Comité de sécurité de l'information
Chambre sécurité sociale et santé

CSI/CSSS/26/308

DÉLIBÉRATION N° 26/154 DU 6 OCTOBRE 2026 PORTANT SUR LE COUPLAGE DE DONNÉES À CARACTÈRE PERSONNEL PSEUDONYMISÉES RELATIVES À LA SANTÉ EN PROVENANCE D'INTEGO - LE RÉSEAU FLAMAND D'ENREGISTREMENT POUR MÉDECINS GÉNÉRALISTES - ET DE LA DIRECTION GÉNÉRALE STATISTIQUE STATISTICS BELGIUM (STATBEL) ET COMMUNICATION À L'UNIVERSITÉ D'ANVERS, LA KU LEUVEN ET LA VRIJE UNIVERSITEIT BRUSSEL DANS LE CADRE DU PROJET « POPULATION HEALTH RESEARCH ON MULTI-MORBIDITY AND SOCIAL DETERMINANTS USING PRIMARY CARE AND POPULATION REGISTER DATA »

Le Comité de sécurité de l'information, chambre sécurité sociale et santé, dénommé ci-après « le Comité » ;

Vu le Règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016 *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE* (Règlement général relatif à la protection des données ou RGPD);

Vu la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel* ;

Vu la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, en particulier l'article 42, § 2, 3°, modifié par la loi du 5 septembre 2018;

Vu la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale,* notamment l'article 15 ;

Vu la demande de l'Universiteit Antwerpen, de la KU Leuven et de la VUB ;

Vu le rapport d'auditorat de la plate-forme eHealth du 15 septembre 2026 ;

Vu le rapport de monsieur Michel Deneyer ;

Émet, après délibération, la décision suivante, le 6 octobre 2026 :

I. OBJET DE LA DEMANDE

1. Le Centrum voor Demografie, Familie en Gezondheid (CDFG) en Eerstelijns- en Interdisciplinaire zorg (ELIZA) de l'Universiteit Antwerpen, l'Academische Centrum Huisartsgeneeskunde (ACHG) de la KU Leuven et le Brussels Institute for Social and Population Studies de la Vrije Universiteit Brussel ont introduit une demande visant à coupler des données à caractère personnel pseudonymisées relatives à la santé provenant du réseau flamand d'enregistrement des médecins généralistes Intego de la KU Leuven et de la Direction générale des statistiques de Statbel, dans le cadre du projet « Population Health Research on Multi-Morbidity and Social Determinants Using Linked Primary Care and Population Register Data ».

2. Le projet porte sur le clustering de conditions de santé, d'inégalités sociales en santé dans une perspective intersectionnelle, ainsi que les parcours de vie individuels en lien avec la santé et le recours aux soins.

3. La présente demande de couplage de données a pour but de coupler des données de patient individuelles provenant de la base de données Intego, un réseau flamand d'enregistrement des médecins généralistes, pour la période 2018 à 2026 à des caractéristiques individuelles de Statbel. Les données Intego ont trait au groupe de contact annuel de personnes âgées de plus de 18 ans, le groupe de contact étant défini comme les patients ayant rendu visite au moins une fois par an à la pratique participante de médecins généralistes. Le couplage est nécessaire pour répondre aux questions de recherche auxquelles il est impossible de répondre si on a recours à une seule des deux banques de données.

4. Ce projet de couplage des données est une collaboration entre les partenaires suivants: KU Leuven (Intego), Universiteit Antwerpen et Vrije Universiteit Brussel. Chaque partenaire apporte une expertise complémentaire, unique au projet et fait partie du consortium Data4PHM qui souhaite promouvoir une approche proactive des soins de santé basée sur des données. Ce projet s'inscrit dans le cadre du « quintuple aim » (5AIM) de la gestion de la santé des populations : (1) améliorer la santé publique, (2) la qualité des soins, (3) le bien-être du personnel soignant et (4) l'égalité au niveau des soins, (5) sans augmentation des coûts.

5. Le groupe de contact annuel de la base de données Intego pour la période 2018-2026 contient les données d'environ 400.000 personnes). Ce groupe de contact annuel est constitué de toutes les personnes qui ont consulté un cabinet de médecine générale au cours d'une année. Le nombre exact de personnes varie d'année en année. La période de sélection est basée sur le recours au logiciel CareConnect en vue de l'enregistrement des données de patient, ce qui a permis une collecte des données plus complète et plus fiable à partir de 2018. 2026 constitue à ce jour l'année la plus récente pour laquelle des données de patient sont disponibles. Les données du groupe de contact annuel complet sont demandées puisque des jeux de données longitudinaux importants sont nécessaires pour étudier (1) la multimorbidité et sa relation avec la position d'individus sur diverses dimensions sociales et ses intersections, (2) le rôle de facteurs environnementaux dans ces relations, (3) analyser les parcours de vie ('life-courses') de patients et (4) l'impact de certains événements dans la vie

('life events') sur les maladies chroniques et leur maîtrise et déroulement. Cela permet aux chercheurs de calculer des modèles plus avancés avec des indicateurs à différents niveaux (période, individu, entourage immédiat), ce qui optimise l'exploitabilité des résultats de la recherche pour la gestion de la santé publique (« public health management ») et la politique de santé. L'objectif est d'accorder une attention particulière au rôle des déterminants sociaux de la santé.

6. Toutes les personnes à partir de l'âge de 18 ans qui ont rendu visite à une pratique de médecins généralistes participant à Intego au cours d'une ou plusieurs années de la période de référence (2018-2026) sont sélectionnées, en compris les données déjà enregistrées concernant leurs antécédents médicaux pour autant qu'elles soient disponibles. Sont demandées pour ces personnes à partir de leur 18e anniversaire, des données longitudinales Demobel et Census provenant de Statbel dans le but de coupler des données socio-démographiques et des données socio-économiques à des données de patient issues de Intego.

II. COMPÉTENCE

7. En vertu de l'article 42, § 2, 3°, de la loi du 13 décembre 2006 portant dispositions diverses en matière de santé, toute communication de données à caractère personnel relatives à la santé, sauf exceptions prévues, requiert une autorisation de principe de la chambre sécurité sociale et santé du Comité de sécurité de l'information.

8. La chambre sécurité sociale et santé du Comité de sécurité de l'information estime par conséquent qu'elle est compétente pour se prononcer sur la communication de données à caractère personnel relatives à la santé.

III. EXAMEN

A. ADMISSIBILITÉ

9. Le traitement de données à caractère personnel relatives à la santé est en principe interdit[1]. Cependant, cette interdiction ne s'applique pas, notamment lorsque le traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques, conformément à l'article 89, paragraphe 1, sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée[2].

10. L'interdiction du traitement de données à caractère personnel relatives à la santé ne s'applique pas lorsque le traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques, conformément à l'article 89, paragraphe 1, sur la base du droit de l'Union ou du droit d'un

[1] Art. 9, §1er, du RGPD.
[2] Art. 9, §2, j) du RGPD.

État membre, qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée[3]. Il s'agit d'une utilisation secondaire de données initialement collectées pour une autre finalité.

11. Intego signifie « Integrated Computer Network ». Il s'agit d'un projet lancé en 1990 au sein de l'Academisch Centrum Huisartsengeneeskunde de la KU Leuven. L'objectif est de constituer une importante base de données de la morbidité dans les soins de première ligne contenant des informations sur l'incidence et la prévalence de maladies en Flandre. Des données relatives aux tests diagnostiques et aux prises en charge médicales dans les cabinets de médecine générale ont également été mises à disposition. Actuellement, Intego est toujours financé par la Communauté flamande et comprend un réseau d'environ 500 médecins généralistes au sein de plus de 125 cabinets de médecine générale partout en Flandre. Le traitement des données de Intego est nécessaire pour des études scientifiques et intervient selon les conditions définies à l'article 9.2, i) du RGPD.

12. Les missions de la Direction générale Statistique (Statbel) sont régies par la loi du 22 mars 2006 *modifiant la loi du 4 juillet 1962 relative à la statistique publique* et la loi du 8 août 1983 *organisant un Registre national des personnes physiques***.** Ce cadre légal définit les compétences de collecte, d'établissement et de diffusion de statistiques fiables qui sont pertinentes pour l'économie, la société et le territoire belge.

13. Le Comité estime par conséquent qu'il existe un fondement pour le traitement des données à caractère personnel relatives à la santé envisagé.

B. FINALITÉ

14. En vertu de l'article 5 du RGPD, les données à caractère personnel doivent être traitées de manière licite, loyale et transparente au regard de la personne concernée. Elles doivent être collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités.

15. Le couplage de (1) données de patient issues des soins de santé de la première ligne de Intego à des (2) données issues des registres de population de Statbel (Domobel et Census) et (3) les données relatives à l'environnement couplées à Intego ont pour objectif d'étudier la vulnérabilité sociale et la multimorbidité en Flandre, Belgique. Cela permet de mener des recherches innovantes sur les inégalités sociales intersectorielles, les événements de vie cruciaux et les facteurs environnementaux pertinents, afin d'approfondir les connaissances sur les principales syndémies en Belgique. Les résultats de ces travaux visent à améliorer les politiques et les interventions qui favorisent la qualité des soins et l'équité en matière de santé. Il est important de noter que cette intégration des données renforcera également la gestion de la santé publique et la qualité des soins dans la pratique clinique. La stratification des risques sociaux et les modèles de prédiction contribueront au développement d'indicateurs indirects de la vulnérabilité des patients et à l'identification des patients à haut risque, offrant ainsi aux professionnels de santé des outils supplémentaires pour suivre les

[3] Art. 9, §2, j) du RGPD.

parcours de soins. Enfin, la microsimulation sera utilisée pour établir des prévisions à long terme en matière de multimorbidité, afin d'étayer des options politiques durables en matière de santé. Le recours à des données de santé de la première ligne à grande échelle est d'une importance cruciale, étant donné que les objectifs de la recherche nécessitent un grand nombre de patients pour identifier et analyser les variations en matière de multimorbi dité, en particulier dans une perspective intersectionnelle et longitudinale. En outre, il est essentiel de compléter les informations relatives aux patients par des données issues des registres de population, étant donné que les premières ne fournissent pas de caractéristiques socio-démographiques détaillées (par exemple, l'état civil et l'origine migratoire), ni des caractéristiques socio-économiques (par exemple, le niveau d'éducation, la situation professionnelle, la situation de logement et les revenus).

1. Connaissances fondamentales: Déterminants sociaux des maladies chroniques et multimorbidité dans une approche syndémique

16. La première partie de ce projet est consacrée à une recherche innovante sur les déterminants sociaux des maladies chroniques et la multimorbidité dans une approche syndémique, complétée par une perspective intersectionnelle et centrée sur le parcours de vie.

1° Le cadre syndémique est un outil utile pour étudier les interactions entre les pathologies concomitantes et les facteurs sociaux, tant au niveau individuel que contextuel. La valeur sociétale de la recherche syndémique réside dans sa capacité à étayer des réponses de santé publique plus efficaces, plus équitables et plus holistiques. Au lieu d'aborder les maladies séparément, la recherche syndémique permet d'identifier des interactions entre les problèmes de santé ainsi que des interactions entre les problèmes de santé et les circonstances sociales, ce qui aide à s'attaquer aux causes sous-jacentes d'une mauvaise santé.

Une syndémie désigne la présence de deux ou plusieurs pathologies néfastes au sein de populations vulnérables, qui interagissent de manière synergique et sont simultanément alimentées par des forces contextuelles — ce qui renforce l'apparition et le clustering de ces pathologies au sein de certains groupes. L'approche syndémique comprend trois éléments clés:

(1) La concomitance de deux ou plusieurs maladies ou affections (par exemple, la dépression et le diabète, ou la BPCO et les maladies cardiovasculaires) au sein d'une même population ou chez les mêmes individus. Ces affections n'existent pas simplement les unes à côté des autres, elles interagissent par contre sur les plans biologique, psychologique ou comportemental.

(2) L'interaction entre ces affections est synergique, ce qui signifie que la présence d'une affection aggrave la gravité ou le bilan de l'autre. Ainsi, la concomitance du diabète et de la dépression peut augmenter le risque d'une mauvaise autogestion de la maladie et de complications supplémentaires.

(3) Ces pathologies qui interagissent sont déclenchées ou aggravées par des facteurs sociaux, économiques et environnementaux, tels que les inégalités structurelles liées au genre, à l'origine migratoire et à la précarité socio-économique.

Les syndémies correspondent donc à : (1) la survenue simultanée de pathologies + (2) l'interaction entre ces pathologies + (3) des motivations sociales/structurelles. Sur la base de ces éléments clés, les questions de recherche (QR) suivantes sont formulées:
QR1.1: Quelles sont les principales syndémies en Flandre (Belgique), à savoir (a) deux maladies ou plus formant un cluster [élément syndémique 1] et (b) qui interagissent de manière synergique, aggravant ainsi la progression, le traitement et le bilan des maladies concomitantes [élément syndémique 2] ?
QR1.2: Quelles caractéristiques individuelles renforcent le clustering de maladies (sélectionnées dans QR1.1) [élément syndémique 3] ?
QR.3: Quels sont les motivations contextuelles des syndémies (sélectionnées dans QR1.1) qui renforcent le clustering et l'interaction entre les maladies [élément syndémique 3] ?[4]

2° L'adoption d'une approche intersectionnelle dans les analyses quantitatives met en lumière les inégalités complexes liées à la présence simultanée de maladies et aux interactions néfastes entre celles-ci, à travers les strates sociales définies par l'intersection de plusieurs dimensions sociales (par exemple, l'âge, le genre, les indicateurs socio-économiques, le passé migratoire, etc.). La pertinence sociale d'une approche intersectionnelle dans la recherche syndémique réside dans sa capacité à mettre en évidence le fait que les inégalités en matière de santé ne sont pas uniquement déterminées par un seul facteur social, tel que le genre ou l'origine migratoire, mais par leurs intersections — par exemple, comment le genre, le statut migratoire et le statut socio-économique créent ensemble des risques sanitaires uniques et des obstacles à l'accès aux soins.

L'intégration d'une approche intersectionnelle donne lieu aux questions de recherche suivantes :
QR2.1: Comment les clusters de maladies identifiés varient-ils selon les différentes strates sociales, en fonction des déterminants sociaux qui se recoupent [éléments syndémiques 1 et 3] ? Par exemple : dans quelles strates sociales, le regroupement des maladies identifiées est-il particulièrement marqué, et dans quelles strates est-il faible, voire inexistant ?
QR2.2: Comment les interactions néfastes entre les maladies formant un cluster varient-elles selon les différentes strates sociales [éléments syndémiques 2 et 3] ? Par exemple :

[4] *Justification des données*. Pour répondre à ces questions de recherche, il est nécessaire de combiner des données administratives à grande échelle provenant de plusieurs sources. L'élément syndémique 1 (QR1.1.a) nécessite des données détaillées relatives aux patients, y compris des informations diagnostiques complètes concernant plusieurs pathologies, car la présence simultanée de plusieurs pathologies est difficile à établir. L'élément syndémique 2 (QR1.1.b) nécessite non seulement des informations diagnostiques, mais aussi des informations sur le parcours de soins, l'évolution de la maladie et les résultats, afin de pouvoir évaluer l'interaction synergique entre les pathologies. L'élément syndémique 3 nécessite une combinaison de données de santé et de données sociales afin d'étudier comment les facteurs individuels (QR1.2) et les facteurs contextuels (QR1.3) agissent en tant que moteurs de la syndémie. Ce dernier point s'appuie sur l'intégration des données du registre de la population de Statbel et des indicateurs environnementaux liés à Intego.

pour quelles strates sociales, la présence simultanée du diabète et de la dépression est-elle le plus fortement associée à une augmentation des complications du diabète ?[5]

3° Le cadre interdisciplinaire « Life Course Health Development » (LCHD) reconnaît les dimensions de santé complexes, non linéaires, multiniveaux et liées au développement. Le cadre LCHD offre un ensemble de principes permettant de réaliser l'étude scientifique du développement de la santé tout au long de la vie. Ce projet s'appuie sur le cadre LCHD et part des hypothèses centrales suivantes : i) la santé individuelle évolue en continu tout au long de la vie et est façonnée par les expériences passées, ii) le développement de la santé résulte de relations adaptatives et réciproques entre les individus et leur environnement social, et iii) le développement de la santé dépend du moment et de la structuration sociale des événements et des expositions. Une approche longitudinale du parcours de vie permet de prendre en compte la chronologie entre les événements du parcours de vie, d'une part, et l'utilisation des soins et les résultats de santé, d'autre part, ce qui permet de remédier aux problèmes potentiels d'endogénéité et aux biais de sélection dans les associations.

Les questions de recherche suivantes occupent donc une place centrale dans ce projet :
QR3.1: (a) Quel est le rapport entre les événements de la vie (par exemple, un divorce ou la perte d'un partenaire) et l'apparition des clusters de maladies identifiés [éléments syndémiques 1 et 3] et (b) leurs interactions néfastes [éléments syndémiques 1 et 2] ?
QR3.2: Comment les clusters de maladies et leurs interactions néfastes évoluent-ils en fonction des événements de la vie (a) au fil du temps et (b) selon les différentes strates sociales [éléments syndémiques 1 à 3] ?[6]

2. Valorisation des connaissances: Stratification des risques sociaux et modèles de prédiction des risques pour la pratique clinique

17. La deuxième partie du projet porte sur la transposition des connaissances fondamentales décrites ci-dessus en connaissances de valorisation — c'est-à-dire leur transposition en procédures et politiques éprouvées et réellement applicables. Plus précisément, cela implique la transposition des connaissances fondamentales en (1) l'identification des patients à risques dans les soins de première ligne, (2) le développement et la validation d'indicateurs de vulnérabilité sociale, et (3) la modélisation par microsimulation pour prédire le fardeau

[5] *Justification des données*. L'intégration de données de patient issues des soins de santé de la première ligne avec celles des registres de population offre aux chercheurs une occasion unique pour documenter et analyser, de manière exhaustive, les inégalités sociales de santé intersectorielles. Cette approche dépasse les limites d'autres sources de données, telles que les enquêtes, qui souvent ne parviennent pas à appréhender la complexité de plusieurs dimensions sociales intersectionnelles dans des échantillons suffisamment représentatifs (par exemple, la sous-représentation des personnes issues de l'immigration). Les données Demobel et Census de Statbel permettent de réaliser des analyses intersectionnelles granulaires, grâce à leurs indicateurs détaillés relatifs aux caractéristiques socio-démographiques et socio-économiques.

[6] J*ustification des données.* Outre la nécessité de disposer de données à grande échelle contenant des informations complètes au niveau du patient (telles que les informations diagnostiques, les détails relatifs au traitement et aux soins, les résultats et les complexités) et des informations détaillées sur les caractéristiques sociales (tant sociodémographiques que socio-économiques), une approche fondée sur le parcours de vie nécessite en outre des informations longitudinales sur le moment où surviennent les événements du parcours de vie, la progression de la maladie et les parcours de soins au niveau individuel.

futur des maladies. Ensemble, les résultats de ces efforts de valorisation contribueront à une élaboration des politiques mieux informée et contextualisée.

1° Il existe un besoin évident d'une identification éclairée des patients à risque dans les soins de santé de première ligne. Le développement d'algorithmes performants de prédiction des risques est essentiel pour aboutir à des outils de prédiction des risques efficaces et faciles à utiliser dans la pratique clinique. Une identification précise des patients à risque élevé ou faible (par exemple, pour le développement d'une cardiopathie ischémique) constitue une étape nécessaire dans l'organisation d'interventions visant à réduire les risques. L'approche de prédiction des risques est utile tant pour la prévention que pour la gestion de la santé dans une perspective syndémique, qui reconnaît la manière dont les maladies se regroupent et interagissent au sein de contextes sociaux spécifiques. La prise en compte des déterminants sociaux de la santé est essentielle pour améliorer ces modèles de stratification des risques, puisqu'elle permet d'intégrer un ensemble plus complet de prédicteurs de risque pertinents. Ces travaux s'appuient donc sur les résultats de recherche de la première partie du projet. Il est important de noter qu'une stratification précise des risques permet une allocation plus efficace des ressources financières et humaines, ce qui aide à établir des priorités parmi les patients et à prévoir l'évolution de leurs besoins. Objectif de valorisation OV1: Le développement de modèles de prédiction de risques de santé dans les soins de santé de la première ligne.[7]

2° La collecte des données dans les soins de santé de la première ligne n'a pas été prévue, en premier lieu, pour suivre la vulnérabilité sociale de patients. Le statut d'intervention majorée est actuellement l'un des rares indicateurs indirects du statut socio-économique (SSE) dans les données de santé de la première ligne (et dans celles de l'assurance maladie belge (AIM)). Cet indicateur est ensuite utilisé dans la quasi-totalité des études d'impact des politiques de santé pour évaluer les résultats en matière d'équité des interventions de santé. La validité de cet indicateur en tant que variable proxy du SES est toutefois inconnue : cet indicateur reflète-t-il réellement le groupe de population avec les niveaux de SES les plus bas, ou d'autres variables proxy ou indicateurs sont-ils nécessaires ? Le couplage unique entre des données sociodémographiques détaillées et des données de santé longitudinales permet aux utilisateurs de données (1) d'évaluer si la variable proxy actuellement dominante (intervention majorée) identifie correctement les personnes en situation de vulnérabilité, tant en termes de sensibilité (capacité à identifier correctement les personnes ayant un faible statut socio-économique) que de spécificité (capacité à identifier correctement les personnes qui n'ont pas un statut socio-économique faible), et (2) de développer d'autres indicateurs de la vulnérabilité sociale. Par ailleurs, le couplage de ce jeu de données combiné à des données environnementales nous permettra également d'examiner comment les contextes caractérisés par un faible statut socio-économique (SES) constituent des environnements dans lesquels la relation entre le SES et la santé pourrait être encore plus forte — ce qui montre à quel point le

[7] *Justification des données*. Les caractéristiques de base des patients font défaut dans les données longitudinales de santé de la première ligne. C'est la raison pour laquelle il est crucial que les données de patient soient complétées par des données socio-démographiques et socio-économiques individuelles de Statbel. Cette combinaison de jeux de données permet aux chercheurs de développer des algorithmes de prédiction des risques dans le temps, compte tenu d'un ensemble de déterminants sociaux granulaires pour améliorer la précision des modèles.

statut de l'« intervention majorée » doit éventuellement être contextualisé. OV2: Le développement de variables proxy pour les vulnérabilités patient.[8]

3° Dans le cadre d'une approche syndémique, toutes les connaissances précédentes relatives aux déterminants sociaux de la multimorbidité, provenant de l'intégration des données de patient issues des soins de santé de la première ligne (évolution de la maladie et efficacité des traitements), les données issues des registres de population (stratification de la population et identification des groupes à risque) et les données environnementales (déterminants environnementaux) peuvent être combinées pour alimenter des modèles de microsimulation qui prédisent les tendances à long terme de la multimorbidité — et donc le fardeau de la maladie futur (et les coûts associés). La force des projections dynamiques réside dans le fait qu'elles permettent de modéliser l'évolution de la multimorbidité selon différents scénarios dans le temps (par exemple, expansion de l'enseignement, tendances migratoires, variations de la participation au marché du travail, etc.) et dans l'espace (par exemple, les expositions environnementales de diverses sous-populations), ce qui permet d'obtenir des estimations robustes dans des contextes incertains. Cette série de projections sur la multimorbidité est considérée comme particulièrement utile pour l'élaboration des futures politiques de santé, étant donné que cette approche proactive permet de mieux se préparer et de mener des interventions plus ciblées, ce qui, à terme, permet de réduire la pression globale exercée sur le système de santé. OV3: L'évaluation des tendances à long terme en matière de multimorbidité à l'aide de modèles de microsimulation.[9]

4° On estime que les soins médicaux ne représentent que 10 à 20 % des facteurs influençables qui contribuent à des résultats de santé favorables au sein d'une population. Les 80 à 90 % restants sont parfois désignés, au sens large, sous le terme de « déterminants sociaux de la santé » — ce qui rend nécessaire le couplage des données de santé avec les données sociodémographiques afin de pouvoir développer des interventions complexes. L'intégration des données permettra le développement et la mise en œuvre d'interventions éprouvées, adaptées aux besoins (sociaux, sanitaires et environnementaux) des personnes les plus à risque. Plus précisément, elle permettra : (a) d'améliorer la prévention : grâce à des données couplées, les prestataires de soins pourront mettre en place des services de prévention plus efficaces, adaptés aux besoins spécifiques de la population. En analysant les tendances et les facteurs à risque, ils pourront identifier les groupes à risque élevé et mettre en place des programmes de prévention ciblés. Ces programmes peuvent inclure des dépistages, des vaccinations et

[8] *Justification des données*. Le couplage des données d'Intego et de Statbel permet d'utiliser des informations relatives aux diagnostics cliniques pour analyser la manière dont les maladies forment des clusters en fonction de caractéristiques socio-démographiques et socio-économiques. Dans une étape suivante, des indicateurs indirects de la vulnérabilité des patients pourront être identifiés dans la base de données Intego. En comparant les indicateurs sociaux détaillés issus des données Statbel avec les informations socio-économiques de base disponibles, qui sont toutefois limitées, il sera possible d'évaluer la validité de ces indicateurs indirects (par exemple, l'intervention majorée).
[9] *Justification des données*. Pour appréhender les dynamiques de population dans la modélisation par microsimulation, il faut disposer, outre des prédicteurs de santé socio-démographiques, socio-économiques et environnementaux pertinents, d'un ensemble complet d'informations sur les événements de la vie (par exemple, la mortalité et la migration). Des informations diagnostiques relatives à une série de pathologies sont nécessaires pour pouvoir appréhender la multimorbidité des patients.

des campagnes de sensibilisation à la santé. L'amélioration de la prévention peut réduire l'incidence des maladies, diminuer les coûts des soins et améliorer la santé générale de la population. (b) Interventions à composantes multiples : la santé dépend d'une interaction complexe entre divers facteurs (déterminants sociaux de la santé, services de santé, contexte environnemental, etc.). La combinaison de données provenant de sources diverses permet de concevoir et de mettre en œuvre des interventions à composantes multiples, qui s'attaquent simultanément à ces différents aspects de la santé et des soins aux patients, sur la base de connaissances reflétant cette complexité. Ces interventions peuvent inclure une combinaison de traitements médicaux, de changements de mode de vie et d'adaptations environnementales. En ciblant plusieurs facteurs à la fois, les interventions à composantes multiples peuvent conduire à des améliorations plus significatives des résultats de santé. Cette approche holistique peut également accroître la satisfaction des patients et réduire les coûts des soins grâce à la prévention des complications et des réadmissions. (c) Meilleure harmonisation entre les soins de santé et les services du bien-être : l'intégration des données relatives aux soins de santé avec les données démographiques et environnementales peut améliorer la coordination entre les services du bien-être et les prestataires de soins. Cette harmonisation garantit que les patients bénéficient de soins complets qui répondent à la fois à leurs besoins médicaux et sociaux. Ainsi, le couplage des données permet d'identifier les personnes dont le profil (sociodémographique, économique ou lié à l'environnement) les expose à un risque de résultats de santé défavorables (par exemple, les veufs ou veuves récents souffrant d'une affection chronique). Cette approche intégrée des données peut donner lieu à de meilleurs résultats de santé et à une utilisation plus efficace des moyens. OV4: Permettre des interventions mieux adaptées grâce à la triangulation des jeux de données intégrés.[10]

18. La présente étude concerne un traitement ultérieur de données qui étaient initialement collectées pour d'autres finalités.

a. INTEGO

Intego est l'acronyme de « Integrated Computer Network ». Il s'agit d'un projet lancé en 1990 au sein de l'Academisch Centrum Huisartsengeneeskunde (www.achg.be) de la KU Leuven. L'objectif est de constituer une importante base de données de la morbidité dans les soins de première ligne contenant des informations sur l'incidence et la prévalence de maladies en Flandre. Des données relatives aux tests diagnostiques et aux prises en charge médicales dans les cabinets de médecine générale ont également été mises à disposition. En 2025, Intego est encore financé par la Communauté flamande et comprend un réseau d'environ 500 médecins généralistes au sein de plus de 130 cabinets de médecine générale partout en Flandre. Les données sont mises à la disposition des chercheurs, des responsables politiques, des étudiants en master et de toute personne qui pense pouvoir en faire un usage utile.

[10] *Justification des données*. Les caractéristiques de base détaillées des patients font défaut dans les données de Intego. C'est la raison pour laquelle il est crucial que les données de patient soient complétées par des données socio-démographiques et socio-économiques individuelles de Statbel. Le couplage des données de Intego et de Statbel permettra le développement d'interventions éprouvées qui abordent simultanément les motivations cliniques et socio-économiques de la santé, ce qui renforce l'impact de ces interventions.

b. Données de Statbel

L'Office belge de statistique collecte, produit et diffuse des chiffres fiables et pertinents sur l'économie, la société et le territoire belges. La collecte intervient sur base de sources de données administratives et d'enquêtes. L'établissement de statistiques intervient d'une manière scientifique et qualitative. En vertu des articles 24bis et 24quater de la loi statistique, Statbel est autorisée à faire usage de données administratives dont le Registre national pour la production et la diffusion de statistiques.

19. Le Comité estime que le traitement de données à caractère personnel envisagé poursuit effectivement des finalités déterminées, explicites et légitimes.

C. PROPORTIONNALITÉ

20. Selon l'article 5 du RGPD, les données à caractère personnel doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées.

21. Le NISS (numéro d'identification de la sécurité sociale, c'est-à-dire soit le numéro de registre national, soit le numéro Banque Carrefour) de la personne est utilisé par les institutions qui sont concernées par la communication et le couplage des données. Les chercheurs auront uniquement accès à des numéros d'identification pseudonymisés.

22. Le lien entre les données pseudonymisées et l'identité des personnes n'est pas conservé.

23. Différentes mesures de pseudonymisation ont été mises en place. (1) Des numéros d'identification personnelles sont communiquées sous forme pseudonymisée par une Tierce partie de confiance (TTP1) aux chercheurs et ont uniquement pour but de coupler des données individuelles provenant de différentes sources de données. (2) Des variables contenant des dates (dans le jeu de données déjà couplé) sont toujours communiquées sous forme pseudonymisée. La méthode décrite ci-après pseudonymise les dates en les glissant sur la base d'une date de naissance fictive aléatoire par individu, tout en conservant correctement les relations temporelles réciproques. Étape 1: pour tout individu, toutes les dates pertinentes (p.ex. diagnostic) sont exprimées en nombre de jours par rapport à la date de naissance originale. Étape 2: une nouvelle date de naissance fictive est générée par individu. À cet effet, le jour et le mois sont choisis de manière arbitraire, tandis que l'année de naissance reste correcte. Étape 3: les intervalles de temps prévues à l'étape 1 sont appliquées à la date de naissance fictive prévue à l'étape 2. Ce qui donne lieu aux dates pseudonymisées finales, tandis que toutes les relations temporelles réciproques demeurent les mêmes. La date de naissance fictive (jour et mois) est identique pour les différents jeux de données sources et est fixée par une Tierce partie de confiance (TTP2). (3) Les variables relatives à l'origine qui indiquent la nationalité ou le pays d'origine initial, dans le cadre des données demandées à Statbel, sont regroupées par groupe de pays avant leur transmission. Cela permet d'éviter que l'origine ne puisse être déterminée à un niveau individuel avec une très grande précision, notamment pour les pays peu représentés.

24. Le Comité fait observer que l'Agence des données de santé désignera une instance au sein du pool des SCRA pour réaliser l'analyse « *small cell risk* » sur les données avant leur

communication aux chercheurs. Toutes les mesures de pseudonymisation nécessaires seront mise en œuvre pour limiter les risques de reidentification des personnes concernées.

25. Conformément à l'article 5, 8°, de la loi du 21 août 2008 relative à l'institution et à l'organisation de la Plate-forme eHealth, la Plate-forme eHealth intervient comme organisation intermédiaire. Les clés de conversion utilisées dans le cadre de la pseudonymisation des numéros de registre national sont conservées par la TTP1 (plateforme eHealth) dans le cadre d'éventuelles demandes de données ou de mises à jour pour des questions de recherche supplémentaires ou des projets de recherche ultérieurs.

26. Il s'agit d'une communication de données ponctuelle.

27. Le Comité estime que le traitement de ces données à caractère personnel est en principe adéquat, pertinent et non excessif.

D. LIMITATION DE LA CONSERVATION

28. Conformément à l'article 5 du RGPD, les données à caractère personnel doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées; les données à caractère personnel peuvent être conservées pour des durées plus longues dans la mesure où elles seront traitées exclusivement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques conformément à l'article 89, paragraphe 1, pour autant que soient mises en œuvre les mesures techniques et organisationnelles appropriées requises par le présent règlement afin de garantir les droits et libertés de la personne concernée (limitation de la conservation).

29. La durée de validité proposée pour la délibération est de quinze ans, c'est-à-dire jusque décembre 2041, après quoi un renouvellement pourra être sollicité auprès du Comité de sécurité de l'information. Ce qui correspond au délai de conservation des données couplées proposé. Elles seront détruites quinze ans après la demande.

30. Les informations couplées seront conservées pendant une période de 15 ans. Ceci signifie que toutes les données devront être détruites au plus tard en décembre 2041.

31. Le Comité estime que le délai de conservation est raisonnable.

E. TRANSPARANCE

32. Lorsque les données à caractère personnel ne sont pas collectées auprès de la personne concernée, le responsable du traitement est tenu de fournir les informations visées à l'article 14 du RGPD à la personne concernée.

33. La personne concernée peut prendre connaissance des informations sur une affiche dans la salle d'attente du médecin généraliste. Le contenu de l'affiche est repris ci-dessous.

F. MESURES DE SECURITE

34. En vertu de l'article 5 du RGPD, les données à caractère personnel doivent être traitées de façon à garantir une sécurité appropriée des données à caractère personnel, y compris la protection contre le traitement non autorisé ou illicite et contre la perte, la destruction ou les dégâts d'origine accidentelle, à l'aide de mesures techniques ou organisationnelles appropriées (intégrité et confidentialité). Ces mesures devront assurer un niveau de protection adéquat compte tenu, d'une part, de l'état de la technique en la matière et des frais qu'entraînent l'application de ces mesures et, d'autre part, de la nature des données à protéger et des risques potentiels.

35. Le traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique, ou à des fins statistiques est soumis, conformément au RGPD, à des garanties appropriées pour les droits et libertés de la personne concernée. Ces garanties garantissent la mise en place de mesures techniques et organisationnelles, en particulier pour assurer le respect du principe de minimisation des données. Ces mesures peuvent comprendre la pseudonymisation, dans la mesure où ces finalités peuvent être atteintes de cette manière. Chaque fois que ces finalités peuvent être atteintes par un traitement ultérieur ne permettant pas ou plus l'identification des personnes concernées, il convient de procéder de cette manière[11].

36. Pour garantir la confidentialité et la sécurité du traitement de données, tout organisme qui conserve, traite ou communique des données à caractère personnel est tenu de prendre des mesures dans les onze domaines d'action suivants liés à la sécurité de l'information: politique de sécurité; désignation d'un conseiller en sécurité de l'information; organisation et aspects humains de la sécurité (engagement de confidentialité du personnel, information et formations régulières du personnel sur le thème de la protection de la vie privée et sur les règles de sécurité); sécurité physique et de l'environnement; sécurisation des réseaux; sécurisation logique des accès et des réseaux; journalisation, traçage et analyse des accès; surveillance, revue et maintenance; système de gestion des incidents de sécurité et de la continuité (systèmes de tolérance de panne, de sauvegarde, …); documentation.

37. Le Comité fait observer que l'Universiteit Antwerpen, la KU Leuven et la VUB ont chacun désigné un délégué à la protection des données.

38. Le Comité fait observer qu'une analyse d'impact relative à la protection des données a été réalisée et communiquée au Comité.

39. Le Comité fait observer que la Plate-forme eHealth et la BCSS interviennent comme tierce partie de confiance (TTP) pour le couplage et la pseudonymisation des données.

40. Le Comité fait observer qu'une analyse de risque « small cell » sera réalisée par le pool SCRA, préalablement à la mise à la disposition des données.

---

[11] Art. 89, §1er du RGPD.

41. Le Comité constate que les collaborateurs de l'Universiteit Antwerpen, de la KU Leuven et de la VUB ont un devoir de confidentialité contractuel par rapport aux données qu'ils traitent dans le cadre de leur fonction.

42. Le Comité rappelle que ni l'Universiteit Antwerpen, la KU Leuven et la VUB, ni ses collaborateurs ne peuvent entreprendre des démarches pour réidentifier les personnes concernées. Les résultats de l'étude doivent être publiés sous forme anonyme.

43. Le Comité constate que l'Universiteit Antwerpen, la KU Leuven et la VUB ont désigné un médecin responsable du traitement des données à caractère personnel relatives à la santé, ainsi qu'un délégué à la protection des données.

44. Le Comité rappelle qu'en vertu de l'article 9 de la loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel, le responsable du traitement prend les mesures supplémentaires suivantes lors du traitement de données génétiques, biométriques ou des données concernant la santé :

1° les catégories de personnes ayant accès aux données à caractère personnel, sont désignées par le responsable du traitement ou, le cas échéant, par le sous-traitant, avec une description précise de leur fonction par rapport au traitement des données visées;

2° la liste des catégories des personnes ainsi désignées est tenue à la disposition de l'autorité de contrôle compétente par le responsable du traitement ou, le cas échéant, par le sous-traitant;

3° il veille à ce que les personnes désignées soient tenues, par une obligation légale ou statutaire, ou par une disposition contractuelle équivalente, au respect du caractère confidentiel des données visées.

45. Le Comité rappelle explicitement les dispositions du Titre 6. Sanctions de la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*, qui prévoient des sanctions administratives et sanctions pénales sévères dans le chef du responsable du traitement et des sous-traitants en cas d'infraction aux conditions prévues dans le RGPD et dans la loi précitée du 30 juillet 2018.

Par ces motifs,

la chambre sécurité sociale et santé du Comité de sécurité de l'information, en ce qui concerne la communication de données à caractère personnel relevant de sa compétence,

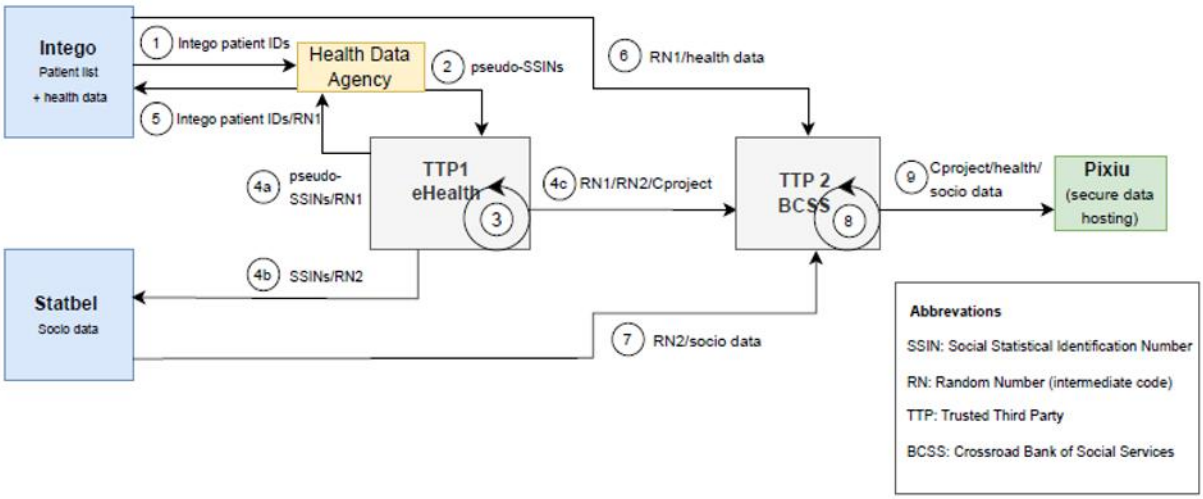
conclut que:

la communication des données à caractère personnel telle que décrite dans la présente délibération est autorisée moyennant le respect des mesures de protection de la vie privée qui ont été définies, en particulier les mesures en matière de limitation de la finalité, de minimisation des données, de limitation de la durée de conservation des données et de sécurité de l'information.

La présente délibération entre en vigueur le 21 octobre 2026.

Michel DENEYER
Président

Bijlage 1– Gegevensstroom



Stap 1: Intego identificeert de studiepopulatie binnen zijn register en bezorgt de geselecteerde lijst van Intego-patiënt-ID's aan het Health Data Agency.[1]
Stap 2: Het Health Data Agency zet de Intego-patiënt-ID's om in een lijst van pseudo-INSZ's en bezorgt deze aan het eHealth-platform (TTP1), zonder enige gezondheidsgegevens over te maken.[1]
Stap 3: eHealth (TTP1) zet de pseudo-INSZ's om in INSZ's ten behoeve van de gegevensselectie bij Statbel.[1] De lijst met pseudo-INSZ's wordt gebruikt in stap 4a (naar het Health Data Agency), terwijl de lijst met INSZ's wordt gebruikt in stap 4b (naar Statbel).
Stap 4a: eHealth (TTP1) genereert een willekeurige, eenmalige tussencode (RN1) voor de geselecteerde Intego-patiënten en bezorgt de lijst pseudo-INSZ/RN1 aan het Health Data Agency. RN1 vervangt de Intego-patiënt-ID's in de verdere verwerking bij Intego.
Stap 4b: eHealth (TTP1) genereert een afzonderlijke willekeurige, eenmalige tussencode (RN2) en bezorgt de lijst INSZ/RN2 aan Statbel. RN2 vervangt het INSZ in de verdere verwerking bij Statbel.
Stap 4c: eHealth (TTP1) bezorgt RN1/RN2/Cproject aan de KSZ (TTP2). Cproject wordt gegenereerd als de finale projectspecifieke pseudonieme identificator. Er wordt geen enkel INSZ overgemaakt.
Stap 5: Het Health Data Agency bezorgt na omzetting de lijst Intego-patiënt-ID/RN1 aan Intego.
Stap 6: Intego bezorgt de gezondheidsgegevens, gelabeld met RN1, aan de KSZ (TTP2).
Stap 7: Statbel bezorgt de sociale gegevens, gelabeld met RN2, aan de KSZ (TTP2).
Stap 8: De KSZ (TTP2) koppelt de datasets aan de hand van RN1 en RN2, verwijdert deze identificatoren en pseudonimiseert datums (bv. diagnose) in de gekoppelde dataset. Koppeling en gepseudonimiseerde identificatie gebeurt door middel van Cproject.
Stap 9: De KSZ (TTP2) bezorgt de volledig gepseudonimiseerde dataset, gelabeld met Cproject, aan de beveiligde analyseomgeving Pixiu.

1 De omzettingen tussen Intego-patiënt-ID's en pseudo-INSZ's (Intego – Health Data Agency, zie stap 2), evenals tussen pseudo-INSZ's en INSZ's (Health Data Agency – eHealth, zie stap 3), zijn vooraf vastgelegd als onderdeel van de bestaande gegevensstromen tussen deze entiteiten. Alle gegevensoverdrachten zijn versleuteld.

Bijlage 3 – informatie met betrekking tot transparantie

De machtiging voor INTEGO vereist het gebruik van een affiche die in de wachtzaal opgehangen wordt met volgende melding: .

"Uw huisarts registreert voor Intego

Uw huisarts registreert gegevens over u en uw gezondheid in het medisch dossier. Zij worden in de eerste plaats bijgehouden opdat de zorg voor u zo goed mogelijk kan gebeuren. En outre, certaines données sont aussi utilisées à des fins de recherches scientifiques. Onderzoekers proberen hiermee ziekten in kaart te brengen en de aanpak ervan te verbeteren.

Dit betekent dat uw arts gegevens over u opstuurt naar het Academisch Centrum Huisartsgeneeskunde (ACHG) van de KU Leuven in samenwerking met de dienst healthdata.be, onderdeel van het Wetenschappelijk Instituut Volksgezondheid (Sciensano). Dit gebeurt op een gecodeerde manier waardoor u niet herkenbaar bent, noch voor ACHG, noch voor Sciensano. Met deze gegevens worden verslagen gemaakt voor de overheid en de volksgezondheid, om bv. toenames of afnames van infectieziekten in kaart te brengen. Met de resultaten kunnen dan bv. campagnes opgezet worden ter voorkoming van verdere verspreiding van ziektes. Zo zullen ook de geografische coördinaten van het adres op een beveiligde manier gebruikt worden voor analyses in het kader van de studie "surveillance van milieu-gerelateerde morbiditeit". U hebt het recht de doorgestuurde gegevens in te kijken en aan te passen. U hebt ook het recht om te weigeren dat het ACHG, in samenwerking met healthdata.be, uw gegevens verzamelt. U kan hiervoor uw huisarts contacteren.

De mededeling van de gecodeerde persoonsgegevens werd gemachtigd door de afdeling gezondheid van het Sectoraal Comité van de Sociale Zekerheid en van de Gezondheid (SCSZG) met referentienr. SCSZG/13/079 bij beraadslaging 13/026 van 13 maart 2013, laatst gewijzigd op 20 maart 2018 en referentienr. 18/176 bij beraadslaging 18/079 van 3 juli 2018, laatst gewijzigd op 7 mei 2019. Het SCSZG maakt onderdeel uit van de Commissie voor de Bescherming van de Persoonlijke Levenssfeer ("Privacycommissie"). Voor elk gebruik van persoonsgegevens dat nog niet beschreven is, moet een nieuwe aanvraag gericht worden aan de Gegevensbeschermingsautoriteit.

De goedkeuringen zijn terug te vinden op de website van de Gegevensbeschermingsautoriteit: www.privacycommission.be. Indien u hieromtrent nog vragen of bemerkingen heeft, kan u dat steeds ter sprake brengen.

Uw huisarts"

Op de website van Intego (https://www.intego.be/) staat bovendien de volgende informatie voor patiënten:

"Uw huisarts registreert gegevens over u en uw gezondheid in het medisch dossier. Zij worden in de eerste plaats bijgehouden opdat de zorg voor u zo goed mogelijk kan gebeuren. Als uw huisarts deelneemt aan het INTEGO-netwerk worden sommige gegevens ook gebruikt voor

wetenschappelijk onderzoek. Onderzoekers proberen hiermee ziekten in kaart te brengen en de aanpak ervan te verbeteren. Het onderzoek wordt verricht in het algemeen belang tot een vermeerdering van kennis en inzicht die de maatschappij (direct of indirect) ten goede komt.

Dit betekent dat uw arts gegevens over u opstuurt naar het Academisch Centrum Huisartsgeneeskunde (ACHG) van de KU Leuven in samenwerking met de dienst healthdata.be, onderdeel van het Wetenschappelijk Instituut Volksgezondheid (Sciensano). Dit gebeurt op een gecodeerde manier waardoor u niet herkenbaar bent, noch voor ACHG, noch voor Sciensano. De gegevens die verzameld worden door INTEGO kan u raadplegen op https://www.healthdata.be/dcd/#/collections.Uw gegevens worden tot 30 jaar na uw overlijden bewaard in de INTEGO databank.

De INTEGO data wordt geanalyseerd door onderzoekers van ACHG – KU Leuven en onderzoekers werkzaam op het INTEGO project binnen een interuniversitaire samenwerking. Met deze gegevens worden verslagen gemaakt voor de overheid en de volksgezondheid, om bv. toenames of afnames van infectieziekten in kaart te brengen. Met de resultaten kunnen dan bv. campagnes opgezet worden ter voorkoming van verdere verspreiding van ziektes. Zo zullen ook de geografische coördinaten van uw adres op een beveiligde manier gebruikt worden voor analyses in het kader van de studie “surveillance van milieu-gerelateerde morbiditeit”.

In het algemeen belang kan de INTEGO-databank ook geaggregeerde gegevens ter beschikking stellen van commerciële partners zoals farmaceutische bedrijven (met het oog op optimalisering van diensten en producten). De commerciële partners krijgen geen rechtstreekse toegang tot de data en krijgen evenmin tot personen herleidbare data. De wetenschappelijke rapporten op hun verzoek gemaakt worden naderhand ook publiek gemaakt op de INTEGO website.

U hebt het recht uw doorgestuurde gegevens in te kijken en te laten aanpassen. U hebt ook het recht om te weigeren dat het ACHG, in samenwerking met healthdata.be, uw gegevens verzamelt in de INTEGO-databank. U kan hiervoor uw huisarts contacteren.

De mededeling van de gecodeerde persoonsgegevens werd gemachtigd door de afdeling gezondheid van het Sectoraal Comité van de Sociale Zekerheid en van de Gezondheid (SCSZG) met referentienr. SCSZG/13/079 bij beraadslaging 13/026 van 13 maart 2013, laatst gewijzigd op 20 maart 2018 en referentienr. 18/176 bij beraadslaging 18/079 van 3 juli 2018, laatst gewijzigd op 7 mei 2019. Het SCSZG maakt onderdeel uit van de Commissie voor de Bescherming van de Persoonlijke Levenssfeer (“Privacycommissie”). Voor elk gebruik van persoonsgegevens dat nog niet beschreven is, moet een nieuwe aanvraag gericht worden aan de Gegevensbeschermingsautoriteit.

De goedkeuringen zijn terug te vinden op de website van de Gegevensbeschermingsautoriteit: www.privacycommission.be.

Als u hieromtrent nog vragen of bemerkingen heeft, kan u dat steeds ter sprake brengen bij uw huisarts.”

https://www.intego.be/patienten

Bijlage 3: overzicht variabelen met omstandige motivatie

BRON: Statbel			
Dataset: Demobel			
Variabele	**Beschrijving**	**Opmerking**	**Motivatie**
ID_DEMO	Coded identification number – DEMO	Gepseudonimiseerd door TTP	Deze variabele is nodig om Statbel data te linken met patiëntendata van Intego via TTP.
CD_SEX	Sex		Geslacht is een cruciale sociodemografische gezondheidsdeterminant en daarom een individueel kenmerk dat de te bestuderen clustering van gezondheidscondities kan beïnvloeden (RQ1.2). Geslacht is ook een proxy om structurele ongelijkheden te meten op basis van gender en zal dus ook als één van de sociale dimensies van ongelijkheid bestudeerd worden vanuit een intersectionele benadering (RQs 2.1, 2.2 en 3.2). Tot slot is geslacht van belang om mee te nemen in alle valorisatie objectieven (VO1 -VO4).
DT_BTH	Date of birth	Dag en maand van de geboortedatum worden per persoon gerandomiseerd voor vrijgave. Op basis van de tijd tussen de originele geboortedatum en andere datumvariabelen worden deze laatsten gepseudonimiseerd	Geboortedatum is nodig om de leeftijd en geboortecohorte van de respondent te berekenen. Leeftijd is een belangrijke sociodemografische gezondheidsdeterminant en daarom een individueel kenmerk dat de te bestuderen clustering van gezondheidscondities kan beïnvloeden (RQ1.2). Leeftijdsverschillen zijn ook belangrijk om mee te nemen bij een intersectionele benadering (RQs 2.1, 2.2 en 3.2) en elementair als tijdschaal voor het in kaart brengen van individuele levenslooppatronen (RQs 3.1 en 3.2). Tot slot is leeftijd van belang om mee te nemen in alle valorisatie objectieven (VO1 -VO4).
DT_DEATH	Date of death	Voor deze variabele wordt een gepseudonimiseerde datum berekend. Dit betreft het aantal jaren, maanden en dagen tussen de originele geboortedatum en de betreffende datum, toegepast op de fictieve geboortedatum.	De datum van overlijden is essentieel om het ultieme gezondheidsrisico vast te stellen en maakt analyses van overleving, mortaliteit en gezondheidsverschillen mogelijk (RQ1.1, RQ1.2). Het is cruciaal voor kosten- en zorganalyses, aangezien overlijden direct de zorggebruik- en kostpatronen beïnvloedt (RQ2.3). Daarnaast ondersteunt het een intersectionele benadering door overlijden te koppelen aan sociodemografische kenmerken zoals leeftijd, geslacht en SES (RQs 2.1, 2.2, 3.2) en vormt het een belangrijke tijdsmarkering voor levensloopperspectieven (RQs 3.1, 3.2). Tot slot draagt het bij aan alle valorisatie-objectieven (VO1–VO4) door inzicht te geven in kritieke gezondheidsuitkomsten.
MS_AGE	Age at reference date		Leeftijd is een belangrijke sociodemografische gezondheidsdeterminant en daarom een individueel kenmerk dat de te bestuderen clustering van gezondheidscondities kan beïnvloeden (RQ1.2). Leeftijdsverschillen zijn ook belangrijk om mee te nemen bij een intersectionele benadering (RQs 2.1, 2.2 en 3.2) en elementair als tijdschaal voor het in kaart brengen van individuele levenslooppatronen (RQs 3.1 en 3.2). Tot slot is leeftijd van belang om mee te nemen in alle valorisatie objectieven (VO1 -VO4).
CD_CNTRY_BTH	Country of birth	Landen en nationaliteiten worden gegroepeerd opgevraagd.	Geboorteland, nationaliteit op de referentiedatum, herkomstnationaliteit en burgerlijke staat zijn belangrijke indicatoren om verblijfsduur, juridische status en culturele identiteit te construeren. Migratie status is een belangrijke sociodemografische gezondheidsdeterminant en daarom een individueel kenmerk dat de te bestuderen clustering van gezondheidscondities kan beïnvloeden
CD_NATLTY	Nationality at reference date		

CD_FST_NATLTY	First registered nationality		(RQ1.2). Migratie status is ook een proxy om structurele ongelijkheden mee te meten op basis van migratie achtergrond en zal dus ook als één van de sociale dimensies van ongelijkheid bestudeerd worden vanuit een intersectionele benadering (RQs 2.1, 2.2 en 3.2), gezien het verband met gezondheidsongelijkheden en onderdrukkingssystemen. Tal van eerdere studies hebben ook aangetoond dat migratie achtergrond relateert aan zorgtrajecten en erkennen daarnaast het belang van verblijfsduur en herkomstland. Tot slot is migratie achtergrond van belang om mee te nemen in alle valorisatie objectieven (VO1-VO4).
CD_CIV	Civil status at reference date		
ID_HH	Identification number of household		Deze informatie is nodig om de individuele data te linken met andere personen uit het huishouden om clustering in kaart te brengen.
HH_TYPE_LIPRO	Household type according to lipro typology		Huishoudtype volgens LIPRO en de individuele LIPRO positie doorheen de tijd zijn elementair in het identificeren van levenslooptransities en dus onontbeerlijk voor het construeren van individuele levenslooppatronen. Gezondheid bestuderen vanuit een longitudinaal en levensloopperspectief geeft een dieper inzicht over hoe gezondheid en zorgtrajecten temporeel samenhangen met verschillende levenslooptransities en -posities (RQs 3.1 en 3.2). Ook voor het in kaart brengen van multi-morbiditeit trends met behulp van micro-simulatie is het modelleren van levenslooptransities cruciaal (VO3), terwijl informatie over levenslooptransities potentieel ook een rol spelen in het ontwerpen van gerichte gezondheidszorginterventies (VO4).
HH_POS_LIPRO	Position within the household according to lipro typology		
CD_DSCNT	Origin	Landen en nationaliteiten worden gegroepeerd opgevraagd.	Afstammingsinformatie is cruciaal om migratie achtergrond in kaart te brengen. Met behulp van deze variabelen kan onderscheid worden gemaakt tussen generaties van migratie en dus individuele migratiegeschiedenis. Deze gedetailleerde informatie is van belang aangezien migratie status een belangrijke sociodemografische gezondheidsdeterminant is (RQ1.2). Migratie status is ook een proxy om structurele ongelijkheden mee te meten op basis van migratie achtergrond en zal dus ook als één van de sociale dimensies van ongelijkheid bestudeerd worden vanuit een intersectionele benadering (RQs 2.1, 2.2 en 3.2), gezien het verband met gezondheidsongelijkheden en onderdrukkingssystemen. Tal van eerdere studies hebben ook aangetoond dat migratie achtergrond relateert aan zorgtrajecten en erkennen daarnaast het belang van verblijfsduur en herkomstland. Tot slot is migratie achtergrond van belang om mee te nemen in alle valorisatie objectieven (VO1-VO4).
CD_DSCNT_NATLTY	Origin nationality		
DR_MAR	Start date mariage	Jaar van huwelijk	De startdatum van een huwelijk wordt gebruikt ter aanvulling van LIPRO om levenslooppatronen te construeren. Gezondheid bestuderen vanuit een longitudinaal en levensloopperspectief geeft een dieper inzicht over hoe gezondheid en zorgtrajecten temporeel samenhangen met verschillende levenslooptransities en -posities (RQs 3.1 en 3.2). Ook voor het in kaart brengen van multi-morbiditeit trends met behulp van micro-simulatie is het modelleren van levenslooptransities cruciaal (VO3), terwijl informatie over levenslooptransities potentieel ook een rol spelen in het ontwerpen van gerichte gezondheidszorginterventies (VO4).
CD_STATUS	Indicator on status of individual (year + 1): alive, death, migration or administrative deletion	Apart aangemaakte variabele voor de specifieke noden van dit project.	Deze variabele geeft de status van een inidividu doorheen de tijd weer op basis van informatie uit het Belgisch rijksregister. Dit is noodzakelijk om te achterhalen waarom een persoon eventueel verdwijnt uit de observatie. CD_STATUS geeft weer of een individu nog in leven is of nog steeds in het land verblijft in het volgend geobserveerde jaar.

CD_STATUS_PTNR	Indicator on status of partner (year + 1): alive, death, migration or administrative deletion		Deze variabele geeft de status van een eventuele partner doorheen de tijd weer op basis van informatie uit het Belgisch rijksregister. Dit is noodzakelijk voor de opvolging van partners van patiënten uit Intego die zelf geen deel uitmaken van de Intego patiëntenpopulatie. CD_STATUS_PTNR geeft weer of een partner nog in leven is of nog steeds in het land verblijft in het volgend geobserveerde jaar. In geval van een verandering in de huishoudsamenstelling en -positie op basis van LIPRO, laat deze informatie toe levenslooptransities accuraat in kaart te brengen door te differentiëren tussen de scheiding of verweduwing van een persoon. Het in kaart brengen van levenslooppatronen is cruciaal met betrekking tot RQs 3.1 en 3.2, en VO3 en VO4.
Dataset: Census 2011/2021			
CAS	Arbeidsmarktsituatie	Voor 1 januari 2011 en 2021. Jaarlijkse updates beschikbaar voor periode 2016-2020.	Arbeidsmarktsituatie is een belangrijk socio-economisch kenmerk, gegeven dat iemands relatie tot de arbeidsmarkt een gezondheidsdeterminant en voorspeller van zorgtrajecten vormt. Toegang tot gezondheidszorg is in belangrijke mate gerelateerd aan arbeidsmarktparticipatie en dus van belang voor het bestuderen clustering van gezondheidcondities (RQ1.2) en een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2). Tot slot is arbeidsmarktsituatie van belang om mee te nemen in alle valorisatie objectieven (VO1-VO4).
EDU	Onderwijsniveau volgens ISCED-2011	Voor 1 januari 2011 en 2021. Jaarlijkse updates beschikbaar voor periode 2012-2017.	Onderwijsniveau is een belangrijk socio-economisch en socio-cultureel kenmerk, gegeven dat iemands onderwijsniveau een gezondheidsdeterminant en voorspeller van zorgtrajecten vormt. Gezondheidsgedrag, -geletterdheid en toegang tot gezondheidszorg is in belangrijke mate gerelateerd aan opleiding en dus van belang voor het bestuderen clustering van gezondheidcondities (RQ1.2) en een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2). Tot slot is onderwijsniveau van belang om mee te nemen in alle valorisatie objectieven (VO1-VO4).
CD_NACE	Economische sector	NACE-code niveau 1: 21 waarden	Economische sector is als component van werkstatus een belangrijk socio-economisch kenmerk, gegeven dat iemands werksituatie betekenisvol is voor de clustering van gezondheidcondities (RQ1.2) en van belang is voor een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2).
CD_SIE_DET	Beroepsstatus		Beroepsstatuut is als component van werkstatus een belangrijk socio-economisch kenmerk, gegeven dat iemands werksituatie betekenisvol is voor de clustering van gezondheidcondities (RQ1.2) en van belang is voor een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2).
CD_TAUXPT	Werktijden (per kwartiel)	per kwartiel ingedeeld	Tewerkstellingspercentage is als component van werkstatus een belangrijk socio-economisch kenmerk, gegeven dat iemands werksituatie betekenisvol is voor de clustering van gezondheidcondities (RQ1.2) en van belang is voor een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2).
TSH	Eigendomssituatie mbt de wooneenheid van huishoudens		Wooneenheid eigendomssituatie is een belangrijk socio-economisch kenmerk, gegeven dat iemands sociale positie betekenisvol is voor de clustering van gezondheidcondities (RQ1.2) en van belang is voor een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2).
YAE	Jaar van aankomst sinds 1980		Jaar van aankomst is noodzakelijk om verblijfsduur als onderdeel van migratie status te berekenen. Migratie status is een belangrijk sociodemografische gezondheidsdeterminant en daarom een individueel kenmerk dat de te bestuderen clustering van gezondheidscondities kan beïnvloeden (RQ1.2). Migratie status is ook een proxy om structurele ongelijkheden mee te meten op basis van migratie achtergrond en zal dus ook als één van de sociale dimensies van ongelijkheid bestudeerd worden vanuit een intersectionele benadering (RQs 2.1, 2.2 en 3.2), gezien het verband met gezondheidsongelijkheden en onderdrukkingssystemen. Tal van eerdere studies hebben ook
YAT	Jaar van aankomst sinds 2000		

			aangetoond dat migratie achtergrond relateert aan zorgtrajecten en erkennen daarnaast het belang van verblijfsduur en herkomstland. Tot slot is migratie achtergrond van belang om mee te nemen in alle valorisatie objectieven (VO1-VO4).
Dataset: IPCAL			
HH_INC_NET_EQ	Equivalized net household income deciles from 2011 onwards	In decielen op basis van ADI (administratief beschikbaar inkomen van het huishouden)	Inkomen is een belangrijk socio-economisch kenmerk, gegeven dat iemands onderwijsniveau een gezondheidsdeterminant en voorspeller van zorgtrajecten vormt. Toegang tot gezondheidszorg is in belangrijke mate gerelateerd aan inkomen en dus van belang voor het bestuderen van clustering van gezondheidcondities (RQ1.2) en een intersectionele analyse van geclusterde ziekten (RQs 2.1, 2.2 en 3.2).
Bron: Intego			
Dataset: all intego datasets			
Patiënt_id	patient identifier		De gepseudonimiseerde patiënt identificator is noodzakelijk om informatie per patiënt met enerzijds Intego data uit verschillende modules en tabellen te koppelen en anderzijds Statbel data.
Practice_id	practice identifier		De gepseudonimiseerde praktijkniveau identificator is noodzakelijk om de clustering van patiënten in eerstelijnszorg in rekening te brengen.
Dataset: patient_master			
firstreg_date	first intego registration date	Voor deze variabele wordt een gepseudonimiseerde datum berekend. Dit betreft het aantal jaren, maanden en dagen tussen de originele geboortedatum en de betreffende datum, toegepast op de fictieve geboortedatum	De startdatum van patiënten registratie is van belang om de mogelijke observatieperiode per patiënt te definiëren.
lastreg_date	last intego registration date		De einddatum van patiënten registratie is van belang om de mogelijke observatieperiode per patiënt te definiëren.
ycg	list of years with at least one clinical encounter (ycg = yearly contact group)		De actieve jaren van patiënten registratie is van belang om de mogelijke observatieperiode per patiënt te definiëren.
Dataset: patient_dynamic			
increased_compensation	indicates whether a patient is in the increased reimbursement system		De indicator van het recht op een verhoogde tegemoetkoming in de gezondheidskosten van een patiënt is onontbeerlijk voor validatie objectieven VO1 en VO2. Deze laat namelijk toe om na te gaan in hoeverre het recht op een verhoogde tegemoetkoming gelinkt is aan socio-demografische en socio-economische kenmerken. Daarnaast is deze oefening instrumenteel om proxies voor sociale kwetsbaarheid bij patiënten te construeren.
change_year	year of change in increased reimbursement		De indicator geeft aan wanneer er een verandering plaatsvond in een tijdsvariërend kenmerk van de patiënt. Dit is van belang om een gewijzigde status van verhoogde tegemoetkoming in de tijd te situeren en accuraat te relateren aan tijdsvariërende achtergrondkenmerken uit de nationale registers van Statbel.
Dataset: patient_lifestyle			

lifestyle_date	date of patient lifestyle record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om de levensstijlkenmerken van een patiënt accuraat te relateren aan verschillende gezondheidscondities, hun clustering en positie in de levensloop.
physical_activity	physical activity status of a patiënt		Fysieke activiteit is een betekenisvolle voorspeller van gezondheidscondities en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
smoking_status	smoking status of a patient		Tabaksgebruik is een betekenisvolle voorspeller van gezondheidscondities en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
alcohol_use	alcohol use of a patient		Alchoholgebruik is een betekenisvolle voorspeller van gezondheidscondities en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
Dataset: patient_body_measures			
body_measures_date	date of patient body measures record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om de levensstijlkenmerken van een patiënt accuraat te relateren aan verschillende gezondheidscondities, hun clustering en positie in de levensloop.
length_cm	length of a patient in cm		Lengte is noodzakelijk om BMI te berekenen als dit ontbreekt. BMI is een relevante risicofactor voor verscheidene chronische (cardiovasculaire en metabole) aandoeningen en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
weight_kg	weight of a patient in kg		Gewicht is een relevante risicofactor voor verscheidene chronische (cardiovasculaire en metabole) aandoeningen en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
bmi	bmi of a patient		BMI is een relevante risicofactor voor verscheidene chronische (cardiovasculaire en metabole) aandoeningen en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
abd_circum	adbdominal circumference of a patient		Buikomtrek is een relevante risicofactor voor verscheidene chronische (cardiovasculaire en metabole) aandoeningen en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2).
Dataset: history_chronic_diseases			
chronic_disease_date	date of patient chronic disease record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om diagnose van chronische aandoeningen van een patiënt accuraat en chronologisch te reconstrueren om verschillende gezondheidscondities, hun clustering en positie in de levensloop in kaart te brengen.
icd10st	ICD 10 code		De ICD-10 code is van belang om de specifieke diagnose van de patiënt zoals het geregistreerd werd door de huisarts te kennen. Deze informatie over chronische aandoeningen waar een patiënt aan lijdt is essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).

icpc2	ICPC2 code		De ICPC-2 code is van belang om de concrete reden van bezoek of diagnose van de patiënt zoals het geregistreerd werd door de huisarts te kennen. Deze informatie over chronische aandoeningen waar een patiënt aan lijdt is essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
ibui	ibui code		De Belgische Unieke Identificator is van belang om de specifieke diagnose van de patiënt zoals het geregistreerd werd door de huisarts te kennen. Deze informatie over chronische aandoeningen waar een patiënt aan lijdt is essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
Dataset: history_procedures			
procedure_date	date of patient procedure record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om de medische procedures van een patiënt accuraat en chronologisch te reconstrueren om verschillende gezondheidscondities, hun clustering en positie in de levensloop in kaart te brengen.
proced_text	medical procedures in text format		Gegevens over medische procedures van een patiënt bevat mogelijk aanvullende informatie over co-morbiditeiten. Dit is belangrijk om de gezondheidsconditie van een patiënt te construeren en bijgevolg essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
Dataset: history_vaccines			
vaccination_date	date of patient vaccine record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om vaccinatie van een patiënt accuraat en chronologisch te reconstrueren om zorgtrajecten, hun clustering en positie in de levensloop in kaart te brengen.
vac_code	history of vaccination	ATC waar beschikbaar, CNK indien niet beschikbaar.	Informatie over vaccinaties is een belangrijke indicator van zorgkwaliteit. Dit is betekenisvol voor RQ1.1 dat het verloop van zorgtrajecten in functie van de interactie tussen aandoeningen in kaart brengt. Daarnaast is informatie over zorgkwaliteit van belang voor het begrip van hoe levenslooptransities relateren aan de clustering van gezondheidcondities (RQ3.1-RQ3.2).
Dataset: diagnoses			
diagnosis_date	date of patient diagnosis record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om diagnoses bij een patiënt accuraat en chronologisch te reconstrueren om verschillende gezondheidscondities, hun clustering en positie in de levensloop in kaart te brengen.
icd10	icd10 code of disease or medical condition diagnosed		De ICD-10 code is van belang om de specifieke diagnose van de patiënt zoals het geregistreerd werd door de huisarts te kennen. Deze informatie over de gezondheidsconditie van een patiënt is essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
icpc2	icpc2 code of disease or medical condition diagnosed		De ICPC-2 code is van belang om de concrete reden van bezoek of diagnose van de patiënt zoals het geregistreerd werd door de huisarts te kennen. Deze informatie over de gezondheidsconditie van een patiënt is essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
ibui	ibui code of disease or medical condition diagnosed		De Belgische Unieke Identificator is van belang om de specifieke diagnose van de patiënt zoals het geregistreerd werd door de huisarts te kennen. Deze informatie over de gezondheidsconditie van een

			patiënt is essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
diag_text	diagnosis information in text format		Aanvullende informatie over de specifieke diagnose van de patiënt is belangrijk om de gezondheidsconditie van een patiënt te construeren en bijgevolg essentieel voor het beantwoorden van alle onderzoeksvragen (RQ1-RQ3) en het bereiken van elk validatie objectief (VO1-VO4).
Dataset: care_approaches			
approach_start_date	start date of care approach	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	De startdatum van een zorgelement is belangrijk om het zorgtraject van een patiënt accuraat en chronologisch te reconstrueren om hun wisselwerking met gezondheidcondities en positie in de levensloop in kaart te brengen.
approach_end_date	end date of care approach	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	De einddatum van een zorgelement is belangrijk om het zorgtraject van een patiënt accuraat en chronologisch te reconstrueren om hun wisselwerking met gezondheidcondities en positie in de levensloop in kaart te brengen.
approach_id	care approach title	Drie mogelijke zorgtrajecten: (1) pre-diabetes, (2) chronische nierinsufficientie, (3) diabetes.	De identificatie van zorgelementen is essentieel voor RQ1.1 dat het verloop van zorgtrajecten in functie van de interactie tussen aandoeningen in kaart brengt. Daarnaast is informatie over zorgtrajecten van belang voor het begrip van hoe levensloopt ransities relateren aan de clustering van gezondheidcondities (RQ3.1-RQ3.2).
Dataset: prescriptions			
prescription_start_date	start date of prescription	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	De startdatum van een voorschrift is belangrijk om het therapeutische traject van een patiënt accuraat en chronologisch te reconstrueren om hun wisselwerking met gezondheidcondities en positie in de levensloop in kaart te brengen.
prescription_end_date	end date of prescription	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	De einddatum van een voorschrift is belangrijk om het therapeutische traject van een patiënt accuraat en chronologisch te reconstrueren om hun wisselwerking met gezondheidcondities en positie in de levensloop in kaart te brengen.
pres_atc	medication ATC code	ATC waar beschikbaar, CNK indien niet beschikbaar.	De ATC-code is van belang om geneesmiddelen voorschriften van patiënten te kennen. Dit is betekenisvol voor RQ1.1 dat het verloop van therapeutische trajecten in functie van de interactie tussen aandoeningen in kaart brengt. Daarnaast is informatie over therapeutische trajecten van belang voor het begrip van hoe levensloopt ransities relateren aan de clustering van gezondheidcondities (RQ3.1-RQ3.2).
pres_status	medication status	Twee mogelijke statussen: (1) actief en (2) passief.	In combinatie met informatie over de start- en einddatum is de status van een voorschrift belangrijk om het therapeutische traject van een patiënt accuraat en chronologisch te reconstrueren om hun wisselwerking met gezondheidcondities en positie in de levensloop in kaart te brengen.
pres_type	medication type	Mogelijke types: (1) eenmalig, (2) acuut, (3) chronisch.	Informatie over de status van voorgeschreven medicatie is van belang om onderscheid te maken tussen een langdurige therapie en beperkt gebruik. Dit is betekenisvol voor RQ1.1 dat het verloop van therapeutische trajecten in functie van de interactie tussen aandoeningen in kaart brengt. Daarnaast is informatie over therapeutische trajecten van belang voor het begrip van hoe levensloopt ransities relateren aan de clustering van gezondheidcondities (RQ3.1-RQ3.2).
dosage_text	dosage in text format		De dosering van voorgeschreven geneesmiddelen van belang om eventuele wijzigingen in de therapeutische trajecten van patiënten te kennen. Dit is betekenisvol voor RQ1.1 dat het verloop van therapeutische trajecten in functie van de interactie tussen aandoeningen in kaart brengt. Daarnaast is

			informatie over therapeutische trajecten van belang voor het begrip van hoe levenslooptransities relateren aan de clustering van gezondheidcondities (RQ3.1-RQ3.2).
Dataset: laboratory_results			
laboratory_test_date	date of laboratory test request	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om labotesten bij een patiënt accuraat en chronologisch te reconstrueren om verschillende gezondheidscondities, hun clustering en positie in de levensloop in kaart te brengen.
medidoc_code	medidoc lab test code		Deze informatie is nodig ter operationalisering van de aangevraagde labotest en de accuratie interpretatie ervan.
lab_result	lab test result (clean)		Resultaten van labotesten zijn noodzakelijk om een beter beeld te krijgen van hoe (geclusterde) ziekten verlopen (RQ1.1). Daarnaast verschaffen ze belangrijke informatie over de aanwezige risicofactoren voor verscheidene chronische (bv. cardiovasculaire en metabole) aandoeningen en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2). Tot slot is het uitvoeren van labotesten een indicator van zorgkwaliteit (cf. RQ1.1, RQ3.1-RQ3.2).
lab_unit	lab test result unit		Deze informatie is nodig voor een accurate interpretatie van labotest resultaten.
Dataset: measurements			
measure_date	date of patient measurement record / registration	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om metingen bij een patiënt accuraat en chronologisch te reconstrueren om verschillende gezondheidscondities, hun clustering en positie in de levensloop in kaart te brengen.
measure_code	measurement code	Mogelijke metingen: bloeddruk, hartslag, lichaamstemperatuur, etc.	Resultaten van metingen zijn noodzakelijk om een beter beeld te krijgen van hoe (geclusterde) ziekten verlopen (RQ1.1). Daarnaast verschaffen ze belangrijke informatie over de aanwezige risicofactoren voor verscheidene chronische (bv. cardiovasculaire en metabole) aandoeningen en daarom belangrijk om in rekening te brengen als controlevariabele in de analyse van de clustering van gezondheidscondities (RQ1.1-RQ1.3 en RQ2.1-RQ2.2) en de link tussen levenslooptransities en gezondheid (RQ3.1-RQ3.2). Tot slot is het uitvoeren van deze metingen een indicator van zorgkwaliteit (cf. RQ1.1, RQ3.1-RQ3.2).
measure_value	measurement value		
measure_unit	measurement unit		
Dataset: imaging_orders			
imaging_order_date	date of patient medical imaging request	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie is essentieel om aanvraag voor medische beeldvorming bij een patiënt accuraat en chronologisch te reconstrueren. Informatie over medische beeldvorming is een belangrijke indicator van zorgkwaliteit. Dit is betekenisvol voor RQ1.1 dat het verloop van zorgtrajecten in functie van de interactie tussen aandoeningen in kaart brengt. Daarnaast is informatie over zorgkwaliteit van belang voor het begrip van hoe levenslooptransities relateren aan de clustering van gezondheidcondities (RQ3.1-RQ3.2).
Dataset: physiotherapy_prescriptions			
physio_prescription_date	date of physiotherapy prescription	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van registratie van een aanvraag voor fysiotherapie is van belang om het therapeutisch traject van een patiënt accuraat en chronologisch te reconstureren. Dit is betekenisvol voor RQ1.1 dat het verloop van therapeutische trajecten in functie van de interactie tussen aandoeningen in kaart brengt.

			Daarnaast is informatie over therapeutische trajecten van belang voor het begrip van hoe levenslooptransities relateren aan de clustering van gezondheidcondities (RQ3.1 -RQ3.2).
Dataset: absence_certificates			
absence_start_date	start date of absence certificate	Voor deze variabele wordt een gepseudonimiseerde datum berekend.	Datum van periode van afwezigheid op het werk is van belang voor RQ1.1 om uitkomsten van clustering van gezondheidscondities in kaart te brengen.
absence_end_date	end date of absence certificate		
Dataset: intego environmental			
train _noise	train noise in decibels (annual mean)	Weighted noise level caused by rail transport over 24h (2016)	Omgevingskenmerken vormen een centraal element van een syndemische benadering van gezondheidscondities. Gelinkte omgevingskenmerken zijn van cruciaal belang om de contextuele factoren die samenhangen met de clustering en interactie van gezondheidscondities te onderzoeken (RQ1.3).
road_noise	road noise in decibels (annual mean)	Weighted noise level caused by rail transport over 24h (2018)	
aviation_noise	aviation noise in decibels (annual mean)	Weighted noise level caused by rail transport over 24h (2016)	
access_residential_greenspace	access to residential green space	Binaire variabelen voor verschillende buffer zones (2016, 2022)	
coverage_residential_greenspace	residential greenspace coverage in percent	Dekking voor verschillende buffer zones (2021)	
view_greenspace	view of green from residence	Viewshed Greenness Visibility Index (2021)	
air_quality_pm10	concentration particulate matter in µg/m³ (annual mean)	Mean of daily concentration (2009-2022)	
air_quality_pm25	concentration particulate matter in µg/m (annual mean)		
air_quality_no2	nitrogen dioxide concentration in µg/m (annual mean)		
air_quality_o3	ozone concentration in µg/m³ (annual mean)		

air_quality_bc	black carbon concentration in µg/m³ (annual mean)	Mean of daily concentration (2010-2022)	
n_heatwave_days	number of days of official heatwaves (yearly)	2000-2021	
heatwave_degree_days	summary measure of heatwave days and average temperature (yearly)		
n_heatalarm_days	number of days of official heatwave alarm (yearly)		
t_mean_daily	daily mean temperature in °C		
relative_humidity_max	daily maximum relative humidity in percentage		
relative_humidity_min	daily minimum relative humidity in percentage		
relative_humidity_mean	daily mean relative humidity in percentage		
distance_to_road	distance in to a major road (in meter) (yearly)	2019	
traffic_load	number of vehicles multiplied by length of roads within 200m around the residence (yearly)	2022	