : assignation, profilage et stigmatisation. Il s'agit donc de bien considérer ces risques pour en limiter et en maîtriser la portée.

1.2. ANGLE CHOISI, CHOIX MÉTHODOLOGIQUE ET ÉTAPES DE LA RÉFLEXION

Cet avis vise d'abord à clarifier le dilemme éthique sous-jacent, à savoir la tension centrale entre la production de connaissances nécessaires pour réduire les inégalités de santé et la protection des personnes contre les risques de discrimination, puis à dresser un état des lieux du cadre juridique et institutionnel français pour examiner

² Comité consultatif national d'éthique (CCNE). 1985. Avis n° 4 : Procréation, parenté et filiation. Paris. 10 p. <https://www.ccne-ethique.fr/sites/default/files/2021-02/avis004.pdf>.

³ Comité consultatif national d'éthique (CCNE). Mai 2019. *Avis n° 130 : Données massives et santé : une nouvelle approche des enjeux éthiques*. Paris. https://www.ccne-ethique.fr/sites/default/files/2021-02/avis_130.pdf

⁴ Comité consultatif national d'éthique (CCNE) et Comité national pilote d'expertise en nutrition (CNPEN). 2023. *Plateformes de données de santé : enjeux d'éthique*. Février 2023. Paris. 73 p. https://www.ccne-ethique.fr/sites/default/files/2023-05/CCNE-CNPEN_GT-PDS_avis_final27032023.pdf.

dans quelle mesure il prend en compte les enjeux éthiques en santé et d'identifier les pistes permettant de mieux les intégrer.

La démarche méthodologique repose sur une combinaison de trois niveaux d'analyse :

- **Éthique et finalités légitimes** : identification des principes en tension, notamment le principe de non-discrimination, et des finalités légitimes pouvant justifier la collecte de données, telles que l'amélioration de la prise en charge, la surveillance épidémiologique ou la réduction des inégalités de santé.
- **Cadre juridique et institutionnel** : examen du droit existant en matière de protection contre les discriminations liées à l'origine, de protection des données personnelles et de règles applicables aux informations de santé ; déclinaison de ce cadre dans le champ spécifique des données de santé pour identifier ce qui est autorisé, ce qui est interdit, et les marges d'interprétation ou d'incertitudes.
- **Pertinence scientifique et usages pratiques** : évaluation de l'existence, de la qualité et de la définition des données relatives à l'origine, ainsi que de leur collecte et mobilisation possibles pour la pratique clinique et la santé publique. Il s'agit d'analyser comment ces données sont utilisées dans la prévention et la prise en charge, en particulier pour certaines pathologies pour lesquelles l'origine présente un intérêt scientifique reconnu, tout en mesurant les risques de mésinterprétation ou d'usage discriminatoire.

Le présent avis comporte quatre étapes. La première consiste à définir la nature des données en question, tant en France qu'à l'étranger, et à clarifier le vocabulaire utilisé.

La deuxième étape porte sur les dilemmes éthiques que soulève la prise en compte des origines dans le domaine de la santé. Il s'agira d'examiner les besoins et les justifications de leur utilisation dans la recherche et le soin, mais également les risques qui y sont associés.

La troisième étape examine la réglementation en vigueur en France, en analysant le cadre légal applicable aux données sensibles et la façon dont il est utilisé.

Enfin, cette réflexion aboutit à la formulation de recommandations destinées à guider l'usage éthique et raisonné des données de santé selon l'origine en vue de concilier la nécessité scientifique et médicale avec la protection des droits individuels.

2. DONNÉES DE SANTÉ RELATIVES À L'ORIGINE : DE QUOI PARLE-T-ON ?

Se poser la question de l'usage de données de santé qui renseignent sur l'origine des personnes suppose, en premier lieu, de clarifier le vocabulaire mobilisé ainsi que la nature des informations utilisées pour caractériser cette origine.

2.1. QUESTIONS DE VOCABULAIRE

Le vocabulaire mobilisé pour désigner les différences d'origines est pluriel et prête fréquemment à polémique. Le panel de termes utilisés (« origine », « ethnie », « race » ou « groupe ethno-racial ») renvoie de fait à des représentations différentes selon les disciplines mais aussi les contextes culturels et nationaux.

Ces notions sont chargées d'histoire et d'usages différenciés entre la recherche, le droit, la statistique et le langage ordinaire⁵.

2.1.1. ORIGINE GÉOGRAPHIQUE, ETHNIQUE, RACIALE

Le terme « origine » recouvre un ensemble de dimensions qui peuvent être de nature géographique, nationale, culturelle ou « perçue ».

- *L'origine géographique* renvoie à des éléments factuels : pays de naissance de l'individu ou de ses ascendants, la nationalité actuelle ou passée, ou encore le parcours migratoire.
- *L'origine ethnique* désigne une appartenance culturelle et sociale fondée sur la référence à un patrimoine commun (langue, religion, histoire, coutumes) et sur un sentiment subjectif d'identité partagée. Ce concept, issu de l'anthropologie et de la sociologie, n'a pas de fondement biologique⁶.

⁵ Comme l'a mis en évidence l'audition d'Ute Kalender, membre du Comité allemand de bioéthique, rappelant que le terme allemand de « Rasse » était inutilisable pour d'évidentes raisons historiques, alors que le mot anglais « race » conservait une acceptabilité relative dans le discours scientifique.

⁶ De nombreux travaux en anthropologie, génétique et sciences sociales soulignent que le concept d'origine ethnique n'a pas de fondement biologique, mais relève d'une construction sociale et culturelle : les différences génétiques humaines ne se regroupent pas en sous-espèces biologiques correspondant aux catégories sociales d'ethnicité ou de race, et ces dernières sont employées plutôt comme des « repères » pour des facteurs sociaux ou environnementaux qu'en raison de bases biologiques objectives, voir par exemple : "Summary." National Academies of Sciences, Engineering, and Medicine. 2025. Rethinking Race and Ethnicity in Biomedical Research. Washington, DC: The National Academies Press. doi: 10.17226/27913

L'expression « ethno-raciale », employée par certains chercheurs⁷, vise à articuler deux dimensions : d'une part, la référence à une appartenance culturelle (« ethno ») et d'autre part, la manière dont les individus sont catégorisés ou perçus socialement à travers des signes visibles ou supposés d'altérité (« racial »)⁸. La frontière entre ces deux notions étant de fait floue, utiliser le terme combiné permet de poser que « l'objet d'étude n'est pas l'ethnicité ou la race dans ce qu'elles ont de spécifique, mais les mécanismes de stratification sociale associés à ces catégories »⁹. Cette approche reconnaît que les processus de racialisation – c'est-à-dire la production de différences perçues comme naturelles – ont des effets concrets sur les trajectoires sociales et sanitaires.

2.1.2. LE TERME DE « RACE » COMME CONSTRUIT SOCIAL

Il n'existe pas de races humaines au sens biologique ou génétique du terme. « La race comme réalité biologique est scientifiquement invalidée depuis la fin des années 1960 »¹⁰. Cette absence de base biologique a été largement confirmée avec les progrès de la génétique des populations et nous y reviendrons dans la section 2.3. Pour autant ce terme conserve une pertinence sociologique pour décrire les processus par lesquels certaines catégories d'individus sont assignées à une différence liée à des marqueurs physiques ou culturels perçus et peuvent subir, de ce fait, des traitements inégalitaires. Le terme renvoie non pas à une essence, mais un processus de catégorisation sociale.

Parler de « race » au sens sociologique ne revient pas à réhabiliter une catégorie biologique, mais à nommer les effets sociaux produits par ces catégorisations parfois inconscientes. Si la génétique a démontré l'inexistence de races humaines, la

⁷ Safi, Mirna. *Les inégalités ethno-raciales*, Paris, La Découverte, 2013.

⁸ Williams DR, Sternthal M. Understanding racial-ethnic disparities in health: sociological contributions. *J Health Soc Behav.* 2010;51 Suppl. Voir aussi : Żuradzki, T., Malinowska, J.K. Ethno-racial categorisations for biomedical studies: the fair selection of research participants and population stratification. *Synthese* 204, 130 (2024). <https://doi.org/10.1007/s11229-024-04769-8>

⁹ Brun, Solène, et Cosquer, Claire. *Sociologie de la race*, Paris, Armand Colin, 2022.

¹⁰ Nathalie Bajos. *La production sociale des inégalités de santé*. Leçon inaugurale au Collège de France. Ed. du Collège de France. 2025 p. 34 ; Adam Hochman, *The Race Illusion. On the Reality of Racialization and the Myth of Race*, Oxford University Press, 2025 ; voir aussi les travaux du généticien Richard C. Lewontin (1972), qui a montré que la variation génétique humaine est plus importante au sein des populations qu'entre elles, rendant obsolète l'idée de races biologiquement distinctes et contribuant à l'abandon du concept de race en biologie. Ces travaux effectués avant les progrès considérables de la génétique moléculaire ont été amplement confirmés depuis par des méthodes beaucoup plus précises (voir Lluis Quintana-Murci, « Les peuples humains », éditions Odile Jacob, 2021).

sociologie rappelle que la « race » continue d'exister comme fait social et produit des effets réels sur les conditions de vie, la santé, l'accès aux soins et les discriminations¹¹.

Cette distinction est essentielle :

- La race biologique est inexistante ;
- La race comme fait social est opérante, car elle structure certaines inégalités ;

On voit certes là un certain paradoxe à considérer la race comme à la fois existante et inexistante, mais le débat public est ainsi construit, de telle sorte qu'il est nécessaire de partir de ce paradoxe pour avancer dans le raisonnement.

2.2. TYPES DE DONNÉES POUR CARACTÉRISER L'ORIGINE

Les données permettant de caractériser l'origine des individus sont multiples et se distinguent selon qu'elles se fondent sur des éléments objectifs, tels que les éléments du parcours migratoire et la nationalité, ou sur des perceptions sociales, telles que l'auto-identification ethno-raciale ou la manière dont l'individu est perçu par autrui.

2.2.1. DONNÉES OBJECTIVES SUR LES ORIGINES NATIONALES ET LE PARCOURS MIGRATOIRE

On considère comme données objectives relatives aux origines la **nationalité** actuelle ou à la naissance, le **pays de naissance** des individus, de leurs parents et grands-parents, ainsi que les données décrivant le **parcours migratoire**, notamment la date d'arrivée dans le pays de résidence. Ces informations permettent de caractériser l'origine géographique et migratoire de manière factuelle mais peuvent aussi masquer la complexité des trajectoires individuelles et produire des interprétations réductrices.

2.2.2. DONNÉES SUBJECTIVES OU DE PERCEPTION SUR LA POSITION ETHNO-RACIALE

Les données relatives à la **position ethno-raciale** reposent sur la perception sociale d'appartenance à un groupe caractérisé par une origine géographique dans l'ascendance, ou des caractères physiques (comme la couleur de peau). Ces éléments peuvent concerner l'auto-identification ou auto-déclaration (la manière dont l'individu

¹¹ D. Fassin et E. Fassin (dir) De la question sociale à la question raciale, ed. La Découverte, 2009.

se définit¹²), la perception par autrui ou hétéro-perception¹³, et l'auto-hétéro-perception (la manière dont la personne pense être perçue par autrui¹⁴).

Cette dernière dimension apparaît particulièrement pertinente pour l'analyse des discriminations, celles-ci étant moins liées à l'origine supposée qu'à la perception sociale des individus. Comme l'ont souligné Solène Brun, sociologue, et Patrick Simon, socio-démographe, au cours de leur audition, les données subjectives, qui reposent essentiellement sur des représentations, permettent de mieux saisir les mécanismes par lesquels certaines personnes sont exposées à des traitements différenciés dans l'accès aux droits, aux soins ou à la qualité de la prise en charge. Elles rendent compte des effets sociaux attachés à des catégories perçues, construites historiquement et socialement.

En France, le recours à ce type de données demeure particulièrement encadré (voir chapitre 4). Leur collecte est assimilée à celle de données sensibles au sens de la loi Informatique et Libertés et du règlement général sur la protection des données (RGPD), et ne peut être autorisée qu'à titre dérogatoire, sous réserve d'un intérêt scientifique ou d'un intérêt général clairement établi¹⁵. Cette contrainte juridique explique que les enquêtes statistiques aient longtemps privilégié des indicateurs indirects (par ex. : pays de naissance ou ascendance migratoire) plutôt que des questions explicites sur l'ethnicité.

Dans ce contexte, certaines enquêtes, *comme Trajectoires et Origines (TeO)*¹⁶ ont introduit, de manière expérimentale et strictement encadrée, des questions d'auto-

¹² L'auto-déclaration peut s'effectuer sur une liste fermée ou ouverte. On trouve ce type de question dans les recensements aux Etats-Unis, au Canada, en Australie et au Royaume-Uni. Voir : Rapport du comité pour la mesure de la diversité et l'évaluation des discriminations (COMEDD) présidé par François HÉRAN. *Inégalités et discriminations - Pour un usage critique et responsable de l'outil statistique*.

¹³ Cette technique est plus rarement utilisée. On la retrouve aux Etats-Unis où les enquêteurs peuvent assigner une catégorie ethno-raciale à la personne interrogée, en se fondant sur le phénotype ou de l'apparence. Une seconde illustration est l'assignation à partir des noms de personnes dont on interprète la consonance ethnique.

¹⁴ Par exemple, l'enquête EU-Midis (European Union Minorities and Discrimination Survey) réalisée en 2008 par la Fundamental Rights Agency/Agence des droits fondamentaux de la Commission européenne s'intéresse aux expériences de discrimination des minorités en ciblant des minorités dans chaque État-membre de l'Union européenne. Le même questionnaire a été utilisé auprès de 23 500 répondants dans les 27 pays. En France, il s'agissait des « Nord-Africains » et des « Subsahariens », selon un principe de résidence en France et d'auto-identification dans le groupe.

¹⁵ Notons que ces méthodes ne résolvent pas pour autant la question de la causalité de la discrimination. Une personne peut déclarer penser être perçue comme appartenant à une minorité, mais en même temps ne pas attribuer à cette appartenance une discrimination ressentie.

¹⁶ L'enquête TeO (Ined, Insee, 2008-2009) qui aborde 5 domaines (école, travail, logement, santé, services) comporte une question sur l'expérience de la discrimination auto-reportée, visant à mesurer, au-delà des discriminations objectives ou objectivées, des discriminations telles qu'elles sont vécues ou ressenties par les personnes qui y sont exposées.

identification pour tenter d'analyser les effets de la racialisation sur les parcours sociaux et de santé.

Le problème éthique qui se pose, notamment pour les chercheurs, réside dans l'utilisation de ces données comme outils d'analyse des discriminations, tout en veillant à ne pas leur attribuer un statut essentialisant ou biologique susceptible de figer les individus dans des catégories identitaires.

2.3. ORIGINE ET GÉNÉTIQUE

2.3.1. INEXISTENCE DE RACES BIOLOGIQUES HUMAINES

La génétique des populations a fermement établi l'absence de fondement biologique à la notion de race humaine. La majorité de la variabilité génétique humaine se situe de fait entre individus, et non entre groupes de populations. Si l'on compare deux individus pris au hasard dans la population mondiale, quelle que soit leur origine géographique, ethnique ou culturelle, on observe environ 0,1 % de différence génétique entre eux, soit près de trois millions de variations sur les trois milliards de bases que compte le génome humain¹⁷. La majorité de ces différences n'a aucune incidence fonctionnelle : elles n'affectent ni les traits phénotypiques (c'est-à-dire ceux qui sont apparents), ni la santé. Une minorité seulement influence certaines caractéristiques observables, comme la taille ou la pigmentation de la peau, et une fraction encore plus faible peut avoir des implications médicales. À cet égard, il convient de rappeler qu'en dehors de maladies liées à une mutation précise bien caractérisée, les tests génétiques produisent des résultats probabilistes : ils indiquent un risque relatif par rapport à une population de référence, sans jamais permettre de prédire avec certitude l'apparition d'une pathologie chez un individu donné.

La structure de la variation génétique humaine s'explique par l'histoire évolutive de l'espèce : migrations, métissages successifs, et diffusion continue des variants au fil du temps. Ainsi les distributions de certaines de ces variations permettent aux généticiens des populations de reconstruire l'histoire des populations humaines, de leurs migrations et de leurs mélanges. Seule une petite partie de ces différences est responsable des caractères apparents utilisés dans la construction sociale des "races" (par exemple, couleur de la peau, des cheveux, des yeux etc.). L'apparence d'une

¹⁷ <https://www.pasteur.fr/fr/journal-recherche/dossiers/genetique-evolutive-humaine-diversite-genetique-est-richeesse>. Voir aussi Lluis Quintana-Murci, « Le peuple des humains », Odile Jacob 2021

personne résulte d'ailleurs de la combinaison de multiples variants génétiques dont les interactions sont complexes et imparfaitement connues¹⁸. Les limites de distribution des variants génétiques intervenant dans les caractères "visibles" ne correspondent pas exactement à celles des groupes de populations qui ne sont jamais complètement homogènes. De plus, la grande majorité des différences génétiques entre individus est indépendante des variations entre populations et des caractères "visibles". Les connaissances de la génétique contredisent donc toute tentative de définir des « races » humaines au sens biologique du terme. Dans une définition stricte, une race suppose une homogénéité génétique interne forte et une nette différenciation entre groupes, ce que l'on observe chez certaines espèces animales domestiques (par exemple entre un caniche et un berger allemand) mais pas chez l'être humain.

Ce rappel scientifique fonde une exigence éthique majeure, en ce qu'il permet de prévenir toute forme de justification biologique des inégalités ou des discriminations.

2.3.2. GROUPES ETHNIQUES ET CARACTÉRISTIQUES GÉNÉTIQUES SPÉCIFIQUES

La notion de groupe ethnique, mobilisée dans le langage courant ou dans certaines recherches, renvoie comme dit plus haut à des critères culturels, linguistiques, historiques ou symboliques (langue partagée, mémoire commune, sentiment d'appartenance). Même si l'étude de la distribution de certains marqueurs génétiques permet d'observer des corrélations statistiques avec celles de traditions culturelles, comme la langue, ces éléments ne correspondent à aucune homogénéité génétique identifiable. De même que la race, un groupe ethnique n'est ni une entité biologique, ni une entité génétique.

Cependant, certaines situations particulières peuvent conduire à une fréquence élevée de certains variants génétiques dans des régions géographiques définies, en lien avec des facteurs historiques ou géographiques. Par exemple, certaines mutations peuvent être maintenues dans certaines régions du globe lorsque, malgré leurs effets délétères, elles confèrent à leurs porteurs une certaine résistance à des maladies infectieuses particulières. Ainsi la mutation HbS responsable de la drépanocytose, présente à l'état hétérozygote (un seul des chromosomes porte la mutation) chez près

¹⁸ Manga P, Loftus S. Genetics of Skin, Hair, and Eye Color in Human Pigmentation Disorders. *Ann Hum Genet.* 2025 Sep;89(5):305-320. doi: 10.1111/ahg.70003.

de 20 % des personnes dans certaines régions d'Afrique centrale, illustre une adaptation locale : cette mutation confère une résistance partielle au paludisme, ce qui explique sa sélection dans certaines zones géographiques touchées par cette maladie. Dans d'autre cas, lorsque des populations sont issues d'un petit groupe de personnes et ont connu une endogamie prolongée, certaines mutations rares ont atteint des fréquences plus élevées : c'est le cas des Québécois d'ascendance ancienne. Encore plus près de nous, on a pu démontrer de façon frappante que la prévalence du rétrécissement aortique calcifié est significativement plus élevée dans quelques départements de l'Ouest de la France et que les personnes affectées descendent toutes d'un très petit nombre d'ascendants qui se retrouvent tous dans un rayon de 30km. Cette pathologie est associée à une mutation génétique qui s'est transmise entre proches, ce qui s'explique par des pratiques matrimoniales, par les histoires des migrations – ou plus exactement de non-migrations - et même par les guerres de Vendée qui ont eu des effets directs sur la structure de la géographie locale. Il peut donc y avoir une inscription très locale du génome¹⁹.

Ces particularités n'impliquent pas que les groupes concernés forment des entités génétiques homogènes. Elles traduisent des différences de fréquence, non de nature : les mêmes variants (ou allèles) peuvent exister partout, mais à des degrés divers. La diversité humaine doit donc être comprise comme un continuum, où les variations sont graduelles et influencées par des facteurs multiples : biologiques, environnementaux, sociaux et culturels.

La génétique des populations invite ainsi à dépasser les classifications ethniques, à éviter toute essentialisation des groupes humains et à promouvoir une approche fine et contextualisée des variations génétiques, articulée avec les déterminants environnementaux et sociaux de la santé.

¹⁹ Karakachoff, Matilde, Nicolas Duforet-Frebourg, Floriane Simonet, et al. 2015. "Fine-Scale Human Genetic Structure in Western France." *European Journal of Human Genetics* 23 (6): 831–36. <https://doi.org/10.1038/ejhg.2014.175>. Et Lindzy Tossé, Thèse de doctorat, *Fabriquer une population locale pour la médecine génomique – Etude sociohistorique et épistémologique des conditions de constitution de la banque d'ADN du Grand Ouest (1983-2017)*, sous la direction du Prof. Richard Redon et de Catherine Bourgain, Université de Nantes et Cermes 3, Inserm. 2023

3. DILEMME ÉTHIQUE : NÉCESSITÉ ET RISQUES DE PRENDRE EN COMPTE LES ORIGINES DANS LE DOMAINE DE LA SANTÉ

L'usage de données relatives à l'origine des individus en santé publique et en recherche médicale soulève un dilemme éthique : il peut être nécessaire de les considérer pour comprendre certaines maladies, mieux cibler les interventions et lutter contre les inégalités, mais leur utilisation comporte des risques de réification et de discrimination. Nous explicitons dans cette partie la nature de ce dilemme éthique, tout d'abord en recherche, puis dans le domaine de la prévention et soin.

3.1. DANS LA RECHERCHE EN SANTÉ

3.1.1. NÉCESSITÉ DE DOCUMENTER LES ORIGINES

a) Pour prendre en compte la diversité des expositions environnementales

L'origine géographique ou l'histoire migratoire d'un individu peut influencer son exposition à différents facteurs environnementaux. Par exemple, les contacts avec certains agents infectieux, les conditions sanitaires et alimentaires ou encore les caractéristiques socio-économiques varient selon les régions et les contextes. Ces expositions induisent des réponses biologiques et ne pas les prendre en compte peut conduire à des interprétations biaisées des données épidémiologiques, notamment dans l'étude des interactions gènes-environnement²⁰.

La diversité de ces facteurs environnementaux, des modes de vie, de l'alimentation ou de l'exposition aux pathogènes explique pourquoi certaines maladies apparaissent plus fréquemment dans certains contextes. En recherche, collecter ces éléments est donc indispensable pour comprendre les déterminants de la santé au-delà de la biologie individuelle, comme le montrent les variations de réponses immunitaires et de vulnérabilité au Covid-19 entre populations exposées à des environnements et antécédents différents²¹. Il est également essentiel de prendre ces facteurs en compte

²⁰ Wild, C. P. (2005). Complementing the genome with an exposome: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* ; Miller, G. W., & Jones, D. P. (2014). The nature of nurture: refining the definition of the exposome. *Toxicological Sciences*, 137(1), 1–2 ; Hunter, D. J. (2005). Gene–environment interactions in human diseases. *Nature Reviews Genetics*, 6(4), 287–298.

²¹ Rotival, M., & Quintana-Murci, L. (2024). Environmental variation and genetic diversity contribute to population differences in immune responses to SARS-CoV-2 and COVID-19 risk. *Genes & Immunity*, 25, 338–340.

pour comprendre comment les réponses aux traitements peuvent varier selon l'origine géographique, sociale ou l'ascendance, ce qui peut influencer la sécurité et l'efficacité des médicaments et dispositifs médicaux.

b) Pour prendre en compte les spécificités génétiques dans la recherche biomédicale

La prise en compte des origines dans le cadre de la recherche biomédicale répond avant tout à une exigence de validité scientifique. Si l'espèce humaine se caractérise par une grande homogénéité génétique, certaines variations rares mais cliniquement significatives sont distribuées de manière inégale selon les histoires migratoires, les contextes géographiques et les dynamiques de brassage génétique des populations. Ces variations peuvent influencer la prévalence de certaines pathologies, la réponse aux traitements ou encore la tolérance aux médicaments. L'un des risques identifiés est celui d'une recherche clinique fondée sur des populations insuffisamment diversifiées, conduisant à des résultats partiels, voire biaisés. Lorsque les essais cliniques ne prennent pas en compte la diversité des origines des personnes, ils ne permettent pas de repérer la variabilité des effets des médicaments selon les individus et les groupes. Cette limitation peut se traduire par une efficacité moindre des traitements, des effets indésirables non anticipés ou une inadaptation thérapeutique pour certaines populations.

Dans ce contexte, l'absence de diversité dans les cohortes de recherche constitue un biais méthodologique, les résultats produits risquant de ne pas être généralisables à l'ensemble de la population, affectant tant l'évaluation de l'efficacité que celle de la sécurité des interventions médicales. La prise en compte raisonnée des origines contribue ainsi à améliorer la conception des essais cliniques, à affiner l'interprétation des résultats et à réduire les angles morts de la production de connaissances biomédicales. Elle impose toutefois un certain nombre d'exigences : diversifier effectivement les populations incluses dans les essais cliniques, informer de manière transparente sur les différences d'efficacité ou de risques observées, et éviter toute forme d'essentialisation.

Des programmes de recherche structurants, tels que France Médecine Génomique, illustrent l'intérêt de ces approches lorsqu'elles sont rigoureusement encadrées. En s'appuyant sur le génotypage de volontaires, ces dispositifs permettent d'identifier des

variations génétiques parfois régionales, de comparer les profils observés et de mieux comprendre certaines maladies rares, contribuant ainsi à l'optimisation des stratégies diagnostiques et thérapeutiques, sans assigner les individus à des catégories figées.

L'impact de ces différences est particulièrement visible en pharmacologie : les individus ne métabolisent pas de manière homogène les substances actives, en raison de variations génétiques affectant notamment les enzymes impliquées dans le métabolisme des médicaments. Ces différences peuvent modifier l'efficacité thérapeutique ou augmenter le risque d'effets secondaires, comme cela a été observé pour certaines classes de médicaments, notamment les antidépresseurs et certains traitements de la tuberculose. Conscients de ces enjeux, certains États, comme le Japon, exigent la réalisation d'études spécifiques sur leur population avant d'autoriser la commercialisation de médicaments développés à partir de données issues majoritairement de populations étrangères. De même, les notices de certains médicaments mentionnent déjà des variations d'efficacité ou de tolérance selon l'origine des patients, comme c'est le cas pour le périndopril – prescrit dans des cas d'hypertension artérielle, de maladie coronaire stable et d'insuffisance cardiaque – ou l'allopurinol – prescrit dans le traitement et la prévention de la goutte ainsi que dans d'autres pathologies liées à l'excès d'acide urique – comme a pu le souligner la professeure de droit privé Elsa Supiot, lors de son audition.

Dans cette perspective, la prise en compte des origines ne vise pas à assigner les individus à des catégories figées, mais à reconnaître la diversité des trajectoires biologiques et contextuelles afin de produire des connaissances médicales plus robustes, plus sûres et plus équitables. Elle s'inscrit ainsi dans les objectifs d'une recherche biomédicale soucieuse à la fois de rigueur scientifique et d'équité en santé.

c) Pour mieux comprendre les mécanismes conduisant aux inégalités sociales de santé

La recherche en santé publique s'attache à analyser comment les conditions de vie et les facteurs sociaux agissent sur la santé des personnes, en particulier via la prévention et l'accès aux soins. Les données qui concernent l'origine sont nécessaires dans ce type de recherche. Ce sont en effet des données sociales, et à ce titre leur connaissance est indispensable pour une compréhension fine des mécanismes qui conduisent à ces inégalités dans la prévention, l'accès aux soins, ou le traitement. Le

pays où on est né, la nationalité, la date d'arrivée dans le pays d'accueil, sont des données qui renseignent sur la situation sociale de la personne, en complément d'autres données sociales comme le niveau d'éducation, le lieu de vie, la profession, le statut marital, etc. Il a ainsi été montré que la prévention du cancer du sein et du col de l'utérus est nettement moins développée en France chez les femmes immigrées que chez les femmes nées en France²², et que ces différences persistent après ajustement sur les caractéristiques socio-démographiques. Autre exemple, les maladies infectieuses comme le VIH ou la tuberculose affectent particulièrement les immigrés africains dans les premières années suivant leur arrivée en France, en lien avec la précarité rencontrée durant cette période d'installation²³. Si la précarité constitue le facteur direct d'exposition aux risques infectieux, le statut d'immigré en est un facteur indirect : il rend plus difficile l'accès à un logement stable et à un titre de séjour, quelle que soit la position professionnelle ou le niveau d'éducation acquis dans le pays d'origine²⁴.

La surmortalité observée chez les immigrés africains lors de la pandémie de Covid²⁵ illustre ces effets : elle résulte d'une surexposition aux risques, liée à leur surreprésentation dans des métiers « de première ligne » (soins, transports, restauration...) doublée d'un accès aux soins plus difficile du fait de leur ségrégation résidentielle dans des quartiers défavorisés avec moins de médecins (par exemple en Seine-Saint-Denis²⁶).

d) Pour mettre à jour de possibles discriminations dans le système de soins

Comme tout phénomène social, celui de l'assignation des personnes à des groupes ethno-raciaux traverse tous les métiers, y compris les métiers de soignants. En effet même les soignants, bien qu'ils soient par leur vocation et leur formation très attachés aux règles éthiques du soin que sont le respect et la justice, et qu'ils puissent eux-mêmes appartenir à des minorités visibles, sont socialement construits et peuvent, de

²² Rondet C, Lapostolle A, Soler M, Grillo F, Parizot I, Chauvin P. Are immigrants and nationals born to immigrants at higher risk for delayed or no lifetime breast and cervical cancer screening? The results from a population-based survey in Paris metropolitan area in 2010. PLoS ONE 2014;9:e87046.

²³ Desgrées du Loû A, Pannetier J, Ravalihasy A, et al. Is hardship during migration a determinant of HIV infection? Results from the ANRS PARCOURS study of sub-Saharan African migrants in France. AIDS 2016;30:64556.

²⁴ Gosselin A, Desgrées du Loû A, Lelièvre É, Lert F, Dray-Spira R, Lydié N. Migrants subsahariens : combien de temps leur faut-il pour s'installer en France ? Popul Sociétés 2016;533.

²⁵ Khlal M et al. Impact of the COVID-19 crisis on the mortality profiles of the foreign-born in France during the first pandemic wave, Social Science & Medicine, Volume 313, 2022, <https://doi.org/10.1016/j.socscimed.2022.115160>.

²⁶ Santé publique France, Inégalités sociales de santé et Covid - Le dossier de La Santé en action n°468, janvier 2025.

façon le plus souvent inconsciente, avoir des préjugés qui peuvent conduire à des traitements différenciés selon l'origine des personnes²⁷²⁸. Il a par exemple été montré que les femmes immigrées africaines, en cas d'hypertension, reçoivent des soins légèrement différents, ce qui peut conduire à retarder chez elles le diagnostic de prééclampsie²⁹. Prendre en compte ces biais implicites dans la recherche sur la qualité des soins permet de lutter contre ce risque de traitements différenciés sans raison médicale³⁰.

Ainsi, pour la recherche sur l'équité en santé et pour lutter contre les discriminations, il peut être utile de recueillir des données sur l'origine, y compris des données de perception comme les données sur la couleur de peau ou le groupe ethno-racial, ainsi que les expériences vécues par des personnes issues des minorités.

3.1.2. RISQUES ASSOCIÉS

a) *Biologisation des problèmes sociaux*

Il est essentiel de distinguer corrélation et causalité biologique. Attribuer des inégalités de santé à l'origine ethnique, plutôt qu'aux déterminants sociaux ou environnementaux, relève le plus souvent d'une biologisation des problèmes sociaux et peut renforcer des stéréotypes ou des préjugés. De nombreuses maladies, comme le diabète ou l'obésité, sont multifactorielles, résultant à la fois de prédispositions génétiques et de facteurs sociaux et environnementaux (nutrition, accès aux soins, conditions de vie). Réduire leur prévalence à la seule génétique est trompeur et risque de masquer les inégalités sociales de santé. Une vigilance particulière est donc nécessaire afin d'éviter toute simplification excessive, susceptible de négliger l'analyse des déterminants sociaux et économiques qui structurent ces inégalités

²⁷ Azria E, Sauvegrain P, Anselem O, Bonnet MP, Deneux-Tharaux C, Rousseau A, Richetin J. Implicit biases and differential perinatal care for migrant women: Methodological framework and study protocol of the BiP study part 3. *J Gynecol Obstet Hum Reprod*. 2022;51. doi: 10.1016/J.JOGOH.2022.102340

²⁸ Linard M, Deneux-Tharaux C, Luton D, Schmitz T, Mandelbrot L, Estellat C, Azria E. Differential rates of cesarean delivery by maternal geographical origin: a cohort study in France. *BMC Pregnancy Childbirth*. 2019;19. doi: 10.1186/s12884-019-2364-x

²⁹ Sauvegrain P, Azria E, Chiesa-Dubruille C, Deneux-Tharaux C. Exploring the hypothesis of differential care for African immigrant and native women in France with hypertensive disorders during pregnancy: a qualitative study. *BJOG An Int J Obstet Gynaecol*. Published Online First: 15 May 2017. doi: 10.1111/1471-0528.14658

³⁰ Par exemple, l'obésité et le diabète sont plus fréquents dans les communautés d'origine latino-américaines ou hispanophones du sud de la Californie parce qu'elles vivent plus souvent dans des environnements socialement et économiquement défavorisés. Les déterminants sociaux de la santé expliquent une grande partie des inégalités observées : Cleveland, J.C., Espinoza, J., Holzhausen, E.A. *et al.* The impact of social determinants of health on obesity and diabetes disparities among Latino communities in Southern California. *BMC Public Health* 23, 37 (2023). <https://doi.org/10.1186/s12889-022-14868-1>

Si l'usage de catégories ethniques ou raciales repose sur des constructions sociales dotées d'une réalité symbolique, leur usage en contexte médical doit être envisagé avec prudence et encadré par des considérations éthiques strictes. Leur définition, dépendante des contextes historiques, sociaux et institutionnels, varie selon les cadres d'analyse et les usages qui en sont faits. Lorsqu'elles sont mobilisées sans clarification de leur statut ni des mécanismes sociaux qu'elles recouvrent, ces catégories peuvent conduire à des biais interprétatifs et favoriser des lectures biologisantes de différences qui relèvent en réalité de déterminants sociaux, environnementaux ou historiques³¹.

b) Ethnisation et utilisation politique

L'utilisation de catégories ethniques dans la recherche peut être instrumentalisée à des fins politiques, idéologiques ou discriminatoires, en essentialisant des inégalités qui relèvent de facteurs économiques ou environnementaux³². Le risque, en employant ces catégories dans la santé, peut-être de légitimer des discriminations ou des stigmatisations (en associant une maladie à un groupe par exemple), voire de les percevoir comme des variables biologiques, ce qui contribuerait à retomber dans des logiques de hiérarchisation entre groupes héritées de l'histoire³³. Cela peut renforcer des idées essentialistes de différences intrinsèquement liées à des groupes, même lorsque l'intention sous-jacente est de promouvoir l'équité. La constitution de « groupes à risque » définis sur une base ethnique peut par exemple donner une apparence scientifique à des inégalités ancrées dans des conditions sociales.

c) Essentialisation de la personne dans son origine ethnique

Réduire une personne à son origine ethnique, qu'elle soit auto-déclarée, perçue ou socialement assignée, limite sa liberté individuelle en la confinant à une catégorie qui ne rend pas compte de la forte variabilité génétique interindividuelle mise en évidence par la génomique humaine. De plus, la couleur de peau ne reflète pas de manière

³¹ Filali A, Osei L, Vignier N. Tenir compte des origines dans le raisonnement médical en maladies infectieuses et tropicales ? Un regard critique [Taking origins into account in medical reasoning in infectious and tropical diseases? A critical look]. *Med Trop Sante Int.* 2024 Jan 29;4(1):mts.v4i1.2024.362. French. doi: 10.48327/mts.v4i1.2024.362. PMID: 38846114; PMCID: PMC11151904.

³² McMillan J, Earp BD, Kong WM, *et al.* The wrong word for the job? The ethics of collecting data on 'race' in academic publishing *Journal of Medical Ethics* 2024;50:149-151.

³³ Jessica P Cerdeña, Vanessa Grubbs, Amy L Non, *Racialising genetic risk: assumptions, realities, and recommendations*, *The Lancet*, Volume 400, Issue 10368, 2147 – 2154.

fiable l'origine géographique, surtout dans des pays marqués par une longue histoire d'immigration.

Comme l'a rappelé le généticien Luis Quintana-Murci lors de son audition, la majeure partie de la diversité génétique humaine est intrapopulationnelle : deux individus appartenant à un même groupe peuvent être génétiquement très différents, tandis que les différences observées entre groupes humains tiennent principalement à des variations de fréquence de certains variants, et non à des mutations spécifiques à un groupe donné.

Dans la recherche comme en clinique, il est donc plus pertinent de se concentrer sur le mécanisme biologique sous-jacent, comme une mutation spécifique ou l'activité d'une enzyme, plutôt que de se baser sur des catégories ethniques telles que « caucasien » ou « noir », qui ne rendent pas compte de la variabilité réelle et peuvent renforcer des idées fausses sur les différences entre groupes.

Cette limite apparaît de manière particulièrement claire en pratique clinique. L'exemple de l'isoniazide, utilisé dans le traitement de la tuberculose, est à cet égard éclairant : certains patients sont des acétyleurs rapides et d'autres des acétyleurs lents en fonction de variants génétiques affectant le métabolisme du médicament. Si des statistiques montrent des différences de fréquence de ces profils selon des groupes ethno-raciaux, ces regroupements restent approximatifs et insuffisants.

d) Biais des algorithmes d'Intelligence artificielle.

De nombreux outils d'intelligence artificielle provenant des Etats-Unis utilisent des données ethnoraciales pour entraîner leurs algorithmes. Le recours à ces données peut permettre une meilleure compréhension des mécanismes produisant les inégalités de santé et inversement ne pas les prendre en compte peut engendrer des biais en défaveur des minorités³⁴. Mais la prise en compte mal contrôlée des origines peut aussi engendrer des biais en défaveur de certaines catégories, effets qui s'expliquent par le fait que la base de données d'entraînement de l'algorithme peut elle-même contenir d'importants biais de collecte et de construction défavorisant certains groupes³⁵. L'usage des données sur les origines dans les bases

³⁴ Sjoding, Michael W., Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, et Thomas S. Valley. « Racial Bias in Pulse Oximetry Measurement ». *New England Journal of Medicine* 383, no 25 (17 décembre 2020): 2477-78.

³⁵ Jindal, Atin. n.d. "Misguided Artificial Intelligence: How Racial Bias Is Built Into Clinical Models." *The Brown Journal of Hospital Medicine* 2 (1): 38021. <https://doi.org/10.56305/001c.38021>.

d'entraînements des algorithmes d'intelligence artificielle n'a pas d'effet mécaniques concernant le traitement équitable des minorités, mais les risques sont importants. L'équité dépend des qualités de ces données et non de leur seule présence. C'est pourquoi le CCNE a recommandé de « **Veiller à limiter les biais potentiels liés à la programmation même des algorithmes** », alors que la réglementation concernant le contrôle externe de ces bases est inexistante.³⁶

Se fier à des catégories ethniques pour interpréter ces résultats, c'est perpétuer l'essentialisation de l'individu, alors qu'une approche précise et personnalisée doit combiner génome, environnement et contexte social.

3.2. DANS LA PRÉVENTION ET LE SOIN

La prise en compte des origines dans la prévention et le soin pose des questions analogues à celles de la recherche, mais avec des implications directes pour la santé des individus. Elle interroge en particulier les principes d'égalité, de non-discrimination et la question de la personnalisation des prises en charge. Si certaines caractéristiques génétiques, expositions environnementales ou contextes sociaux peuvent orienter la prévention et adapter les traitements, l'usage de catégories ethniques comporte des risques de biais, de diagnostics erronés et de stigmatisation.

3.2.1. NÉCESSITÉ DE PRENDRE EN COMPTE LES ORIGINES

a) *Pour mieux dépister et soigner en prenant en compte les contextes d'exposition :*

Certaines maladies infectieuses ou chroniques varient selon les populations en raison de facteurs environnementaux ou historiques. L'origine géographique constitue alors un indicateur pertinent d'exposition à certains pathogènes, comme pour la tuberculose, les hépatites ou le VIH, dont la prévalence est plus élevée en Afrique par exemple : une personne venant d'Afrique a plus de risque d'être porteuse d'une de ces infections qu'une personne née en France, ce qui implique une vigilance accrue pour le dépistage de ces maladies dans ce groupe. En revanche, la couleur de peau ou l'ethnicité n'apportent aucune information médicale intrinsèque³⁷. Ainsi, s'agissant des maladies infectieuses et tropicales, il est important de prendre en compte l'origine

³⁶ Diagnostic Médical et Intelligence Artificielle : Enjeux Éthiques. Avis commun du CCNE et du CNPEN, Avis 141 du CCNE, Avis 4 du CNPEN. Novembre 2022, p. 41

³⁷ Filali et al, 2024, op cité

géographique et les parcours de vie des patients pour mieux évaluer l'exposition à certains pathogènes, et adapter la prévention et le suivi médical.

Dans la pratique clinique, connaître la provenance et le parcours de vie des patients permet de détecter plus rapidement des pathologies spécifiques, d'éviter des retards de prise en charge, d'adapter la prévention aux risques réels d'exposition.

b) Pour prendre en compte les contextes sociaux

Au-delà des facteurs biologiques ou géographiques, les différences sociales et culturelles influencent fortement la santé et l'accès aux soins. Précarité, isolement, niveau d'éducation ou expérience de discrimination peuvent créer des obstacles aux traitements ou à la prévention. Par exemple, lors de la pandémie de Covid-19, la vulnérabilité face à l'infection était largement liée aux conditions de logement, de transport, d'exercice professionnel et aux pathologies chroniques non ou mal prises en charge (obésité, HTA, diabète)^{38,39}.

Le risque pour un professionnel de santé est de s'en tenir aux indicateurs biologiques et de ne pas considérer ces conditions sociales. Pourtant, repérer les obstacles sociaux à la prévention et à la prise de traitements et anticiper leurs impacts sur le parcours de soins est essentiel à une prise en charge de qualité. C'est en particulier vrai pour les groupes les plus défavorisés et les personnes récemment immigrées, vulnérables du fait du changement de repères, de la perte de liens, de la méconnaissance du système français. Cela implique que les soignants soient formés aux dimensions sociales et culturelles du soin, et à leurs possibles différences selon les origines, afin de garantir des soins équitables et personnalisés.

c) Pour améliorer la médecine personnalisée et la vigilance pharmacologique

La pharmacovigilance et la médecine de précision impliquent de prendre en compte les variations génétiques individuelles. Certaines variations sont en effet connues pour influencer la tolérance aux médicaments, leur efficacité ou le risque d'effets secondaires. Par exemple, l'enzyme CYP2C19, impliquée dans le métabolisme de plusieurs médicaments comme certains antidépresseurs ou inhibiteurs de la pompe à

³⁸ McGowan VJ, Bambra C. COVID-19 mortality and deprivation: pandemic, syndemic, and endemic health inequalities. *Lancet Public Health*. 2022 Nov;7(11):e966-e975. doi: 10.1016/S2468-2667(22)00223-7. PMID: 36334610; PMCID: PMC9629845.

³⁹ Gosselin A, Warszawski J, Bajos N; EpiCov Study Group. Higher risk, higher protection: COVID-19 risk among immigrants in France-results from the population-based EpiCov survey. *Eur J Public Health*. 2022 Aug 1;32(4):655-663. doi: 10.1093/eurpub/ckac046. PMID: 35478253; PMCID: PMC9341671.

protons, varie d'un individu à l'autre, et certaines mutations sont plus fréquentes dans certaines populations, sans que cela permette de classer les individus de manière absolue⁴⁰.

De même, la tolérance au lactose est liée à des mutations génétiques fréquentes dans certaines populations européennes⁴¹, mais ces variations restent propres à chaque individu et ne définissent pas un groupe homogène. La génétique permet donc d'anticiper certaines maladies et d'adapter les traitements en fonction des profils individuels. La pharmacogénétique vise ainsi à personnaliser les soins sur la base de caractéristiques biologiques précises, au-delà de toute catégorisation ethnique ou sociale. Cette approche contribue à améliorer l'efficacité thérapeutique et à réduire les effets indésirables, tout en respectant la diversité génétique individuelle.

3.2.2. RISQUES ASSOCIÉS

a) *Prisme ethnique inadéquat : mauvais diagnostics, biais implicites*

Pour le clinicien, ce qui importe réellement est le statut métabolique du patient, et l'essentialiser par sa supposée origine peut conduire à des erreurs ou à des traitements inadaptés. En effet, l'utilisation de l'origine ethnique ou de la couleur de peau comme critère au cours de la consultation médicale peut induire des biais implicites et des erreurs diagnostiques si elle remplace une approche analytique rigoureuse (hypothético-déductive), mais ces informations restent utiles dans certains contextes épidémiologiques et pour documenter les discriminations⁴². L'approche doit être individualisée, tenant compte à la fois de facteurs génétiques, environnementaux et socio-économiques⁴³. Cette prise en compte peut aussi conduire à des raccourcis cliniques, à des pratiques différenciées injustifiées ou à une altération de la relation de confiance. Un mauvais usage des critères d'origine dans le soin peut ainsi contribuer à stigmatiser certaines populations, en attribuant à tort des prédispositions ou des

⁴⁰ Par exemple, l'étude des variations génétiques du gène CYP2C19 dans la population japonaise montre comment les spécificités génétiques peuvent influencer le métabolisme des médicaments, soulignant l'importance de la vigilance pharmacologique pour réduire les risques d'effets secondaires : <https://pubmed.ncbi.nlm.nih.gov/16141610/>

⁴¹ Anguita-Ruiz A, Aguilera CM, Gil Á. Genetics of Lactose Intolerance: An Updated Review and Online Interactive World Maps of Phenotype and Genotype Frequencies. *Nutrients*. 2020 Sep 3;12(9):2689. doi: 10.3390/nu12092689. PMID: 32899182; PMCID: PMC7551416.

⁴² Filali, Amel, Lindsay Osei et Nicolas Vignier. 2024. Op. cit.

⁴³ S'appuyer uniquement sur des tendances observées à l'échelle de groupes peut masquer la diversité individuelle et conduire à des conclusions erronées, il est donc essentiel de sortir de l'approche collective des données pour privilégier une analyse personnalisée. Néanmoins, l'usage de l'origine dans le soin peut également conduire à attribuer à tort des prédispositions à un groupe, alors que les déterminants sont multifactoriels et contextuels.

comportements à un groupe, alors que les déterminants réels sont multifactoriels et contextuels.

b) Risques d'inégalité économique et d'injustice dans l'accès aux soins

Ces outils sont précieux pour la prévention et le suivi médical, mais leur utilisation peut aussi générer une discrimination, notamment lorsque certains groupes sont ciblés par des tests plus coûteux ou exclusifs. C'est le cas de l'affaire Myriad Genetics : cette entreprise a développé un test de dépistage des mutations BRCA1 et BRCA2, associées au cancer du sein et fréquentes dans la population ashkénaze⁴⁴. Alors qu'un test similaire était déjà accessible en Europe, Myriad a obtenu un brevet pour des mutations spécifiques aux femmes ashkénazes, rendant le test beaucoup plus cher et créant un risque d'exclusion économique pour ce groupe.

⁴⁴ <https://pubmed.ncbi.nlm.nih.gov/10413426/>

4. RÉGLEMENTATION

La collecte et l'utilisation des données de santé relatives aux origines des personnes s'inscrivent dans un cadre juridique défini par le règlement européen général sur la protection des données (RGPD) et le droit national, en particulier la Constitution et les dispositions de la loi « Informatique et Libertés ». Ce cadre juridique vise à protéger les personnes, tout en soulevant un enjeu éthique majeur d'équilibre entre la protection des libertés individuelles et la lutte contre les inégalités de santé.

4.1. DONNÉES SENSIBLES : CADRE LÉGAL EN FRANCE ET DANS L'UNION EUROPÉENNE

La qualification des données d'origine comme données sensibles impose des garanties fortes, tout en limitant potentiellement leur mobilisation à des fins d'intérêt général.

4.1.1. LE CADRE GÉNÉRAL : LE PRINCIPE D'UNIVERSALISME RÉPUBLICAIN...

Le principe d'universalisme républicain, inscrit à l'article 1er de la Constitution, garantit l'égalité des citoyens sans distinction d'origine, de race ou de religion.

Une décision du Conseil constitutionnel de 2007⁴⁵ autorise néanmoins les études sur la diversité, la discrimination ou l'intégration, à condition qu'elles ne reposent pas sur des critères ethniques ou raciaux et que les données utilisées soient objectives. Les données « objectives » acceptées concernent la nationalité, le pays de naissance ou l'ascendance familiale. Sur le plan juridique, la collecte de la nationalité ou de l'origine géographique ne présente donc pas de difficulté, contrairement aux catégories ethniques qui se heurtent à l'article 1er de la Constitution.

⁴⁵ Conseil constitutionnel. Décision n° 2007-557 DC du 15 novembre 2007 concernant la loi relative à la maîtrise de l'immigration, à l'intégration et à l'asile. Conseil constitutionnel. Décision n° 2007-557 DC du 15 novembre 2007 concernant la loi relative à la maîtrise de l'immigration, à l'intégration et à l'asile. Dans son rapport au président de la République de décembre 2008, le Comité de réflexion sur le préambule de la Constitution présidé par Simone Veil s'exprime ainsi sur cette décision : « Certains ont déduit de cette décision, à la vérité surprenante, que le Conseil avait, de la sorte, complètement fermé la porte au principe même de statistiques permettant de connaître les handicaps dont souffrent les minorités visibles. Le commentaire officiel de la décision aux *Cahiers du Conseil constitutionnel* invite toutefois à une lecture moins radicale, le Conseil n'ayant en fait pas entendu, si l'on se fie à cette explication, s'opposer à la collecte de données objectives telles que le nom, l'origine géographique ou la nationalité antérieure à la nationalité française, ni même au traitement de données subjectives comme celles fondées sur le « ressenti d'appartenance ».

Décision du Conseil constitutionnel n° 2007-557 DC du 15 novembre 2007

Considérant 29.

Les traitements nécessaires à la conduite d'études sur la mesure de la diversité des origines des personnes, de la discrimination et de l'intégration peuvent porter sur des données objectives mais ne sauraient, sans méconnaître le principe énoncé par l'article 1er de la Constitution, reposer sur l'origine ethnique ou la race.

CC – Décision 2007-557 DC, 15 novembre 2007, cons. 29.

Le texte de la décision est accompagné, comme de coutume, sur le site du Conseil constitutionnel, d'un commentaire officiel qui éclaire la portée de la décision en s'inspirant des débats qui ont conduit à son adoption. S'agissant de ce point, le commentaire est le suivant :

Le Conseil « a jugé que ces traitements peuvent porter sur des données objectives mais ne sauraient, sans méconnaître le principe énoncé par l'article 1er de la Constitution, reposer sur l'origine ethnique ou la race. Ces données objectives pourront, par exemple, se fonder sur le nom, l'origine géographique ou la nationalité antérieure à la nationalité française. Le Conseil n'a pas jugé pour autant que seules les données objectives pouvaient faire l'objet de traitements : il en va de même pour des données subjectives, par exemple celles fondées sur le « ressenti d'appartenance ». En revanche, serait contraire à la Constitution la définition, a priori, d'un référentiel ethnoracial. Telle est la limite constitutionnelle qui a été posée par la décision du 15 novembre 2007. Cette décision s'inscrit ainsi dans la droite ligne de la jurisprudence du Conseil constitutionnel relative à l'article 1er de la Constitution. Le Conseil veille au respect des principes annoncés dans la Déclaration de 1789 qui interdisent tout communautarisme ».

4.1.2. COMPLÉTÉ PAR UN RÉGIME D'EXCEPTION...

Les données relatives aux origines font partie des données sensibles dont le traitement est interdit par le RGPD : « *Le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie*

sexuelle ou l'orientation sexuelle d'une personne physique sont interdits. » (RGPD, article 9). Cependant, ce règlement prévoit plusieurs exceptions, notamment⁴⁶ :

- Consentement explicite de la personne concernée pour une ou plusieurs finalités déterminées.
- Sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne, lorsque la personne concernée ne peut pas donner son consentement.
- Médecine préventive ou du travail, évaluation de la capacité de travail, diagnostic, soins ou gestion des systèmes de santé ou de protection sociale, fondés sur la loi ou un contrat avec un professionnel de santé et sous garanties spécifiques
- Santé publique : protection contre les menaces sanitaires graves, maintien de normes de qualité et de sécurité des soins ou des médicaments/dispositifs médicaux, avec garanties appropriées.
- Archivage dans l'intérêt public, recherche scientifique ou historique, ou statistiques, avec mesures spécifiques, proportionnalité et respect de l'essence du droit à la protection des données.

En s'appuyant sur ce cadre du RGPD, la Commission Nationale Informatique et libertés (CNIL) autorise donc, par exception, la collecte et le traitement de données à caractère ethno-racial ainsi que des données génétiques, lorsque cette collecte et ce traitement sont à des fins de recherche ou de soin. Cette autorisation est donnée au cas par cas en appréciant les critères suivants :

- *Démonstration claire de la nécessité scientifique* de collecter et traiter ces données.
- *Principe de minimisation des données* : Les données collectées doivent être strictement nécessaires au regard de l'objectif scientifique poursuivi.
- *Responsabilisation des acteurs* : Le RGPD met l'accent sur la responsabilité du responsable de traitement⁴⁷.

⁴⁶ Article 9 du RGPD ayant pour objet le traitement portant sur des catégories particulières de données à caractère personnel.

⁴⁷ Audition de la CNIL : Les refus d'autorisation concernent surtout les dossiers incomplets ou mal justifiés.

4.1.3. RÉUTILISATION DES DONNÉES SENSIBLES ET CONSENTEMENT : L'EXEMPLE DES DONNÉES GÉNÉTIQUES

La réutilisation de données sensibles, notamment génétiques ou relatives à l'origine, soulève des questions complexes de consentement d'information des personnes, du contrôle des usages secondaires et d'encadrement juridique. Les analyses génétiques peuvent en effet révéler non seulement l'histoire biologique de l'individu, mais aussi celle des parents, avec des implications inattendues sur l'origine géographique ou l'ascendance. Ces informations sont utiles pour la médecine personnalisée et la prévention de certaines pathologies (drépanocytose, thalassémie, mucoviscidose,) ainsi qu'à la pharmacogénétique, mais elles sont sensibles sur le plan éthique. Leur réutilisation doit donc s'accompagner de garde-fous : interdiction de certaines finalités (par ex. : calcul de risque santé par une compagnie assurance à partir de données génétiques), information claire des participants et encadrement scientifique précis des projets.

Le cadre juridique français en la matière permet la réutilisation des données collectées, y compris des données sensibles telles que les données génétiques ou relatives à l'origine, sous réserve d'une information transparente et d'un consentement approprié au moment de la première étude⁴⁸. En effet, l'information et le consentement ne concernent pas seulement le moment initial de la collecte des données, mais aussi toute réutilisation pour une finalité nouvelle ou différente de celle pour laquelle elles ont été collectées. Pour qu'un traitement ultérieur des données soit licite, la finalité de ce nouveau traitement doit être compatible avec celle initialement communiquée aux personnes concernées, ou la personne doit être à nouveau informée et, si nécessaire, donner son consentement explicite à cette nouvelle finalité⁴⁹. Cela est particulièrement important pour les données sensibles, qui révèlent par exemple des informations génétiques pouvant concerner non seulement l'individu, mais aussi ses proches, ou des informations sur l'origine géographique ou l'ascendance.

⁴⁸ Cnil « Donnée sensible », <https://cnil.fr/fr/definition/donnee-sensible?utm>, consulté le 03/02/2026.

⁴⁹ Article 6 du RGPD, « Licéité du traitement » ; Article 9 du RGPD, « Traitement portant sur des catégories particulières de données à caractère personnel » ; <https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04/fra>, consulté le 03/02/2026. L'avis commun 143 du CCNE et 5 du CNPEN intitulé « Plateformes de données de santé : enjeux d'éthique » montre qu'une réflexion riche sur les nouvelles modalités du consentement est en cours. Pour résumer, l'avis recommande de « favoriser un consentement dynamique en garantissant la confiance dans la PDS par une information régulière et transparente, sans trop de lourdeur. » (p. 49)

Pourtant, cette approche permet de protéger les droits individuels tout en tirant parti des avancées scientifiques, en évitant les usages abusifs ou discriminatoires, et en garantissant la validité scientifique des recherches.

Dans le contexte de la recherche, des exemples comme le programme « France Médecine Génomique »⁵⁰ ou la cohorte « Constances »⁵¹ montrent qu'il est possible de concilier exploitation scientifique et protection des droits individuels. Les participants donnent un consentement éclairé pour le prélèvement d'échantillons et leur analyse, et sont régulièrement informés des usages de leurs données. Ils peuvent retirer leur consentement s'ils sont opposés à certains usages. Ce suivi participatif favorise l'adhésion et la transparence : dans Constances, le taux d'acceptation de la réutilisation génétique atteint environ 70 %, grâce à des lettres d'information, webinaires et contacts réguliers. Les chercheurs doivent également prévoir la possibilité pour les participants de s'opposer à la réutilisation future de leurs données ou à de nouveaux projets.

L'exemple de Constances montre qu'un cadre participatif, transparent et juridiquement solide est possible et constitue un modèle pour la recherche sur les données sensibles.

4.2. LIMITES ET DÉFAUTS DU SYSTÈME ACTUEL

Malgré le cadre expliqué plus haut, le système actuel reste limité dans sa capacité à documenter les inégalités selon l'origine et à appuyer des politiques publiques efficaces pour un certain nombre de raisons :

- a) *Les variables nécessaires sont parfois absentes des grandes bases de données publiques*

De grandes bases de données de santé comme le SNDS (système national des données de santé) ne contiennent pas la variable « pays de naissance », pourtant essentielle pour identifier les immigrés et analyser leurs trajectoires de santé ou de mortalité⁵². Cette absence ne facilite pas l'étude des inégalités de santé selon l'origine,

⁵⁰ <https://pfmtg2025.fr/presentation-2/>

⁵¹ <https://www.constances.fr/>

⁵² Le principe de minimisation explique pourquoi certaines bases comme le SNDS ne retiennent pas le pays de naissance : ce n'est pas interdit, mais souvent non pertinent. Par ailleurs, dans la pratique, de nombreuses demandes de laboratoires invoquent des motifs externes, notamment une exigence présumée des autorités étrangères (comme la FDA) visant à garantir la représentativité ethnique des participants. La CNIL examine alors si cette exigence constitue réellement une nécessité scientifique pour l'étude envisagée en France. Lorsque ce n'est pas le cas, la collecte de données relatives à l'origine n'est pas jugée proportionnée et se heurte au principe de minimisation.

comme les différences de recours aux soins ou d'évolution des pathologies entre personnes nées en France et à l'étranger. Or, le pays de naissance est une donnée objective et légale. Pour sa part, le recensement collecte le pays de naissance mais ne collectait pas jusqu'en 2025 le pays de naissance des parents, pourtant utile pour étudier les situations des immigrés de deuxième génération. Depuis 2025, et au terme d'un débat sur la pertinence de l'introduire dans le recensement plutôt que dans des enquêtes partielles, cette variable est incluse mais elle est facultative, ce qui pourrait limiter la qualité et la représentativité des données⁵³.

b) Difficulté d'appariement entre les bases de données publiques

Le second ensemble de limites tient aux difficultés d'appariement entre les différentes bases de données publiques utilisées pour analyser la population et celles relatives à la santé. Le groupe de travail a auditionné Myriam Khlaf et Michel Guillot, démographes, pour saisir les tensions et enjeux rencontrés par les chercheurs lorsqu'ils tentent de suivre les trajectoires de santé et de mortalité des populations immigrées et de leurs descendants, dans un cadre méthodologique et juridique contraint.

Aujourd'hui, cette analyse des inégalités de santé selon l'origine repose sur un croisement complexe de sources hétéroclites : l'Échantillon démographique permanent (EDP)⁵⁴ est un outil permettant de lier, sous réserve d'une autorisation de la CNIL, les bulletins de recensement, de naissance et de décès. D'autres bases, comme l'Échantillon Longitudinal de Mortalité, permettent de suivre la survie sur plusieurs décennies et d'inclure des informations complémentaires, telles que l'origine familiale ou la langue parlée dans l'enfance, afin de distinguer les immigrés des rapatriés. Des données issues de la Caisse nationale d'assurance vieillesse (CNAV) permettent un suivi spécifique des retraités, y compris ceux retournés dans leur pays d'origine. Croiser ces différentes bases de données est actuellement très difficile. Si des progrès récents – comme l'appariement naissance-décès par l'INSEE – améliorent la situation, les limites restent importantes. Les contraintes techniques de l'état civil limitent actuellement les appariements aux seules années 2022-2024, et le

⁵³ <https://blog.insee.fr/en-2025-le-questionnaire-du-recensement-evolue/>

⁵⁴ Il s'agit d'un échantillon représentatif de la population, dans lequel les bulletins de recensement, naissance et décès sont liés.

suiivi des décès à l'étranger demeure partiel, ne concernant que certains retraités suivis par la CNAV.

La mise en relation de ces bases de données et la standardisation des variables renseignant sur l'origine – notamment le pays de naissance de la personne et celui de ses parents – apparaissent comme des conditions nécessaires pour renforcer la qualité des analyses et suivre de manière fiable les disparités de santé à travers les générations. L'ajout du pays de naissance au Système National des Données de santé (SNDS), techniquement réalisable, représenterait en ce sens une avancée significative pour la recherche sur les inégalités de santé.

c) Un accès inégal aux données sensibles selon les institutions ?

Une troisième limite, entendue dans nos auditions, est que l'autorisation de collecte et traitement de données en lien avec l'origine peut sembler variable selon les institutions. Les structures qui utilisent fréquemment ce type de données tendent à obtenir plus rapidement les autorisations. Elles sont habituées à justifier scientifiquement le besoin des données. Les dossiers moins mûrs peuvent nécessiter des échanges plus longs ; cette variabilité dépend du niveau de maturité des projets⁵⁵.

Par ailleurs, il semble que les données sur l'origine ethno-raciale soient traitées par les instances chargées de la réglementation avec plus de prudence que d'autres données sensibles, comme celles sur l'orientation sexuelle. Enfin la décentralisation des décisions vers les délégués à la protection des données (DPO), qui a favorisé une approche plus souple, a aussi introduit une diversité de pratiques entre les institutions, avec des applications de la règle et de ses exceptions plus ou moins rigoristes.

d) Une auto-censure sur les données relatives à l'origine

Par ailleurs, on observe une auto-censure des chercheurs en santé dans la collecte de données sur l'origine, qui semble en grande partie liée à une interprétation très restrictive du RGPD, du fait d'une mauvaise connaissance du règlement européen et du régime dérogatoire pour la recherche prévu par la CNIL, et parfois même de la méconnaissance du cadre légal, qui permet de collecter le pays de naissance. Cette prudence peut conduire au recours à des catégories imprécises et à des compromis

⁵⁵ La CNIL a développé des référentiels spécifiques à la recherche en santé, intégrant la collecte de données relatives à l'origine ethnique. Elle privilégie le dialogue et l'accompagnement des porteurs de projet, en vérifiant que les justifications scientifiques sont validées par des comités éthiques et scientifiques (notamment d'un comité de protection des personnes pour les recherches impliquant la personne humaine)

méthodologiques qui limitent l'analyse fine des inégalités. C'est le cas par exemple quand le niveau de détail sur le lieu de naissance est uniquement donné par région (Afrique, Asie, Europe etc...), ce qui empêche d'analyser les hétérogénéités par pays au sein de ces régions de naissance, qui peuvent être pourtant très importantes pour la compréhension de l'exposition à des maladies. Cela souligne la nécessité de mieux faire connaître le cadre juridique et la réglementation qui en France permet de concilier exigences scientifiques et protection des droits.

e) Les limites des « proxys »

Pour éviter le recours à des données subjectives et d'identification ethno-raciale, les chercheurs en France utilisent souvent des données objectives, comme le pays de naissance de la personne ou de ses parents, pour approcher (on parle de « proxys ») la perception ethno-raciale ou apparence physique. Par exemple, dans les enquêtes françaises telles que Trajectoires et Origines (TeO), ces variables ont permis de définir des groupes de population – natifs sur deux ou trois générations, personnes issues des départements et régions d'outre-mer, immigrés – et d'analyser les discriminations ou inégalités subies. Cependant, l'usage de la région de naissance pour identifier les populations qui peuvent être discriminées du fait de leur couleur de peau, par exemple, deviendra moins pertinent au fil de l'intégration des générations. Pour interroger spécifiquement des discriminations sur la perception ethno-raciale, il est donc nécessaire de recourir à des études dédiées, qui collectent cette information par dérogation par la CNIL. Un autre proxy utilisé dans les études sur les inégalités ou la discrimination peut être les noms de famille (données onomastiques). Comme le pays de naissance, utiliser ces données pour créer des catégories comporte des risques d'amalgame entre différentes situations, et requiert la prudence pour éviter des interprétations essentialisantes ou discriminatoires.

f) Conclusion : un cadre protecteur mais souple

Ainsi, le système français qui encadre l'utilisation des données sur l'origine repose sur un cadre qu'on peut qualifier de protecteur et souple : il protège l'accès à ces données sensibles que sont les données sur l'origine et ne permet pas a priori leur recueil ou leur traitement. Cependant, il l'autorise au cas par cas, par régime dérogatoire, à des fins de recherche ou de santé, après vérification par la CNIL de la pertinence de la demande et de la sécurité du traitement.

Ce système semble cependant aujourd'hui pâtir de certaines limites : en grande partie parce que ce cadre est mal connu et mal compris, l'accès aux données selon l'origine est limité et inégal, ce qui constitue un obstacle à la documentation des inégalités et des discriminations, et de ce fait ne permet pas de soutenir pleinement des politiques publiques fondées sur des données robustes.

4.3. ÉLÉMENTS DE COMPARAISON DANS LE MONDE

Les comparaisons internationales font apparaître deux groupes de pays dont les pratiques sont nettement distinctes. Le premier groupe inclut les Etats-Unis, le Royaume-Uni et, de manière moins nette, le Canada. Ces pays utilisent, en routine, des catégories ethno- raciales dans leurs études de santé publique. Le second groupe, formé de plusieurs pays européens, adopte, au contraire, une pratique universaliste proche de celle de la France.

Les grandes caractéristiques de ces deux groupes sont décrites dessous. Une description plus détaillée figure en annexe au présent avis.

Le premier groupe utilise les catégories ethno- raciales de manière normative sur l'ensemble du territoire des pays concernés, qu'il s'agisse d'un Etat fédéral (les Etats-Unis), ou d'un Etat national (le Royaume-Uni). Il est frappant de constater que les catégories retenues par les systèmes statistiques de ces pays ne sont pas définies de manière homogène. Aux Etats-Unis, ces catégories additionnent des critères d'apparence physique (Blanc, Noir), des critères géographiques (« Hawaïen ou Pacifique », « Asiatique ») et des critères culturels (« Hispanique ou Latino »). Au Royaume-Uni, les catégories additionnent ces mêmes critères mais aussi des critères juridiques traduisant la nationalité des personnes (« Asian British » ou « Black British »). En outre, ces systèmes prévoient la possibilité pour les personnes remplissant les formulaires de recueil statistique de cocher une case « mixed » ou « multiple ethnic group ».

Le caractère hétérogène - d'aucuns diront « indéterminé » - de ces catégories explique sans doute qu'elles ont connu de nombreuses variations depuis leur introduction dans les systèmes statistiques en cause. Aux Etats-Unis, la dernière « actualisation » date de 2024. Les autorités publiques justifient la conception et la mise en œuvre de ces systèmes par l'idée qu'il ne s'agit pas de catégories destinées à ouvrir des droits mais à analyser des comportements et notamment d'éventuelles discriminations auxquelles

les populations en cause se heurteraient. Il ne s'agit donc pas, selon les autorités publiques, d'un outil juridique mais, à titre principal, d'un outil d'analyse sociologique. Il convient cependant d'ajouter que les autorités américaines et britanniques utilisent ces catégories ethno-raciales, voire des sous-groupes démographiques, au sein des systèmes de santé, notamment dans le cadre d'essais cliniques de médicaments, pour tenter d'apprécier d'éventuelles divergences d'efficacité des produits selon chaque catégorie. Le caractère hétérogène de ces catégories et la forte évolution de leur périmètre au cours du temps interrogent cependant sur leur pertinence dans le champ de la science médicale.

Le second groupe poursuit une politique statistique proche de celle de la France, telle qu'elle a été décrite dans les pages qui précèdent sous la dénomination de régime d'universalisme républicain avec exceptions justifiées. Les pays recensés (Allemagne, Belgique, Pays-Bas, Suède) sont tous à la recherche d'un compromis entre mémoire historique, protection juridique et objectifs de santé publique. L'histoire de ces pays, incluant également la France, explique le bannissement absolu du terme de race dans leurs systèmes statistiques. L'usage de la catégorie « origine ethnique » est parfois autorisée, comme en Belgique ou en Suède, mais de manière strictement conditionnée à des justifications scientifiques, notamment dans le domaine des essais cliniques ou à des objectifs de lutte contre les discriminations. Des « proxys » de l'origine ethnique sont parfois mis en œuvre comme, par exemple, le pays d'origine. Il présente l'avantage d'être un critère objectif et stable dans le temps mais n'est pas toujours très opérant pour définir des objectifs de santé publique.

En définitive, ces pays partagent avec la France le rejet de toute catégorisation statistique qui serait de nature à essentialiser des populations mais sont à la recherche d'outils utiles aux politiques de lutte contre les discriminations.

5. FAUT-IL FAIRE ÉVOLUER LE SYSTÈME FRANÇAIS EN MATIÈRE DE DONNÉES DE SANTÉ RELATIVES À L'ORIGINE ?

A l'issue de cette réflexion sur le système français en matière de données de santé relatives à l'origine, nous parvenons à la conclusion que, globalement, le cadre actuel répond bien à la recherche d'un équilibre entre deux objectifs : renforcer la connaissance des spécificités de santé, inégalités ou discriminations liées à l'origine, tout en respectant les principes de l'universalisme républicain. Cette conclusion écarte l'idée d'un rapprochement avec la conception américaine de ces statistiques, étrangère à notre tradition historique et inutile à la recherche des deux objectifs cités. Il demeure une responsabilité éthique collective face aux inégalités de santé persistantes et à la capacité de l'État à y répondre, étant entendu que l'action publique ne peut se limiter à la réglementation, mais doit également assurer une mise en œuvre équitable et réfléchie. Afin de rendre ce cadre plus effectif, nous proposons quatre recommandations.

5.1. RECONNAITRE ET UTILISER PLEINEMENT LA FLEXIBILITÉ DU CADRE LÉGAL EXISTANT EN MATIÈRE D'USAGE DES DONNÉES SUR L'ORIGINE DANS LA SANTÉ

Il existe une perception répandue selon laquelle le cadre légal français serait rigide, voire prohibitif, concernant l'usage des données relatives à l'origine des individus dans le domaine de la santé. En réalité, ce cadre s'avère plus flexible qu'il n'y paraît, dès lors que l'utilisation des données est justifiée par des finalités de soin ou d'intérêt général, dans le respect des conditions de protection des personnes et de leurs données fixées par la CNIL et le RGPD.

Les enjeux principaux ne résident pas dans une interdiction absolue de collecte ou d'usage, mais dans la qualité des données recueillies, la protection des droits des patients et le respect de protocoles. Des dispositifs de recherche, comme les cohortes épidémiologiques ou certaines enquêtes ad hoc, montrent que, sous réserve de justification scientifique et de sécurisation des informations, il est possible d'intégrer des variables relatives à l'origine pour analyser des différences de santé et identifier des inégalités.

Dans le prolongement des interrogations formulées lors de l'avis relatif aux plateformes de données de santé, il convient de souligner que certaines bases de données contiennent déjà des informations relatives aux origines, collectées dans un cadre clinique ou médico-scientifique légal. Toutefois, leur existence et leur enrichissement ne sauraient être considérés comme allant de soi : leur légitimité repose sur le respect strict des finalités de soin ou de recherche, sur la proportionnalité des données collectées, ainsi que sur des garanties renforcées en matière de transparence, de gouvernance et de prévention des usages discriminatoires.

Reconnaître cette flexibilité et mieux informer chercheurs et soignants sur le cadre légal et ce qu'il permet ou ne permet pas contribuerait à dépasser les scrupules qui freinent un usage raisonné de ces données. Dans le domaine de la santé, une telle reconnaissance éclairerait les professionnels et les chercheurs sur les possibilités offertes par le cadre légal, toujours dans l'intérêt du patient.

5.2. FACILITER L'ACCÈS AUX DONNÉES OBJECTIVES DISPONIBLES

La recherche sur les inégalités sociales et migratoires en santé repose sur l'accès à des données objectives et fiables, notamment sur l'origine géographique des individus et de leurs parents, permettant d'analyser discriminations et inégalités dans l'accès aux soins.

Faciliter cet accès aux données objectives sur l'origine géographique et le parcours migratoire aidera à la production de connaissances robustes sur les inégalités de santé. Cependant l'amélioration de l'accès aux données pose la question du juste équilibre entre connaissance, protection des personnes et prévention des mésusages. Cela implique de clarifier l'interprétation du cadre juridique existant, d'harmoniser les pratiques de collecte et de systématiser l'existence de ces variables dans les bases de données nationales. Ces avancées permettront de mieux cibler les politiques publiques pour réduire les inégalités.

5.3. INTÉGRER LE PAYS DE NAISSANCE DANS LE SNDS

Le Système national des données de santé (SNDS) constitue une ressource majeure pour la recherche en santé publique en raison de la richesse des informations qu'il rassemble, ainsi qu'il a été dit. Il présente néanmoins la limite de ne pas inclure la variable « pays de naissance ». Or, cette information est essentielle pour analyser les

dynamiques de santé des populations migrantes et de leurs descendants. L'intégration du pays de naissance dans le SNDS serait techniquement faisable et compatible avec le cadre juridique français.

Son inclusion offrirait également la possibilité de mieux relier les données de santé à l'origine géographique. Elle faciliterait l'étude des causes de décès, l'identification de groupes à risque spécifiques et l'élaboration de politiques publiques visant à réduire les inégalités de santé entre générations.

5.4. RENFORCER LA FORMATION DES PROFESSIONNELS DE SANTÉ AUX ENJEUX LIÉS À L'ORIGINE : EXPOSITIONS, INÉGALITÉS, DISCRIMINATIONS

La formation des professionnels de santé doit insister sur la compréhension des enjeux liés à l'origine des patients, non pas pour raisonner en termes ethniques ou raciaux dans la pratique clinique, mais pour mieux appréhender les expositions, les déterminants sociaux et les inégalités de santé. Les biais de perception liés à l'origine – explicites ou implicites – peuvent affecter la qualité des soins en France, que ce soit par des différences dans les diagnostics, le choix des traitements, ou la stigmatisation de certaines populations, y compris parmi les migrants ou leurs descendants.

Renforcer la formation des soignants consiste à distinguer l'origine géographique, facteur d'exposition à certaines maladies infectieuses, et les déterminants sociaux, qui influencent la santé de manière transversale. Une formation renforcée devrait permettre aux professionnels de reconnaître ces biais, de dépasser les stéréotypes et de moduler la prise en charge selon les besoins réels des patients, sans stigmatisation.

ANNEXES

ANNEXE 1 : COMPARAISONS INTERNATIONALES

1. EN AMERIQUE : ÉTATS-UNIS ET CANADA

Aux Etats-Unis

La collecte de données relatives à la race et à l'ethnicité dans le domaine de la santé repose sur un cadre réglementaire fédéral harmonisé depuis plusieurs décennies. Le socle de cette standardisation est constitué par la *Statistical Policy Directive No. 15*⁵⁶ de l'Office of Management and Budget (OMB). Cette directive gouvernementale, révisée en 2024, établit des normes uniformes pour la collecte, la classification et la présentation des données sur la « race » (notion acceptée aux USA) et l'ethnicité dans toutes les agences fédérales. Son objectif est de garantir que les statistiques reflètent la diversité de la population américaine envisagée au prisme de catégories communes comme « Blanc », « Noir ou Afro-Américain », « Asiatique », « Hispanique ou Latino », « Amérindien ou Autochtone d'Alaska », « Hawaïen ou Pacifique », et « Moyen-Orient ou Afrique du Nord »⁵⁷. La directive encourage l'utilisation d'une question combinée sur la « race » et l'ethnicité, avec la possibilité de sélectionner plusieurs catégories⁵⁸, et précise que ces données servent uniquement à des fins statistiques et analytiques, non à déterminer l'éligibilité à des programmes. Les catégories raciales et ethniques utilisées dans le recensement ont beaucoup varié au cours du temps. La version de 2024 vise à moderniser les normes pour mieux représenter la diversité actuelle.

La *Food and Drug Administration* (FDA) a traduit ces standards fédéraux en recommandations pour la recherche clinique sur les produits de santé. En octobre 2016, la FDA a publié un guide définissant ses attentes concernant l'approche standardisée de collecte des données « raciales » et ethniques dans les essais cliniques pour les médicaments et dispositifs médicaux⁵⁹. Ce guide répondait aux

⁵⁶ <https://spd15revision.gov/content/spd15revision/en/2024-spd15.html>

⁵⁷ <https://spd15revision.gov/content/spd15revision/en/2024-spd15/categories-definitions.html>

⁵⁸ Les agences fédérales doivent présenter les données sur la race et l'ethnicité selon la directive SPD 15 révisée en 2024. Selon cette directive, les individus peuvent appartenir à plusieurs catégories. La directive précise que les catégories doivent être traitées équitablement, que toute combinaison ou regroupement doit être clairement étiqueté, et vise à produire des données détaillées et respectueuses de la diversité. Voir : <https://spd15revision.gov/content/spd15revision/en/2024-spd15/data-on-race-ethnicity.html>

⁵⁹ <https://www.regulations.gov/docket/FDA-2016-D-3561>

exigences de la section 4302 de l'*Affordable Care Act* (ACA) de 2010⁶⁰ qui imposait l'établissement de standards de collecte de données démographiques dans le secteur de la santé, ainsi qu'au plan d'action prévu par la section 907 du *Food and Drug Administration Safety and Innovation Act* (FDASIA) de 2012⁶¹ visant à améliorer la qualité des données démographiques. En janvier 2024, la FDA a publié une version révisée de ce guide intitulé "Collection of Race and Ethnicity Data in Clinical Trials and Clinical Studies for FDA-Regulated Medical Products »⁶². La recommandation centrale est que les données raciales et ethniques soient auto-déclarées par les participants, ces derniers pouvant sélectionner plusieurs catégories raciales pour refléter une identité multiraciale. Dans les situations où l'auto-déclaration n'est pas possible, le guide recommande de solliciter cette information auprès d'un parent ou d'une autre source informée, plutôt que de laisser l'équipe de recherche assigner une race ou une ethnicité au participant.

En standardisant la collecte des données sur la race et l'ethnicité dans les essais et études cliniques portant sur les produits médicaux réglementés, la FDA encourage les chercheurs à obtenir ces informations directement auprès des participants. L'objectif est de mieux refléter la diversité, entendue dans le sens ethno-raciale, des populations étudiées et d'améliorer la compréhension des variations de réponse aux traitements selon les groupes raciaux et ethniques, ce qui peut avoir, selon elle, un impact sur la sécurité et l'efficacité des médicaments et dispositifs médicaux⁶³. En uniformisant ces pratiques, la FDA vise à assurer une évaluation des produits médicaux plus inclusive et précise ; il n'est pas sûr qu'elle y parvienne par ce moyen.

⁶⁰ Ce guide fournit des normes uniformes pour la collecte et la déclaration de données démographiques dans les enquêtes et programmes de santé financés par le gouvernement fédéral américain. Il se concentre sur les catégories de race, ethnicité, sexe, langue principale et handicap, afin d'assurer la cohérence des données à l'échelle fédérale. Ces standards issus de la Section 4302 de l'*Affordable Care Act* servent de référence pour la formulation des questions et la codification des réponses. L'objectif est d'améliorer l'analyse des disparités de santé, la planification des services et la prise de décision politique, tout en permettant aux enquêtes de collecter des informations plus détaillées si nécessaire. Voir : <https://aspe.hhs.gov/sites/default/files/private/pdf/76331/index.pdf>

⁶¹ Il s'agit d'une initiative législative visant à améliorer l'inclusion des données démographiques dans les essais cliniques en demandant à la FDA d'analyser comment les sous-groupes démographiques (le sexe, l'âge, la race et l'ethnicité) sont inclus dans les essais cliniques pour l'approbation de médicaments et dispositifs médicaux. L'objectif est de favoriser une recherche clinique respectant la diversité et plus représentative des populations qui utiliseront ces produits. Voir : <https://www.fda.gov/regulatory-information/food-and-drug-administration-safety-and-innovation-act-fdasia/fdasia-section-907-inclusion-demographic-subgroups-clinical-trials>

⁶² <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/collection-race-and-ethnicity-data-clinical-trials-and-clinical-studies-fda-regulated-medical>

⁶³ <https://www.medicept.com/new-draft-guidance-from-fda-collection-of-race-and-ethnicity-data-in-clinical-trials-and-clinical-studies-for-fda-regulated-medical-products/>

Au Canada

Au Canada, la collecte de données relatives à la race et à l'ethnicité en santé n'est pas harmonisée au niveau fédéral. Contrairement aux États-Unis, il n'existe pas de directive nationale standardisée encadrant la manière dont ces données doivent être recueillies, classifiées et utilisées. Les enquêtes de santé et les bases de données administratives contiennent rarement des informations précises, et lorsqu'elles existent, elles sont souvent difficiles à relier entre elles⁶⁴. Les pratiques identifiées dans d'autres pays, comme l'utilisation de modèles statistiques pour identifier les groupes sous-représentés, et l'inclusion de questions sur le racisme et la discrimination, sont encore peu appliquées au Canada. Les chercheurs et les agences publiques soulignent la nécessité de normaliser la collecte de données raciales et ethniques et de recourir à l'auto-déclaration des participants pour améliorer la représentativité, notamment des populations noires, dans les études de santé⁶⁵.

Ces modèles privilégient ainsi la reconnaissance explicite des catégories ethno-raciales, tout en cherchant à encadrer les risques de discrimination.

2. EN EUROPE : ROYAUME-UNI⁶⁶, ALLEMAGNE, BELGIQUE, PAYS-BAS, SUÈDE

Au Royaume-Uni

Le Royaume-Uni se distingue en Europe par son approche explicite et légalement contraignante de la collecte de données ethniques en santé. Le *Ethnicity Recording Improvement Plan* du National Health Service vise à améliorer la complétude, la qualité et l'utilisation des données relatives à l'origine ethnique des patients dans le système de santé anglais⁶⁷ en s'appuyant sur des catégories standardisées alignées sur celles du recensement national. Celles-ci regroupent les individus en grands

⁶⁴ Margaret Jamieson, Alexandra Blair, Beth Jackson, Arjumand Siddiqi, Evidence synthesis – Race-based sampling, measurement and monitoring in health data: promising practices to address racial health inequities and their determinants in Black Canadians

⁶⁵ Ibid. Les chercheurs soulignent que le Canada manque de données représentatives et détaillées sur les populations en cause, ce qui limite la capacité d'évaluer les déterminants sociaux et les disparités en matière de santé. L'article identifie les limites des données administratives canadiennes, qui ne contiennent généralement pas d'information sur la race sans recourir à des liaisons de données complexes, et recommande de réviser les stratégies d'échantillonnage, de standardiser les mesures de race et d'ethnicité, et d'adopter des pratiques novatrices pour améliorer la surveillance des déterminants sociaux et des inégalités de santé. L'objectif est de fournir des outils et lignes directrices permettant au Canada de mieux mesurer et réduire les inégalités de santé, en particulier chez les populations noires.

⁶⁶ Le Royaume-Uni qui n'est plus membre de l'Union européenne n'est pas soumis au règlement général sur la protection des données (RGPD).

⁶⁷ <https://www.england.nhs.uk/long-read/ethnicity-recording-improvement-plan/>

ensembles tels que « White », « Mixed », « Multiple ethnic groups », « Asian », « Asian British », « Black », « African », « Caribbean » ou « Black British », ainsi que « Other ethnic group », avec des sous-catégories permettant de refléter la diversité des trajectoires et des appartenances. Cette harmonisation vise à produire des données comparables à l'échelle nationale, tout en améliorant l'identification des inégalités de santé et l'adaptation des politiques publiques. Elle part du constat que des données incomplètes empêchent d'identifier les inégalités de santé et d'y répondre efficacement. Le plan fixe une approche nationale pour standardiser les catégories d'ethnicité, renforcer les systèmes informatiques afin de réduire les données manquantes ou erronées, et former le personnel à demander ces informations de façon claire et respectueuse⁶⁸. Il insiste sur la transparence envers les patients, en expliquant pourquoi ces données sont collectées et comment elles servent à améliorer l'équité, la qualité des soins et la planification des services. Enfin, le plan vise à garantir que ces données soient effectivement utilisées par le NHS pour concevoir, cibler et évaluer des politiques et des services de santé capables de réduire les inégalités entre groupes ethniques.

En Allemagne

En Allemagne, la prise en compte des données relatives à l'origine en santé se caractérise par une prudence extrême à l'égard de toute catégorisation raciale, liée au poids de l'histoire nationale et à l'horreur du national-socialisme. Le terme allemand « Rasse » est largement proscrit dans l'espace scientifique et politique, en raison de son association avec les lois raciales de Nuremberg, l'eugénisme et les politiques de persécution fondées sur la prétendue race.

Cette mise à distance se traduit par l'absence du terme « race » dans les cadres statistiques et de recherche en santé, sans pour autant exclure toute prise en compte des différences entre populations. Les travaux mobilisent alors d'autres variables, reflétées par les notions « ethnicité », « migrant », « immigrant », « réfugié » ou, plus

⁶⁸ Le NHS définit l'ethnicité comme une caractéristique auto-identifiée, reconnue par l'Equality Act 2010, et souligne que les minorités ethniques peuvent rencontrer davantage de barrières (linguistiques, informationnelles, délais, discrimination) ayant un impact négatif sur leur santé mentale et leur relation aux services. Le NHS insiste sur l'importance d'une collecte standardisée et de qualité, basée sur les catégories du recensement, privilégiant l'auto-déclaration plutôt que la supposition par les professionnels, prévoyant des options « non déclaré » ou « inconnu », et rappelant que ces données servent avant tout à la planification des services, à l'analyse statistique et à l'amélioration de l'équité, tout en nécessitant une approche éthique, transparente et respectueuse pour instaurer la confiance des patients. Voir : <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set/submit-data/data-quality-of-protected-characteristics-and-other-vulnerable-groups/ethnicity>

récemment, « origine migratoire ». Introduite dans le recensement allemand pour mieux identifier les populations issues de l'immigration, y compris après acquisition de la nationalité, cette catégorie est aussi utilisée en épidémiologie et en santé publique. Cette stratégie d'évitement terminologique s'accompagne néanmoins d'une ambiguïté conceptuelle : des notions renvoyant à des réalités sociales, culturelles ou administratives distinctes sont parfois utilisées de manière interchangeable, ce qui peut favoriser des confusions, des généralisations excessives et des risques de stigmatisation⁶⁹. Par ailleurs, la « race » peut être inscrite dans certains dispositifs techniques, bases de données internationales ou algorithmes médicaux importés, sans être toujours explicitement nommée⁷⁰. Le cas allemand illustre une tension spécifique entre la volonté de rompre avec tout passé racialisant et la volonté, pour la recherche et la santé publique, de disposer de catégories suffisamment précises pour documenter les inégalités de santé et y répondre de manière adéquate

En Belgique

En Belgique, l'approche relative aux données de santé relatives à l'origine se caractérise par une position pragmatique et explicitement encadrée, articulant exigences scientifiques, garanties éthiques et rejet de toute biologisation des différences humaines. Dans son avis par lettre n° 17 du 6 mai 2025, le Comité consultatif de Bioéthique reconnaît que la collecte de données relatives à l'origine ethnique peut, dans certains contextes, être justifiée par la nécessité de garantir la solidité scientifique des essais cliniques et d'assurer que ceux-ci bénéficient à l'ensemble des populations, notamment en tenant compte de la variabilité inter et intra-populationnelle dans la réponse aux médicaments⁷¹. Cette possibilité est toutefois strictement conditionnée : la collecte doit être justifiée dans le protocole, proportionnée à l'objectif poursuivi, conforme au principe de minimisation des données et pleinement transparente à l'égard des participants, tant au stade du consentement éclairé que dans les publications scientifiques. Le Comité opère par ailleurs une distinction nette

⁶⁹ Bartram, I., Schnieder, L., Ellebrecht, N. et al. Categorizing people in the German life sciences: a systematic literature review of classifications of human diversity. *Discov Soc Sci Health* 3, 4 (2023). <https://doi.org/10.1007/s44155-023-00033-5> :

⁷⁰ Certains algorithmes médicaux en provenance des Etats-Unis médicaux intègrent la race dans leurs calculs, ce qui peut influencer l'évaluation des risques et les décisions de prise en charge. Par exemple, des outils utilisés pour diagnostiquer des calculs rénaux tiennent compte de la classification raciale des patients : <https://algorithmwatch.ch/fr/racial-health-bias/>

⁷¹ Avis par lettre n°17- Opportunité de mentionner l'origine ethnique dans les protocoles d'essais cliniques, Comité consultatif de Bioéthique de Belgique.

entre l'origine ethnique et la « prétendue race ». Si les données relatives à l'origine ethnique peuvent être envisagées sous réserve d'une auto-déclaration libre par les participants (excluant toute assignation par autrui) le Comité rejette la collecte de données relatives à la « race », y compris sous forme auto-déclarée. Ce refus s'inscrit dans le cadre juridique belge et européen, qui proscrit toute légitimation scientifique ou juridique du concept de race et rappelle son absence de fondement biologique. Lorsque des protocoles émanent de partenaires anglo-saxons utilisant encore ce terme, le Comité laisse aux comités d'éthique locaux une marge d'appréciation quant à l'acceptabilité de la terminologie, sans pour autant en reconnaître la validité.

En Suède

En Suède, le terme « race » n'est pas utilisé car considéré comme lié au racisme. L'expression « origine ethnique » (*etniskt ursprung*) est davantage acceptée et utilisée dans le domaine de la lutte contre les discriminations. Le terme employé dans les documents publics est « origine étrangère », par opposition à « origine suédoise »⁷². Ces deux termes sont officiellement définis par Statistics Sweden⁷³ comme suit : Les personnes d'origine suédoise sont les personnes nées en Suède dont l'un des deux parents au moins est né en Suède. Les personnes d'origine étrangère sont les personnes nées à l'étranger, ou les personnes nées en Suède dont les deux parents sont nés à l'étranger. Ce qui est mesuré, par conséquent, dans certaines statistiques relève plutôt de l'origine étrangère (foreign background) définie par la naissance de la personne ou de ses parents hors de Suède, et non d'appartenances ethniques. Pour certaines études ciblées comme la santé des populations autochtones (Sámi⁷⁴), la recherche requiert des méthodes spécifiques car il n'existe pas de variable ethnique dans les registres⁷⁵.

⁷² European Commission, Data collection in the field of ethnicity, analysis and comparative review of equality data collection practices in the European Union.

⁷³ Statistics on persons with foreign background: guidelines and recommendations. Reports on Statistical Coordination for the Official Statistics of Sweden 2002:3. Voir: <http://www.scb.se/Statistik/OV/AA9999/2003M00/X11OP0203.pdf>

⁷⁴ Les Samis constituent une minorité ethnique et le seul peuple autochtone de l'Union européenne. Les études de santé basées sur la population chez les Samis de Suède sont rares et obsolètes. L'objectif de cette étude était d'analyser les différences ethniques, entre Samis et non-Samis, en matière de santé chez les hommes et les femmes vivant en Suède.

⁷⁵ Comme la législation suédoise n'autorise pas l'enregistrement de l'ethnicité, trois registres ont été utilisés pour identifier les Sámi et constituer un échantillon de population en Suède : le registre électoral Sámi, le registre des marques de rennes et les statistiques du travail basées sur des sources administratives pour identifier les entreprises d'élevage de rennes. Voir : Miguel San Sebastián, Jon Petter Stoor, The self-reported health of the Sámi in Sweden: the SámiHET study, *European Journal of Public Health*, Volume 33, Issue 3, June 2023, Pages 366–371, <https://doi.org/10.1093/eurpub/ckad040>

Aux Pays-Bas

Les données de santé aux Pays-Bas sont essentiellement fournies sur la base du pays d'origine⁷⁶. De nombreuses bases de données de santé sont mises à disposition des chercheurs en santé par le Bureau Central des Statistiques des Pays-Bas⁷⁷. Ces bases de données peuvent être liées par l'enregistrement de la population en utilisant le numéro de sécurité sociale, ou une combinaison de la date de naissance, du code postal et du sexe. Ce système permet aux chercheurs de collecter très facilement des informations supplémentaires, telles que le pays d'origine, l'état civil, l'emploi, les prestations sociales et le revenu⁷⁸.

Le système néerlandais de « pays d'origine » est considéré comme objectif et uniforme aux fins de la conduite de recherches en santé : l'idée derrière la classification par pays de naissance est que, étant donné son caractère stable, elle permet une identification uniforme des groupes, ce qui favorise la comparabilité des résultats d'étude⁷⁹. Le système n'est pas exempt de critiques, certains chercheurs estimant que l'utilisation routinière et largement acceptée de ces classifications pourrait éroder la conscience de leur statut sociopolitique et naturaliser ces catégories de recensement⁸⁰.

Les approches européennes illustrent ainsi une diversité de compromis entre mémoire historique, protection juridique et objectifs de santé publique.

⁷⁶ Helberg-Proctor, A., Krumeich, A., Meershoek, A. *et al.* The multiplicity and situationality of enacting 'ethnicity' in Dutch health research articles. *BioSocieties* 13, 408–433 (2018). <https://doi.org/10.1057/s41292-017-0077-9>

⁷⁷ Ibid. Many health databases are made available to health researchers through the Central Bureau of Statistics of the Netherlands linked system StatLine, including: the Netherlands Hospital Discharge Register, the Netherlands Perinatal Registry, death certificates, the Netherlands Information Network of General Practitioners, the Health Occupations Register, data from insurers, data on home care and nursing home patients, and Dutch hospital and mental health data

⁷⁸ Desrosières, Alain. 1999. "La commission et l'équation: une comparaison des Plans français et néerlandais entre 1945 et 1980." *Genèses. Sciences sociales et histoire* 34 (1): 28–52. <https://doi.org/10.3406/genes.1999.1550>.

⁷⁹ Stronks, K., Kulu-Glasgow, I. and Agyemang, C. (2009) The utility of 'country of birth' for the classification of ethnic groups in health research: The Dutch experience. *Ethnicity and Health* 14(3): 255–269.

⁸⁰ Smart, A., Tutton, R., Martin, P., Ellison, G.T. and Ashcroft, R. (2008) The standardization of race and ethnicity in biomedical science editorials and UK biobanks. *Social Studies of Science* 38(3): 407–423.

ANNEXE 2 : COMPOSITION DU GROUPE DE TRAVAIL ET PERSONNES AUDITIONNÉES

Composition du groupe de travail

Annabel Desgrées du Loû (Rapporteuse)
Emmanuel Didier (Rapporteur)
Marie-Pierre Gariel
Isabelle de Gaulmyn
Fabrice Gzil
François Stasse (Rapporteur)
Ruddy Valentino

Personnes auditionnées

- Amel Filali et Nicolas Vignier, Médecins spécialistes en maladies infectieuses
- Solène Brun (Docteure en sociologie et chargée de recherche au CNRS) et Patrick Simon (Socio-démographe, Directeur de recherche à l'Ined)
- Myriam Khlal (Directrice de recherche à l'Ined) et Michel Guillot (Démographe directeur de recherche à l'Ined)
- Erik Sauleau, Professeur de biostatistique en santé publique
- Elsa Supiot, Professeur de droit privé, en détachement au CNRS
- Hélène Guimiot (Cheffe du service de la santé à la CNIL), Marie Zins (Professeur des Universités en santé publique à l'Université Paris-Cité et Membre de la CNIL) et Vincent Villette (Maître des requêtes au Conseil d'Etat et Secrétaire général de la CNIL)
- Maria Melchior, Epidémiologiste, Directrice de recherche à l'Inserm
- Lluís Quintana-Murci, Généticien des populations, Institut Pasteur et Collège de France
- Elie Azria (gynécologue obstétricien et chef de service de la maternité du groupe hospitalier Paris Saint Joseph) et Priscille Sauvegrain (Sage-femme et sociologue, Université de la Sorbonne)
- Ute Kalender (Spécialiste des questions de genre et Membre du Comité d'éthique allemand) et Aldo Faisal (Professeur d'Intelligence artificielle et de neuroscience aux départements d'informatique et de bio-ingénierie de l'Imperial College de Londres)

AVIS 151⁺



ccne.

COMITÉ CONSULTATIF NATIONAL
D'ÉTHIQUE POUR LES SCIENCES
DE LA VIE ET DE LA SANTÉ