

<p>Comité de sécurité de l'information</p> <p>Chambre sécurité sociale et santé</p>

CSI/CSSS/21/054

DÉLIBÉRATION N° 21/032 DU 7 SEPTEMBRE 2021 RELATIVE À LA COMMUNICATION DE DONNÉES À CARACTÈRE PERSONNEL PSEUDONYMISÉES RELATIVES À LA SANTÉ PAR LA FONDATION REGISTRE DU CANCER, L'AGENCE INTERMUTUALISTE, LE SPF SANTÉ PUBLIQUE ET LE DATAWAREHOUSE DE LA BANQUE CARREFOUR DE LA SÉCURITÉ SOCIALE À LA VRIJE UNIVERSITEIT BRUSSEL, GROUPE DE RECHERCHE INTERUNIVERSITY CENTRE FOR HEALTH ECONOMICS RESEARCH (I-CHER), DANS LE CADRE DE SON PROJET DE RECHERCHE

Le Comité de sécurité de l'information;

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 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, notamment l'article 37;

Vu la loi du 3 décembre 2017 *relative à la création de l'Autorité de protection des données*, en particulier l'article 114, modifié par la loi du 25 mai 2018;

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 5 septembre 2018 *instituant le comité de sécurité de l'information et modifiant diverses lois concernant la mise en œuvre du Règlement (UE) 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*, notamment l'article 97;

Vu la loi du 21 août 2008 *relative à l'institution et à l'organisation de la plate-forme eHealth et portant dispositions diverses*;

Vu la demande de la Vrije Universiteit Brussel, groupe de recherche Interuniversity Centre for Health Economics Research (I-CHER);

Vu le rapport d'auditorat de la Plate-forme eHealth du 12 août 2021;

Vu le rapport de monsieur Bart Viaene.

Émet, après délibération, la décision suivante, le 7 septembre 2021:

I. OBJET DE LA DEMANDE

1. La Vrije Universiteit Brussel, groupe de recherche Interuniversity Centre for Health Economics Research (I-CHER), a introduit une demande afin d'être autorisé par le Comité de sécurité de l'information à obtenir la communication de données à caractère personnel relatives à la santé par la Fondation registre du cancer, l'Agence intermutualiste, le SPF Santé publique et le datawarehouse de la Banque Carrefour de la sécurité sociale dans le cadre de son projet de recherche.
2. En Belgique, différentes organisations disposent chacune d'une source importante d'informations concernant les diagnostics, la consommation de soins et le coût des soins administrés, la mortalité par cause et la situation sociale des femmes atteintes de cancer du sein. Le couplage de ces données offre une valeur ajoutée d'un point de vue scientifique puisque ceci permet de réaliser des études au niveau de la population belge.
3. Le projet de recherche du I-CHER vise à retracer le parcours des femmes atteintes de cancer du sein en Belgique sur une période de vingt ans à compter du diagnostic. Le projet se concentre sur la question de savoir quelles interventions sont susceptibles de réduire la charge de morbidité d'un traitement de cancer du sein chez les femmes en Belgique.
4. Toutes les femmes belges qui ont reçu un diagnostic de cancer du sein au cours de l'année d'incidence 2010 sont incluses dans l'étude. De manière spécifique, il s'agit de femmes avec les codes ICD-10 C50 (malignant neoplasms of breast) et D05 (carcinoma in situ of breast). La date de la première confirmation microscopique de la tumeur ou, à défaut, la date de l'examen clinique qui a abouti au diagnostic de la tumeur (variable : « date d'incidence ») est considérée comme la « date indice ». Il est estimé que la série de données porterait sur 13.000 personnes.
5. L'étude comporte aussi plusieurs sous-objectifs :

1) Etude du coût sociétal du traitement du cancer en Belgique jusque 1,5 ans après le diagnostic de cancer du sein.

Sur la base des diagnostics de cancer du sein, les femmes appartenant à la cohorte sont identifiées. Le but est de dresser la carte de la consommation de soins de ces femmes de sorte à se former une idée du trajet parcouru par ces femmes jusque 1,5 ans après le diagnostic.

Les données relatives aux prestations en provenance des mutualités belges constituent à cet égard une source d'information importante pour calculer les frais médicaux directs.

Tant les coûts à charge de l'assurance maladie que la part personnelle du patient sont étudiés en détail. Les frais médicaux directs suite au diagnostic de cancer du sein doivent être corrigés en ce qui concerne les soins réguliers du patient. Cette correction est effectuée sur la base du calcul des frais au cours de la période précédant la date indice. Le patient fonctionne à cet égard comme sa propre personne de contrôle.

A condition de tenir compte de la consommation de soins accrue au cours de la phase pré-diagnostique en raison des consultations et examens pré-diagnostiques, la différence dans les frais avant et après la date indice peut être attribuée à la consommation de soins liée au traitement du cancer du sein. Les informations relatives à la situation d'emploi avant et après la date indice fournissent une indication des frais indirects (absentéisme). Le calcul des frais médicaux directs et l'évaluation des frais indirects permettent de calculer le coût sociétal global du traitement du cancer du sein en Belgique.

2) *Etude de l'incidence et de la mortalité de maladies cardiovasculaires et de tumeurs pulmonaires dans une cohorte de femmes belges atteintes de cancer du sein sur une période de 20 ans après le diagnostic de cancer du sein.*

Le trajet parcouru par les patients de la cohorte sera suivi pendant 20 ans à compter de la date indice. Les données des hôpitaux (RHM et AIM) et de la Fondation registre du cancer seront utilisées pour déterminer l'incidence de maladies cardiovasculaires et de tumeurs pulmonaires (radio-induites) dans la cohorte. Dans ce cadre, le coût des soins administrés pour ces nouveaux diagnostics peuvent à nouveau être déterminés. Par ailleurs, le projet vise aussi à examiner la mortalité par cause chez ces femmes sur une période de 20 ans après la date indice. Sur la base de ces données, le coût sociétal de tous les trajets de ces femmes atteintes de cancer du sein sur une période de 20 ans à compter du diagnostic peut être calculé.

3) *Etude du développement de nouvelles stratégies permettant de réduire la charge de morbidité liée au cancer du sein.*

Les résultats de cette étude sont susceptibles de permettre le développement de méthodologies pour calculer les coûts et effets attendus de nouvelles interventions préventives. Ces nouvelles stratégies peuvent réduire la charge de morbidité sur l'ensemble du trajet de maladie de la population.

6. Les données pseudonymisées suivantes sont demandées à la **Fondation registre du cancer** : le numéro NISS pseudonymisé, le sexe, l'âge au moment du diagnostic, la date d'incidence convertie en nombre de jours par rapport à la date indice, ICD10 (numérique), cTNM, pTNM, combined stage, sous-type histologique, ER, PR, HER2, nombre de tumeurs, multiple, latéralité, statut vital, date statut convertie en nombre de jours par rapport à la date indice, cause de décès directe, cause de décès intermédiaire (1 à 2), cause de décès initiale, cause de décès secondaire (1 à 3).
7. L'**Agence intermutualiste** fournira les données suivantes à partir de *Pharmanet* : identification pseudonymisée du bénéficiaire (SS00010), date de délivrance convertie en nombre de jours par rapport à la date indice (basée sur SS00015), code catégorie médicament (SS00020), quantité (SS00050), intervention AMI 1 (SS00060), code bénéficiaire 1 {CG1} (SS00095), code bénéficiaire 2 {CG2} (SS00100), numéro produit

(SS00135), part personnelle (SS00160), réduction intervention de l'assurance / cotisation des pharmaciens (SS00165), intervention AMI 2 (SS00195).

Les données suivantes sont transmises à partir de la banque de données *Soins de santé* : identification pseudonymisée du bénéficiaire (SS00010), date de début prestation convertie en nombre de jours par rapport à la date indice (basée sur SS00015), code de nomenclature (SS00020), nombre de cas (SS00050), nombre de jours (SS00055), montant remboursement (SS00060), code d'identification pseudonymisé du prestataire de soins (SS00065A), qualification prestataire de soins (SS00065B), identification pseudonymisée de l'établissement (SS00075), code service (SS00080), lieu de la prestation (SS00085), code bénéficiaire 1 {CG1} (SS00095), code bénéficiaire 2 {CG2} (SS00100), numéro tiers pseudonymisé (SS00105), date d'admission convertie en nombre de jours par rapport à la date indice (basée sur SS00110), date de sortie convertie en nombre de jours par rapport à la date indice (basée sur SS00115), numéro produit (SS00135), codes de nomenclature facturés (SS00150), intervention personnelle (SS00160), montant supplément ou montant pour produits non-remboursables, prestations ou services (SS00165), code implant (SS00170), régime tiers payant (SS00175).

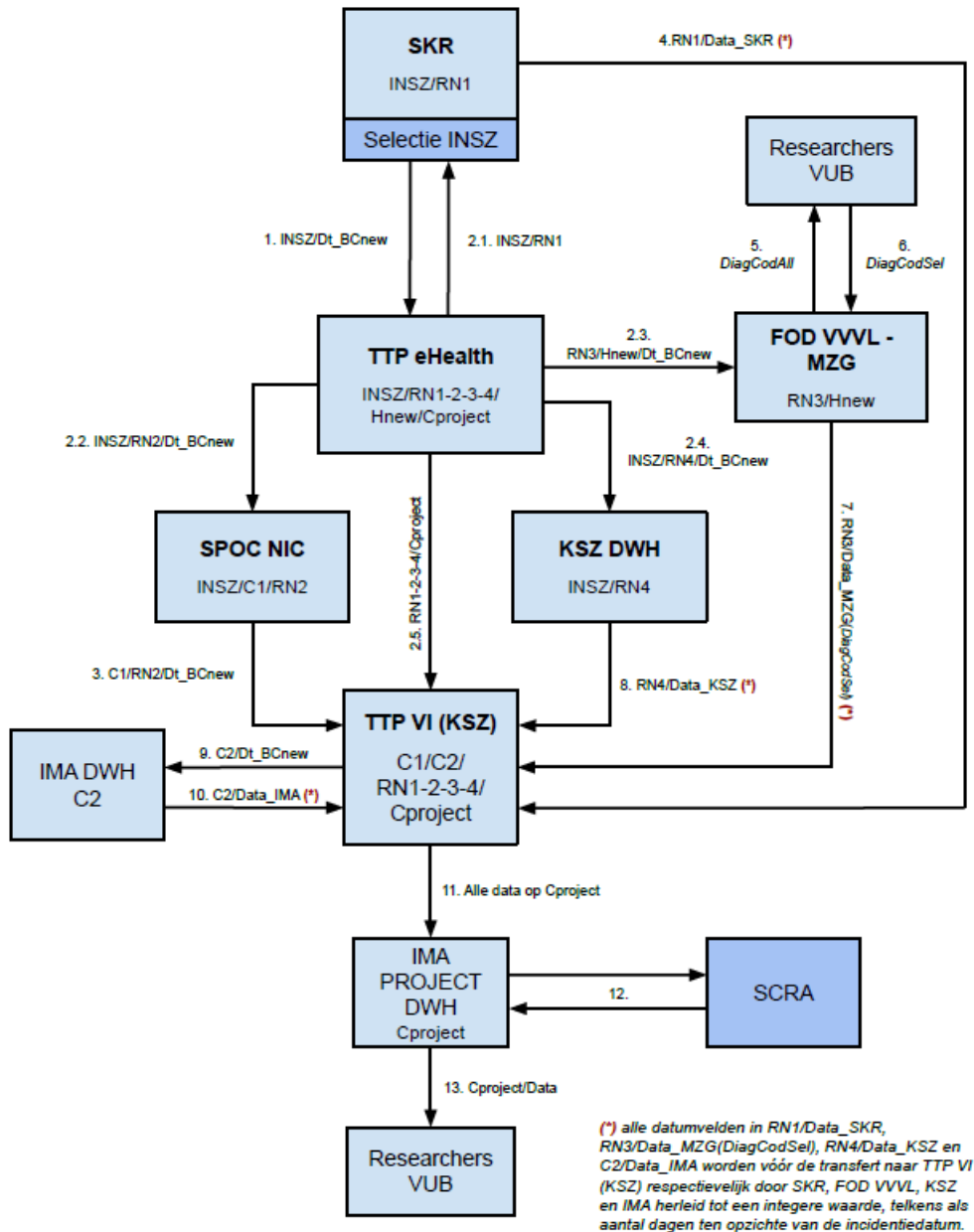
Les données suivantes sont transmises à partir de la banque de données *Population* : identification pseudonymisée du bénéficiaire (PP0010), nature IM/OMNIO (PP1010), forfait B soins de santé (PP2001), forfait C soins de santé (PP2002), kinésithérapie E ou physiothérapie (PP2003), allocations familiales majorées (PP2004), allocation d'intégration pour personnes handicapées (PP2005), allocation pour l'aide aux personnes âgées {cat III, IV, V} (PP2006), allocation aide de tiers (PP2007), allocation d'invalidité majorée en raison de l'aide de tiers (PP2008), allocation forfaitaire pour l'aide d'une tierce personne (PP2009), critère hospitalisation {120 jours} (PP2010), critère hospitalisations {6 hospitalisations} (PP2011), droit à l'allocation pour personnes handicapées (PP3011), code bénéficiaire 1 {CG1} et 2 {CG2} (PP0030 et PP0035), droit MAF ménage (PP3001), remboursement et plafond ménage (PP3004), remboursement et plafond individu ou partie du ménage (PP3005), nombre de jours de chômage (PP4001), nombre de jours d'invalidité (PP4002), CHRONICAL_YN, MAJOR_COVERAGE_YN, ATC_ANAT/THER/PHAR/CHEM/PROD_L.

8. Les données suivantes sont demandées au **SPF Santé publique** : numéro de patient anonyme (PATNUM), numéro d'agrégation pseudonymisé de l'hôpital (CODE_AGR), numéro de séjour (STAYNUM), date d'admission et de sortie converties en nombre de jours par rapport à la date indice (basées sur A2_YEAR/MONTH/DAY_HOSP_IN/OUT), catégorie de séjour hospitalier (A2_HOSPTYPE_CAT), catégorie de séjour hospitalier basée sur la facturation (A2_HOSPTYPE_FAC), nombre de journées d'hospitalisation à facturer (A2_TOTAL_NUMBER_DAY_FAC), code réadmission (A2_CODE_READMISSION), code statut d'assurance au cours du séjour (A2_CODE_STAT_INSURANCE), lieu d'admission (A2_CODE_PLACE_BEFORE_ADM), type d'admission (A2_CODE_ADM), destination (A2_CODE_DESTINATE), type de sortie (A2_CODE_DISCHARGE), numéro d'ordre indice lit (ORDER_BEDINDEX), code de l'indice lit basé sur la facturation (A4_CODE_BEDINDEX_FAC), date d'admission indice lit convertie en nombre de jours par rapport à la date indice (basée sur

A4_YEAR/MONTH/DAY_BEDINDEX_IN), nombre de journées d'hospitalisation indice lit à facturer au cours de l'année d'enregistrement actuelle (A4_NUMBER_DAY_FAC), nombre de journées d'hospitalisation à facturer au cours de l'année d'enregistrement précédente (A4_NUMBER_DAY_FAC_PREV), numéro d'ordre spécialité (ORDER_SPEC), type de diagnostic {principal ou secondaire} (TYPE_DIAGNOSE), système de codage des diagnostics (CODE_SYS_DIAGNOSE), code ICD diagnostic (CODE_DIAGNOSE), système de codage pour les interventions réalisées (M2_CODE_SYS_PROCEDURE), code ICD de la procédure (M2_CODE_PROCEDURE), date d'exécution de la procédure convertie en nombre de jours par rapport à la date indice (basée sur M2_YEAR/MONTH/DAY_PROCEDURE), nombre d'exécutions de la même procédure le même jour (M2_NUMBER_PROC_1_DAY).

9. Le **datawarehouse de la Banque Carrefour de la sécurité sociale** communique les données suivantes: le NISS pseudonymisé, la nomenclature de la position socio-économique et la position socio-économique supplémentaire convertie en une variable cluster {actif et en interruption de carrière partielle/crédit-temps partiel, actif et en prépension à mi-temps, actif et en prépension à temps plein, actif et en interruption de carrière complète/crédit-temps complet, actif et en mise en disponibilité partielle préalable à la pension, actif et en mise en disponibilité complète préalable à la pension, actif et à mi-temps dans le régime de chômage avec complément d'entreprise, actif et complètement dans le régime de chômage avec complément d'entreprise} (*source*: BCSS), niveau d'enseignement détaillé (*source*: STATBEL) niveau de formation ISCED (*source*: BCSS); catégorie d'études enseignement non universitaire (cat_shu), catégorie d'études (cat_études) (*source*: CREF), code du diplôme (CO_DIPLOME) (*source*: Saturn), régime de travail (T_Prest), pourcentage d'occupation (Eqtpsp), pourcentage d'occupation (Tauxaa) (*source*: ONSS); date de début et date de fin incapacité de travail converties en nombre de jours par rapport à la date indice (Begindatum, Einddatum), nombre de jours d'incapacité de travail (aantal_dagen_ao), composition du ménage (samenst_gezin), (*source*: CIN) ; système de codage pour la maladie à l'origine de l'invalidité (Type_ziekte), cause de l'invalidité (Gr_ziekte), code ICD pour la cause de l'invalidité (Diag_ziekte), date de début de l'incapacité de travail primaire convertie en nombre de jours par rapport à la date indice (B_ziekte), dates de début et de fin de l'invalidité converties en nombre de jours par rapport à la date indice (B_inval, E_inval), variable de correction pour la fusion des trimestres (R_exclus) (*source* : INAMI).
10. Les flux de données se dérouleront selon le schéma ci-dessous.

**Data Traject Koppeling SKR, TCT-MZG, IMA en KSZ data:
Studie Maatschappelijke Kost Borstkanker (P_CVAL01)**



Dans le cadre de ce couplage, la Fondation registre du cancer (FRC) constitue la liste des numéros NISS uniques et des dates de diagnostic (Dt_BCnew) de la population de l'étude (toutes les femmes avec un diagnostic de cancer du sein invasif et/ou in situ dans le courant de l'année 2010) et transmet la liste NISS/Dt_BCnew (*) au TTP eHealth (étape 1). Dans l'hypothèse où il y a plus d'une date d'incidence pour la sélection du cancer du sein au cours de l'année indice 2010, la date indice pour ce patient est fixée à la première date d'incidence. La liste fournie ne contient qu'un seul enregistrement par patient.

Après réception de la liste, le TTP eHealth:

- attribuera un RN1 à chaque NISS unique de la liste FRC et transmettra la liste de conversion NISS/RN1 à la FRC (2.1.) ;
- attribuera un RN2 à chaque NISS unique de la liste FRC et transmettra la liste de conversion NISS/RN2 y compris les dates de diagnostic (NISS/RN2/Dt_BCnew) au conseiller en sécurité du Collège intermutualiste national (SPOC NIC) (2.2.) ;
- attribuera un RN3 à chaque NISS unique de la liste FRC. Le TTP eHealth pseudonymise chaque NISS unique en Hnew, l'identifiant utilisé de manière standard par eHealth dans le flux RHM depuis 2012. La liste de conversion RN3/Hnew y compris les dates de diagnostic (RN3/Hnew/Dt_BCnew) est transmise au responsable du SPF Santé publique - RHM (2.3.) ;
- attribuera un RN4 à chaque NISS unique de la liste FRC et transmettra la liste de conversion NISS/RN4 y compris les dates de diagnostic (NISS/RN4/Dt_BCnew) au responsable DWH (datawarehouse) de la Banque Carrefour de la sécurité sociale (étape 2.4.) ;
- attribuera un Cprojet à chaque NISS unique de la liste FRC et transmettra la liste de conversion RN1/RN2/RN3/RN4/Cprojet au TTP OA (BCSS) (2.5.).

Le SPOC CIN remplace chaque NISS unique de la liste NISS/RN2/Dt_BCnew par C1 et transmet la nouvelle liste de conversion C1/RN2/Dt_BCnew au TTP OA (BCSS) (étape 3).

La FRC sélectionne les données (Data_FRC) de la période de sélection demandée (*) sur la base du NISS. La FRC réduit tous les champs de dates à une valeur entière par rapport à la date d'incidence et remplace chaque NISS par le RN1 correspondant. La FRC transmet ensuite RN1/Data_FRC au TTP OA (BCSS) (étape 4).

Le responsable du SPF Santé publique – RHM établit, sur la base de Hnew, une liste de tous les codes¹ associés aux diagnostics principaux et secondaires des admissions de ces patients au cours de la période de sélection demandée (*), et transmet la liste (DiagCodAll) aux chercheurs de la VUB (étape 5).

Les chercheurs de la VUB sélectionnent uniquement les codes qui ont potentiellement un rapport avec les objectifs de l'étude (codes maladies cardiovasculaires et tumeurs pulmonaires radio-induites) et transmettent cette liste (DiagCodSel) au responsable du SPF Santé publique - RHM (étape 6).

Le responsable du SPF Santé publique - RHM sélectionne, sur la base de Hnew, uniquement les données d'admission au sein de la période de sélection demandée (*) qui sont associées à un code de la liste DiagCodSel des chercheurs de la VUB. Le responsable du SPF Santé publique - RHM réduit tous les champs de dates à une valeur entière par rapport à la date d'incidence et met les données sur RN3. Le responsable du SPF Santé publique - RHM transmet les données RN3/Data_RHM(*DiagCodSel*) au TTP OA (BCSS) (étape 7).

¹ Les variables RHM incluses dans la liste DiagCodAll sont :

- CODE_DIAGNOSE: code ICD pour les diagnostics principaux et secondaires pendant l'hospitalisation.
- TYPE_DIAGNOSE: pour distinguer les diagnostics principaux et secondaires.
- CODE_SYS_DIAGNOSE: type de système de codage (ICD-9-CM ou ICD-10-BE) pour le diagnostic.
- M2_CODE_PROCEDURE: code des interventions réalisées pendant l'hospitalisation.

Le responsable du DWH BCSS sélectionne, sur la base du NISS, les données BCSS pour la période de sélection (*) demandée, réduit tous les champs de dates à une valeur entière par rapport à la date d'incidence et met les données sur RN4. Le responsable du DWH BCSS transmet ensuite les données RN4/BCSS_Data au TTP OA (BCSS) (étape 8).

Le TTP OA (BCSS) attribue C2 à chaque C1 unique de la liste C1/RN2/Dt_BCnew (voir l'étape 3) et place C2/Dt_BCnew sur le DWH AIM (étape 9).

L'AIM sélectionne, sur la base de C2 et des dates de diagnostic Dt_BCnew, les données pour la période de sélection demandée (*), réduit tous les champs de dates à une valeur entière par rapport à la date d'incidence et transmet les données C2/Data_AIM au TTP OA (BCSS) (étape 10).

Le TTP OA (BCSS) met toutes les données sur Cprojet et place les données sur le DWH AIM PROJET (étape 11).

Une analyse de risque "small cell" est réalisée par le KCE (étape 12).

Les séries de données Cprojet/Data_FRC, Cprojet/Data_AIM, Cprojet/Data_BCSS et Cprojet/Data_RHM (*DiagCodSol*) sont mises à la disposition des chercheurs sur la base de Cprojet (étape 13).

(*) le trajet doit être parcouru 4 fois pour la sélection des données des périodes suivantes :

	<i>Période de sélection 1</i>	<i>Période de sélection 2</i>	<i>Période de sélection 3</i>	<i>Période de sélection 4</i>
<i>FRC</i>	<i>01/07/2008-31/06/2012</i>	<i>01/07/2012-31/12/2020</i>	<i>01/01/2021- 31/12/2025</i>	<i>01/01/2026-31/12/2030</i>
<i>BCSS</i>	<i>01/07/2008-31/06/2012</i>	<i>01/07/2012-31/12/2020</i>	<i>01/01/2021- 31/12/2025</i>	<i>01/01/2026-31/12/2030</i>
<i>RHM</i>	-	<i>01/01/2012-31/12/2020</i>	<i>01/01/2021- 31/12/2025</i>	<i>01/01/2026-31/12/2030</i>
<i>IMA</i>	<i>01/07/2008-31/06/2012</i>	<i>01/07/2012-31/12/2020</i>	<i>01/01/2021- 31/12/2025</i>	<i>01/01/2026-31/12/2030</i>

La liste NISS/Dt_BCnew doit chaque fois être transmise à nouveau au TTP eHealth pour chaque période de sélection (étape 1). Les étapes suivantes (2 à 13) se déroulent comme décrit.

Aux étapes 4, 7, 8 et 10, les fournisseurs de données réduisent les champs de dates à une valeur entière par rapport à la date d'incidence sur la base de la date d'incidence fournie. Les différents fournisseurs de données peuvent supprimer la liste Dt_BCnew puisque celle-ci est transmise à nouveau à chaque extraction de données par la FRC et le TTP eHealth.

11. Le Comité constate que le Comité éthique de l'UZ Brussel a formulé un avis positif pour cette étude.

II. COMPÉTENCE

12. En vertu de l'article 42, § 2, 3° de la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, la chambre sécurité sociale et santé du Comité de sécurité de l'information est compétente pour rendre une autorisation de principe concernant toute communication de données à caractère personnel relatives à la santé.

13. En vertu de l'article 42, § 2, 2°, a) de la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, la chambre sécurité sociale et santé du Comité de sécurité de l'information est également compétente pour rendre une autorisation de principe pour le couplage de données à caractère personnel issues du Registre du cancer à d'autres données externes.
14. En vertu de la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, la Banque Carrefour de la sécurité sociale recueille des données à caractère personnel auprès des institutions de sécurité sociale, les enregistre, procède à leur agrégation et les communique aux personnes qui en ont besoin pour la réalisation de recherches pouvant être utiles à la connaissance, à la conception et à la gestion de la protection sociale (article 5, § 1^{er}). Une telle communication de données à caractère personnel requiert une délibération de la chambre sécurité sociale et santé du Comité de sécurité de l'information (article 15, § 1^{er}). Etant donné que les chercheurs souhaitent avoir recours en l'occurrence à des données à caractère personnel socio-économiques du datawarehouse marché du travail et protection sociale, la chambre sécurité sociale et santé est compétente pour se prononcer.
15. 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, l'intervention de la Plate-forme eHealth en tant qu'organisation intermédiaire pour le couplage et la pseudonymisation de données à caractère personnel requiert une délibération du Comité de sécurité de l'information.
16. Le Comité de sécurité de l'information estime par conséquent qu'il est compétent.

III. EXAMEN

A. ADMISSIBILITÉ

17. Le traitement de données à caractère personnel est uniquement autorisé pour des finalités déterminées, explicites et légitimes et le traitement de données à caractère personnel relatives à la santé est en principe interdit.²
18. 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 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.
19. Le Comité estime que le traitement de données à caractère personnel est admissible.

B. FINALITÉ

² Art. 9, point 1 RGPD.

20. Conformément à l’art. 5, b) du RGPD, le traitement de données à caractère personnel est uniquement autorisé pour des finalités déterminées, explicites et légitimes.
21. L’objectif de l’étude est de dresser la carte du parcours des femmes atteintes de cancer du sein sur une période de vingt ans à partir du diagnostic. Dans ce projet, l’accent est mis sur deux aspects : (1) la consommation de soins de ces femmes en Belgique et (2) l’étude des effets à long terme (à savoir la cardiotoxicité et les tumeurs pulmonaires radio-induites) du traitement dans cette cohorte.
22. A la lumière de ce qui précède, le Comité est par conséquent d’avis qu’il existe un fondement admissible pour le traitement des données à caractère personnel pseudonymisées relatives à la santé envisagé.

C. PROPORTIONNALITÉ

23. Conformément à l’art. 5, b) et c) du RGPD, les données à caractère personnel doivent être adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont obtenues et pour lesquelles elles sont traitées ultérieurement.
24. Les données en provenance de la **Fondation registre du cancer** concernent :
 - *Les diagnostics de cancer du sein au cours de l’année d’incidence 2010* : Ces données sont nécessaires pour l’identification de la cohorte.
 - *Les diagnostics de cancer du poumon chez les femmes de cette cohorte* : L’étude de l’apparition de tumeurs pulmonaires (radio-induites) dans la cohorte de femmes atteintes de cancer du sein constitue un aspect important de l’étude (voir le sous-objectif 2) ;
 - *Décès par cause sur la base des certificats de décès chez les femmes de la cohorte* : L’étude de la mortalité au sein de la cohorte de femmes atteintes de cancer du sein constitue un aspect important de l’étude (voir les sous-objectifs 2 et 3) ;
25. L’analyse des **données AIM** permet de déterminer le coût direct du traitement primaire de cancer du sein. Ceci est un composant important du calcul du coût sociétal d’un traitement de cancer du sein (voir les sous-objectifs 1 et 3). Ensuite, l’analyse des données AIM permet d’identifier les effets à long terme attendus dans la cohorte et d’en calculer les coûts directs. Ceci est un composant important du calcul du coût sociétal suite à la cardiotoxicité et les tumeurs pulmonaires radio-induites dans le cadre du cancer du sein (voir les sous-objectifs 2 et 3).
26. L’analyse des **données RHM** (Source: SPF Santé publique) permet d’identifier les effets à long terme attendus dans la cohorte et d’en calculer les coûts directs. Ceci est un composant important du calcul du coût sociétal suite à la cardiotoxicité et les tumeurs pulmonaires radio-induites dans le cadre du cancer du sein (voir les sous-objectifs 2 et 3).

27. L'analyse des données relatives à la situation socio-économique et à l'emploi en provenance du **datawarehouse de la BCSS** permettent d'évaluer les coûts indirects d'une traitement du cancer du sein (voir les sous-objectifs 1, 2 et 3).
28. Le numéro NISS / numéro d'identification unique sera uniquement utilisé à des fins de sélection et de couplage. Cette variable est convertie en numéro d'identification pseudonymisé (variable : Cprojet) qui est mis à la disposition des chercheurs.
29. Cette étude prévoit la réalisation de nombreuses analyses de séries chronologiques ainsi que d'analyses de régression Cox qui doivent permettre aux chercheurs d'établir des modèles de prédiction 'time-to-event'. Ce type d'analyses statistiques réfèrent à des analyses de durée jusqu'à un événement déterminé (p.ex. admission, événement cardiovasculaire, diagnostic de tumeur pulmonaire). Ceci requiert que les dates du point de départ et de l'évènement soient connues, afin de non seulement pouvoir prédire si un évènement aura lieu, mais aussi quand il se produira. Pour éviter que les chercheurs aient accès à des dates exactes, les fournisseurs de données convertiront ces dates en valeur entière par rapport à la date d'incidence, sur la base de la date d'incidence transmise.

Concrètement, il s'agit des variables suivantes :

	Variable	Remarque
FRC	<i>Date d'incidence (cancer du sein)</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>Date d'incidence (tumeur pulmonaire)</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>Date du statut</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
BCSS	<i>DWH_CIN : Date de début</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>DWH_CIN : Date de fin</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>DWH_IMANI_CMI : B_ziekte</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>DWH_INAMI_CMI : B-ival</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>DWH_INAMI_CMI : E_ival</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
RHM	<i>A2_YEAR_HOSP_IN</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>A2_MONTH_HOSP_IN</i>	
	<i>A2_DAY_HOSP_IN</i>	
	<i>A2_YEAR_HOSP_OUT</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>A2_MONTH_HOSP_OUT</i>	
	<i>A2_DAY_HOSP_OUT</i>	
	<i>A4_YEAR_BEDINDEX_IN</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>A4_MONTH_BEDINDEX_IN</i>	
	<i>A4_DAY_BEDINDEX_IN</i>	
	<i>M2_YEAR_PROCEDURE</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>M2_MONTH_PROCEDURE</i>	
<i>M2_DAY_PROCEDURE</i>		
AIM	<i>SS00015 - Date début de la prestation / date de la délivrance</i>	<i>Définie comme « nombre de jours par rapport à la date d'incidence »</i>
	<i>SS00110 – Date d'admission</i>	<i>(1) Définie comme « nombre de jours par rapport à la date d'incidence »</i>

		(2) Limitée à la date d'extraction 1 (en remplacement des variables du SFP Santé publique - RHM)
	SS00115 – Date de sortie	(1) Définie comme « nombre de jours par rapport à la date d'incidence » (2) Limitée à la date d'extraction 1 (en remplacement des variables du SFP Santé publique - RHM)

- 30.** Une liste de toutes les données demandées ainsi qu'une motivation circonstanciée par variable sont jointes en annexe.
- 31.** Le principe de proportionnalité implique que le traitement doit en principe être réalisé au moyen de données anonymes, ce qui est le cas en l'espèce. Cependant, si la finalité ne peut être réalisée au moyen de données anonymes, des données à caractère personnel pseudonymisées peuvent être traitées. Vu la nécessité de réaliser des analyses très détaillées à partir de ces données, les chercheurs ont besoin d'avoir accès à des données pseudonymisées afin d'être en mesure de réaliser ces analyses qu'ils ne pourraient pas réaliser à l'aide de données anonymes. Cette finalité justifie donc le traitement de données à caractère personnel pseudonymisées.
- 32.** Les données à caractère personnel ne peuvent être conservées sous une forme permettant l'identification des personnes concernées que pendant la durée nécessaire à la réalisation des finalités pour lesquelles les données à caractère personnel sont traitées.
- 33.** La période étudiée porte sur une période de 20 ans à compter de l'année d'incidence 2010 pour le diagnostic de cancer du sein (jusqu'au 31/12/2030). Quatre transmissions de données sont prévues, réparties dans le temps. Il s'agit des données pour les périodes suivantes = 1,5 ans pré-indice à 1,5 ans post-indice, 1,5 ans post-indice à 10 ans post-indice, 10 ans post-indice à 15 ans post-indice et 15 ans post-indice à 20 ans post-indice.
- 34.** Pour permettre les analyses pour l'ensemble de la période étudiée, les données des transmissions de données précédentes sont également mises à disposition à chaque nouvelle transmission de données. Dans le cadre du suivi scientifique du projet³, la durée de la disponibilité des données, à savoir la nouvelle transmission de données plus toutes les données des transmissions de données antérieures, est systématiquement fixée à quatre ans, à compter du moment où tous les fournisseurs de données ont mis leurs données à la disposition des chercheurs. Par conséquent, la série de données complète est conservée pendant 4 ans après la dernière transmission de données.
- 35.** Etant donné que le couplage des données entre les divers fournisseurs reste possible, il est important que le rapport entre la série de données pseudonymisées et l'identité des personnes ne se perde pas. L'extraction de données en phases permet d'éviter des retards dans le délai prévu pour le projet. Le Comité est dès lors d'avis que la Plate-forme eHealth peut conserver le lien entre les données pseudonymisées et l'identité des personnes pour cette finalité.

³ Par exemple pour le contrôle des données statistiques suite à des publications scientifiques à comité de lecture.

36. Compte tenu des finalités de l'étude, le Comité estime que la communication envisagée est adéquate, pertinente et non excessive.

D. TRANSPARENCE

37. Le responsable du traitement de données à caractère personnel, collectées à des fins déterminées, explicites et légitimes ou l'organisation intermédiaire doit, préalablement à la pseudonymisation des données, en principe communiquer certaines informations à la personne concernée.
38. Le responsable du traitement est dispensé de cette obligation lorsque la communication de telles informations se révèle impossible ou exigerait des efforts disproportionnés, en particulier pour le traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques sous réserve des conditions et garanties visées à l'article 89, paragraphe 1, du RGPD⁴.
39. Les données ne sont pas obtenues auprès des intéressés mêmes et elles sont pseudonymisées, de sorte que les chercheurs n'ont pas accès aux données à caractère personnel des personnes concernées. Ensuite l'ensemble de données est tellement volumineux (environ 13.000 patients) qu'une communication d'informations à tous les intéressés est irréalisable au niveau pratique⁵ et au niveau financier.
40. Le Comité estime par conséquent que la demande répond aux exigences de transparence.

E. MESURES DE SÉCURITÉ

41. Conformément à l'article 5, f) du RGPD, le demandeur doit prendre toutes les mesures techniques et organisationnelles appropriées pour protéger les données à caractère personnel. Ces mesures doivent garantir un niveau de protection adéquat, compte tenu, d'une part, de l'état de la technique en la matière et des frais qu'entraîne l'application de ces mesures et, d'autre part, de la nature des données à protéger et des risques potentiels.
42. 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.

⁴ Art. 14, alinéa 5, b) du RGPD.

⁵ Il est irréaliste d'obtenir une liste actualisée avec les données de contact de tous les intéressés compte tenu de l'ampleur de la série de données.

43. Conformément à l'article 9, point 3, du RGPD, le traitement de données à caractère personnel relatives à la santé peut uniquement être effectué sous la surveillance et la responsabilité d'un professionnel des soins de santé. C'est le cas en l'espèce.
44. Le Comité rappelle que lors du traitement de données à caractère personnel, le professionnel des soins de santé ainsi que ses préposés ou mandataires sont soumis au secret.
45. Il est interdit d'entreprendre toute action visant à convertir les données à caractère personnel pseudonymisées qui ont été communiquées en données à caractère personnel non pseudonymisées.
46. L'accès au serveur sécurisé sur lequel sont conservées les données fait l'objet de loggings par l'AIM. Une liste des chercheurs habilités est mise à la disposition de l'AIM par le chef de service du I-CHER.
47. Le Comité prend acte du fait qu'une analyse d'impact relative à la protection des données a été réalisée.
48. Le Comité constate enfin qu'une « small cell risk analysis » (SCRA) est prévue et que celle-ci sera réalisée par l'AIM.

Par ces motifs,

la chambre sécurité sociale et santé du comité de sécurité de l'information

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 et à la condition qu'une « Small Cell Risk Analysis » (SCRA) soit réalisée par le KCE afin d'exclure toute possibilité de réidentification des personnes concernées.

conclut que la Plate-forme eHealth est autorisée à conserver le lien entre la série de données pseudonymisées et le numéro d'identification réel de sorte que le couplage des données entre les différents fournisseurs reste possible.

Bart VIAENE
Président

Le siège de la chambre sécurité sociale et santé du comité de sécurité de l'information est établi dans les bureaux de la Banque Carrefour de la sécurité sociale, à l'adresse suivante: Quai de Willebroeck 38 - 1000 Bruxelles.

Bijlage: Selectie van variabelen

A. Toelichting en motivering

i. Algemeen

In deze studie zitten vier data-extracties vervat. Dit laat toe om de verschillende doelstellingen trapsgewijs in de tijd te realiseren in plaats van alles op het einde van de follow-up periode te analyseren. Dit maakt het mogelijk om de looptijd van het project efficiënt te gebruiken. Deze stapsgewijze benadering heeft als bijkomend voordeel dat de data incrementeel toeneemt in plaats van dat alles in een keer beschikbaar wordt gesteld. Dit alles zorgt ervoor dat, inzake datamanagement, alle procedures en protocollen beter beheersbaar en controleerbaar zijn.

De studie ambiëert het opstellen van predictiemodellen naar ‘time to event’ via tijdreeksanalyses en Cox regressie analyse. Deze statistische modellen analyseren de duurtijd vanaf indexdatum (diagnose borstkanker) tot een bepaald event (bijvoorbeeld: opnamedatum, cardiovasculair event en diagnose longkanker). Een vereiste hierbij is dat het beginpunt (diagnose borstkanker) en de gebeurtenis/event gekend zijn zodoende niet enkel predicties te berekenen naar het voorkomen van events maar ook naar het tijdstip (vanaf indexdatum) waarop het event plaatsvindt. Echter, om te vermijden dat de onderzoeker toegang krijgt tot exacte datums, zullen de dataleveranciers de exacte datums herleiden tot een integere waarde ten opzichte van de incidentiedatum op basis van de aangeleverde incidentiedatum.

ii. Met betrekking tot aanvraag van socio-economische variabelen

Deze aanvraag kadert in een onderzoek naar de maatschappelijke kost van borstkanker. Het inschatten van zowel directe als indirecte kosten zijn inherent bij dergelijke economische evaluaties vanuit het maatschappelijk perspectief. De sociaal-economische positie (bijvoorbeeld: arbeidsmarktpositie) laten toe om de indirecte kosten van zorg in te schatten. Daarnaast heeft de sociaal-economische positie -en de evolutie hiervan in de tijd- een belangrijk effect op de verschillende sociale (financiële) beschermingsmaatregelen in België. De uitgebreidheid van socio-economische variabelen binnen dit onderzoek is noodzakelijk om die sociaal-economische posities en sociale beschermingsmaatregelen correct te bepalen zodoende de indirecte kosten correct in te schatten en op te volgen in de tijd.

B. Variabelen die noodzakelijk zijn voor de selectie- en/of koppelingsdoeleinden

Data-leverancier	Variabele	Bewijs van proportionaliteit	Selectieperiode			
			1	2	3	4
SKR	INSZ	Het INSZ-nummer/uniiek identificatienummer zal enkel gebruikt worden voor selectie- en koppelingsdoeleinden. Deze variabele wordt omgezet naar een gepseudonimiseerd identificatienummer (variabele: C _{project}) dat ter beschikking gesteld wordt voor de onderzoekers.	X	X	X	X
AIM	SS00010 / PP0010 – Identificatie rechthebbende		X	X	X	X
KSZ	INSZ – Identificatienummer van de sociale zekerheid		X	X	X	X
FOD VVVL (MZG)	PATNUM		X	X	X	X

SKR	Geslacht	Deze variabele is noodzakelijk als selectievariabele voor het samenstellen van de cohort. Op basis van de combinatie van variabelen “Geslacht”, “Incidentiedatum” en “ICD10 (numeriek)” kunnen alle vrouwelijke patiënten met een diagnose van borstkanker tijdens incidentiejaar 2010 geïdentificeerd worden.	X	X	X	X
	Incidentiedatum (borstkanker) = Indexdatum	Deze variabele is noodzakelijk als selectievariabele voor het samenstellen van de cohort. In kader van het dataminimalisatieprincipe wordt geopteerd voor een methode waarbij geen exacte datums ter beschikking gesteld worden van de onderzoekers in de aangeleverde dataset. De exacte incidentiedatum (borstkanker) wordt door SKR aangeleverd aan alle dataleveranciers opdat zij op hun beurt alle datumvariabelen kunnen omzetten naar ‘aantal dagen t.o.v. indexdatum’. De lijst met incidentiedatums en patient-ID’s worden niet bewaard door de dataleveranciers, maar worden bij elke nieuwe datalevering opnieuw aangeleverd door SKR waarbij eHealth codeersleutels kan voorzien. In geval van meer dan één diagnose van borstkanker tijdens het indexjaar, geldt de eerste diagnose (incidentiedatum) als indexdatum. De aangeleverde lijst met incidentiedatums omvat slechts één record per patiënt. De indexdatum wordt gefixeerd bij de eerste data-extractie. Bij laattijdige correcties binnen indexjaar 2010 (cfr. wijziging in maand en/of dag), blijft de initiële indexdatum behouden tijdens data-extracties 2-4. Laattijdige correcties buiten het indexjaar (cfr. wijziging in jaar) resulteren in het verwijderen van de patiënt uit de dataset. Er worden geen nieuwe diagnoses (laattijdige correcties m.b.t. type tumor) t.o.v. de eerste data-extractie opgenomen in de dataset. Concreet betekent dit dat er ten opzichte van de eerste datalevering geen nieuwe patiënten toegevoegd kunnen worden aan de initiële selectie, het verwijderen van patiënten n.a.v. correcties is wel mogelijk. De conversiesleutel INSZ/Cproject blijft steeds dezelfde over het volledige verloop van het project. Deze sleutels dienen -evenals de aangeleverde databeschikbaar te blijven tot vier jaar na de aanlevering van de laatste dataset. Dit is noodzakelijk voor het voorkomend geval dat patiënten verwijderd worden uit de dataset.	X	X	X	X
	ICD10 (numeriek)	Deze variabele is noodzakelijk als selectievariabele voor het samenstellen van de cohort. Op basis van de combinatie van variabelen “Geslacht”, “Incidentiedatum” en “ICD10 (numeriek)” kunnen alle vrouwelijke patiënten met een diagnose van borstkanker tijdens incidentiejaar 2010 geïdentificeerd worden. Concreet betreft het ICD-10 codes C50.x en D05.x. <i>(!) Deze variabele is daarnaast ook noodzakelijk voor het identificeren van vrouwen in de cohort die</i>	X	X	X	X

		<i>gediagnosticeerd worden met een longtumor tot 20 jaar na incidentiedatum (cfr. infra).</i>				
--	--	---	--	--	--	--

Tabel 1. Variabelen die noodzakelijk zijn voor koppelings- en/of selectiedoeleinden

C. Selectie van variabelen vanuit de databank van SKR

Dataleverancier	: SKR
Aantal gevraagde variabelen (excl. variabelen voor koppeling en selectie)	: 22

Variabele	Bewijs van proportionaliteit	Selectieperiode			
		1	2	3	4
Leeftijd (bij diagnose)	Deze variabele is van belang voor het omschrijven van de cohort, en eventueel voor het corrigeren voor leeftijd (jaar) als risicofactor voor langetermijneffecten (zoals cardiovasculaire aandoeningen).	X	X	X	X
Incidentiedatum (longtumor) Incidentiedatum (borstkanker)	Incidentiedatum (longtumor of borstkanker) wordt omgezet naar een gecodeerde datum ('aantal dagen t.o.v. indexdatum'). Deze variabele is noodzakelijk om secundaire tumoren te kunnen identificeren in de cohort.	X	X	X	X
ICD10 (numeriek)	Naast het belang van deze variabele als selectievariabele (cfr. supra), dient deze meegenomen te worden over de volledige onderzoekstermijn voor de identificatie van de diagnose van longtumoren in de cohort. Concreet betreft het ICD-10 code C34.x, alsook ICD-10 codes C50.x en D05.x.	X	X	X	X
cTNM (cT + cN + cM)	Deze variabelen laten toe om de stadiëring te bepalen van de tumoren waardoor het mogelijk wordt om de verschillende behandelingstrajecten uit te zetten in functie van de stadia van ziekte.	X	X	X	X
pTNM (pT + pN + pM)		X	X	X	X
Combined stage		X	X	X	X
Histologisch subtype	Deze variabele laat subanalyses toe volgens histologisch subtype wat van invloed kan zijn op het behandelingstraject van de patiënt.	X	X	X	X
ER	Deze variabele laat toe om de ER-receptorstatus te bepalen wat van invloed kan zijn op het behandelingstraject van de patiënt.	X			
PR	Deze variabele laat toe om de PR-receptorstatus te bepalen wat van invloed kan zijn op het behandelingstraject van de patiënt.	X			
HER2	Deze variabele laat toe om de HER2-receptorstatus te bepalen wat van invloed kan zijn op het behandelingstraject van de patiënt.	X			
Aantal tumoren	Deze variabele geeft het aantal geregistreerde tumoren weer bij de patiënt en maakt het (in combinatie met variabele "Multiple") mogelijk om na te gaan of het borstcarcinoom de eerste gediagnosticeerde tumor is bij de patiënt.	X	X	X	X
Multiple	Deze variabele geeft inzicht in het chronologisch voorkomen van tumoren waardoor het borstcarcinoom met zekerheid als voorafgaande tumor aangeduid kan worden. Indien andere tumortypes gediagnosticeerd worden na het borstcarcinoom, zullen deze ook mee opgenomen worden in de telling, zonder dat de records van deze tumoren zullen worden meegedeeld.	X	X	X	X
Lateraliteit	Deze variabele geeft inzicht in de lateraliteit van de tumor. Op deze manier kan een onderscheid gemaakt worden tussen	X	X	X	X

	linker- en rechterborstcarcinomen, wat van belang kan zijn voor het inschatten van de langetermijneffecten.				
Vitale status	Deze variabelen maken het mogelijk om te bepalen of (en tot welke datum) patiënten uit de cohort nog in leven, dan wel overleden, of <i>lost to follow-up</i> zijn. Status datum wordt geconverteerd naar een gecodeerde datum ('aantal dagen t.o.v. indexdatum').	X	X	X	X
Status datum		X	X	X	X
Onmiddellijke doodsoorzaak	Deze variabelen zijn noodzakelijk om de sterfte en de oorzaak-specifieke doodsoorzaak in de cohort te bepalen.	X	X	X	X
Intermediaire doodsoorzaak 1		X	X	X	X
Intermediaire doodsoorzaak 2		X	X	X	X
Oorspronkelijke doodsoorzaak		X	X	X	X
Bijkomende doodsoorzaak 1		X	X	X	X
Bijkomende doodsoorzaak 2		X	X	X	X
Bijkomende doodsoorzaak 3		X	X	X	X

Tabel 2. Selectie van variabelen vanuit de databank van Stichting Kankerregister

D. Selectie van variabelen vanuit de databank van IMA⁶

Dataleverancier	: AIM
Aantal gevraagde variabelen (excl. variabelen voor koppeling en selectie)	: 21 (Gezondheidszorgen), 10 (Farmanet), 23 (Populatiegegevens)

Bestand(en)	Variabele	Bewijs van proportionaliteit	Selectieperiode			
			1	2	3	4
Gezondheidszorgen en Farmanet	SS00015 – Begindatum verstrekking / Datum van aflevering	Deze variabele laat toe na te gaan wanneer een bepaalde verstrekking (voor de eerste keer) gefactureerd werd. Het betreft verpleegdagen, (toediening van) farmaceutische producten en prestaties. Deze datum wordt omgezet naar een gecodeerde datum ('aantal dagen t.o.v. indexdatum'). In combinatie met variabelen SS00020, SS00050 en SS00055 kunnen de directe zorgkosten op verschillende tijdsintervallen bepaald worden.	X	X	X	X
Gezondheidszorgen en Farmanet	SS00020 – Nomenclatuur / Code categorie geneesmiddel	De nomenclatuurcode of de categorie van het geneesmiddel (zoals voorgeschreven door het RIZIV) laten toe om na te gaan voor welke	X	X	X	X

		verstrekkingen de patiënt een vergoeding ontvangen heeft.				
Gezondheidszorgen en Farmanet	SS00050 – Aantal gevallen / Hoeveelheid	Deze variabele geeft het aantal keer weer dat een verstrekking verleend is, het aantal leveringen/afgeleverde eenheden of hoeveelheden of het aantal keer dat een eenheidstarief aangerekend werd op een bepaalde dag/periode. Deze variabele laat toe om de totale kost voor een specifieke verstrekking te berekenen.	X	X	X	X
Gezondheidszorgen	SS00055 – Aantal dagen	Deze variabele geeft - vergelijkbaar met SS00050- het aantal gefactureerde dagen of forfaits weer in een bepaalde periode. Het IMA adviseert om beide variabelen op te nemen omdat de informatie niet altijd op correcte wijze geregistreerd wordt.	X	X	X	X
Gezondheidszorgen en Farmanet	SS00060 – Bedrag terugbetaling / ZIV-tegemoetkoming 1	Deze variabele geeft het bedrag weer dat terugbetaald werd door de verzekeringsinstelling aan de patiënt. Deze variabele vormt, in combinatie met variabelen SS00160 en SS00165, de belangrijkste bron van informatie over de directe medische kosten van het zorggebruik.	X	X	X	X
Gezondheidszorgen	SS00065(A) – Zorgverstrekkers-identificatiecode	Deze variabelen laten toe om de zorgverstrekker die verantwoordelijk is voor het uitvoeren van de prestatie te identificeren. Dit vormt een belangrijke bron van informatie omtrent het voorkomen van ziektes/aandoeningen. Variabele SS00065(A) wordt in gecodeerde vorm aangeleverd.	X	X	X	X
Gezondheidszorgen	SS00065(B) – Zorgverstrekkers-bekwaming		X	X	X	X
Gezondheidszorgen	SS00075 – Identificatie instelling	Deze variabele is noodzakelijk om te traceren in welke instelling de rechthebbende van de verleende prestaties opgenomen is (of verblijft). Deze variabele wordt voorzien in een unieke gepseudonimiseerde code per voorziening.	X	X	X	X

Gezondheidszorgen	SS00080 – Dienstcode	Deze variabele laat toe om in detail de kostenplaats/dienstcode voor de verleende zorg binnen een bepaalde zorgvoorziening te bepalen.	X	X	X	X
Gezondheidszorgen	SS00085 – Plaats van verstrekking	Deze variabele laat toe om - op gecodeerde wijze- de instelling te identificeren waar de prestatie werkelijk uitgevoerd werd.	X	X	X	X
Gezondheidszorgen en Farmanet	SS00095 – Code gerechtigde 1 (CG1)	Deze variabelen laten toe om het sociaal statuut van de patiënt te bepalen. Deze variabelen laten toe om het sociaal statuut gelinkt aan de primaire borstkankerbehandeling in detail te bestuderen. In functie van doelstelling 2 is dergelijk detail minder noodzakelijk waardoor deze variabelen bij de volgende data-extracties (2 t.e.m. 4) worden vervangen door PP0030 en PP0035.	X			
	SS00100 – Code gerechtigde 2 (CG2)		X			
Gezondheidszorgen	SS00105 – Nummer derde	Deze variabele laat toe om - op gecodeerde wijze- de instelling of verstrekker die de prestatie gefactureerd heeft te identificeren.	X	X	X	X
Gezondheidszorgen	SS00110 – Datum opname	Deze variabelen geven aan wanneer een patiënt opgenomen/ontslagen is uit een zorgvoorziening. Beide variabelen worden geconverteerd naar gecodeerde variabele ('aantal dagen t.o.v. indexdatum'). Deze variabelen worden enkel opgevraagd tijdens de eerste data-extractie aangezien MZG-variabelen niet beschikbaar zijn tijdens deze extractie. Deze variabele wordt enkel opgevraagd tijdens de eerste data-extractie aangezien de MZG-variabelen niet beschikbaar zijn tijdens deze extractie.	X			
	SS00115 – Datum ontslag		X			
Gezondheidszorgen en Farmanet	SS00135 – Nummer product	Deze code staat voor de code van het gefactureerde farmaceutisch product. Hierdoor kan exact bepaald worden welke geneesmiddelen de patiënt	X	X	X	X

		gebruikt. Hieruit kan o.a. afgeleid worden aan welke aandoeningen de patiënt lijdt.				
Gezondheidszorgen	SS00150 – Gefactureerde nomenclatuurcodes	De gefactureerde nomenclatuurcode geeft de (pseudo-) nomenclatuurcode weer voor de verstrekking die de zorgverstrekker aangerekend heeft aan de patiënt. Het geeft informatie over pseudocodes waardoor de specifieke aard van de verstrekking bepaald kan worden.	X	X	X	X
Gezondheidszorgen en Farmanet	SS00160 – Persoonlijke tussenkomst / Persoonlijk aandeel	Deze variabele geeft het reëel persoonlijk aandeel (wettelijk remgeld) weer dat de patiënt heeft betaald voor de verstrekking. In combinatie met variabelen SS00060 en SS00165 kan hierdoor de directe medische kost berekend worden.	X	X	X	X
Gezondheidszorgen	SS00165 – Bedrag supplement of bedrag voor niet-vergoedbare producten, verstrekkingen of diensten	Deze variabele geeft informatie over supplementen die aangerekend werden aan de patiënt bovenop het persoonlijk aandeel. Deze supplementen vormen, in combinatie met variabelen SS00060 en SS00160, de basis voor de berekening van de directe medische kosten.	X	X	X	X
Farmanet	SS00165 – Vermindering verzekeringstegemoetkoming/Bijdrage van de apothekers	Deze variabele geeft de vermindering op de verzekeringstegemoetkoming van geneesmiddelen weer. Naar analogie van variabele SS00165 in het bestand Gezondheidszorgen, is deze variabele noodzakelijk om de kost van het geneesmiddelengebruik te berekenen.	X	X	X	X
Gezondheidszorgen	SS00170 – Code implant	Deze variabele geeft de identificatiecode voor implantaten weer (zoals gefactureerd in variabele SS00150). Dit kan van belang zijn om het zorggebruik in kaart te brengen bij vrouwen die reconstructieve heelkunde ondergaan na een borstamputatie.	X	X	X	X
Gezondheidszorgen	SS00175 – Derdebetalersregeling	Deze variabele laat toe om te bepalen of een patiënt al dan niet geniet van een	X	X	X	X

		voorkeursregeling volgens het derdebetalerssysteem. Dit kan relevant zijn voor patiënten die recht hebben op een verhoogde tegemoetkoming (identificeerbaar in variabele SS00160).				
Farmanet	SS00195 – ZIV-tegemoetkoming 2	Deze variabele geeft de som weer van de aangerekende honoraria en laat een gedetailleerde berekening van de kost van het geneesmiddelengebruik toe.	X	X	X	X
Populatiedata	PP1010 – Aard VT/OMNIO	Deze variabele geeft weer of de patiënt recht heeft op een verhoogde tegemoetkoming (voorheen OMNIO-statuut). Deze variabele laat toe om het sociaal statuut gelinkt aan de primaire borstkankerbehandeling in detail te bestuderen. In functie van doelstelling 2 is dergelijk detail minder noodzakelijk waardoor deze variabelen bij de volgende data-extracties worden vervangen door CHRONICAL_YN.	X			
Populatiedata	PP2001 – Forfait B geneeskundige zorgen	Deze variabelen laten toe om te bepalen of de patiënt recht heeft op de omschreven tegemoetkomingen/verstrekkingen. Deze variabele laat toe om het sociaal statuut, gelinkt aan de primaire borstkankerbehandeling, in detail te bestuderen. In kader van de berekening van de incrementele kost (doelstelling 1) is het van belang om exact te bepalen welke sociale maatregelen een impact hebben op de vermindering in de directe en indirecte kosten van zorg (tot 1,5 jaar pre- en post-indexdatum). In functie van doelstelling 2 is dergelijk detail minder noodzakelijk waardoor deze variabelen bij de volgende data-extracties worden vervangen door CHRONICAL_YN.	X			
	PP2002 – Forfait C geneeskundige zorgen					
	PP2003 – Kinesithérapie E of fysiotherapie					
	PP2004 – Verhoogde kinderbijslagen					
	PP2005 – Toelage voor integratie van gehandicapten					
	PP2006 – Toelage voor hulp aan ouderen (cat III, IV of V)					

	PP2007 – Uitkering hulp aan derden					
	PP2008 – Verhoogde invaliditeitsuitkering wegens hulp aan derden					
	PP2009 – Forfaitaire tegemoetkoming voor hulp aan derden					
	PP2010 – Criterium hospitalisatie (120 dagen)					
	PP2011 – Criterium hospitalisaties (6 hospitalisaties)					
	PP3011 – Recht op toelage voor gehandicapten					
Populatiedata	PP0030 – Code gerechtigde 1 (CG1)	Deze variabelen laten toe om het sociaal statuut van de patiënt te bepalen. Dit kan een verklarende factor zijn voor (verhoogde) tegemoetkomingen voor verleende zorg.	X	X	X	X
	PP0035 – Code Gerechtigde 2 (CG2)		X	X	X	X
Populatiedata	PP3001 – Recht MAF – gezin	Deze variabelen laat toe te bepalen welk type MAF toegekend kan worden aan de patiënt en/of het gezin. Deze variabelen dienen samen geïnterpreteerd te worden.	X	X	X	X
Populatiedata	PP3004 – Terugbetaling en plafond - gezin	Deze variabelen omschrijft de plafonnering die de basis vormt voor de hoogte van de terugbetaling.	X	X	X	X
Populatiedata	PP3005 – Terugbetaling en plafond - individu of gedeelte van het gezin	Deze variabele omschrijft het recht, en de plafonnering op basis waarvan de hoogte van de terugbetaling bepaald wordt.	X	X	X	X
Populatiedata	PP4001 – Aantal dagen werkloosheid	Deze variabelen laten toe een inschatting te maken van het productiviteitsverlies. Deze variabele is noodzakelijk voor de berekening van de indirecte kosten.	X	X	X	X
	PP4002 – Aantal dagen invaliditeit		X	X	X	X
Populatiedata	CHRONICAL_YN	Deze flag laat toe om te bepalen of de patiënt in het afgelopen jaar recht had op	X	X	X	X

		een attest/statuut Chronische aandoening en/of een toelage voor gehandicapten. Deze variabele kan verklarend zijn voor verschillen in de kosten van zorg (als gevolg van de toekenning van verhoogde tegemoetkoming). Deze variabele wordt enkel opgevraagd tijdens data-extracties 2, 3 en 4 aangezien voor de eerste data-extractie een selectie van variabelen opgevraagd wordt die meer detail toelaat.				
Populatiedata	MAJOR_COVERAGE_YN	Deze flag laat toe om te bepalen of een patiënt al dan niet geniet van een voorkeursregeling. Dit kan relevant zijn voor patiënten die recht hebben op een verhoogde tegemoetkoming (identificeerbaar in variabele SS00160).	X	X	X	X
Populatiedata	ATC_ANAT_L + ATC_THER_L + ATC_PHAR_L + ATC_CHEM_L + ATC_PROD_L	Deze flag is een afgeleide van variabele SS00135 (nummer product) volgens de referentietabellen van de ATC codes zoals voorgesteld door het KCE.	X	X	X	X

Tabel 3. Selectie van variabelen vanuit de databank van het Inter mutualistisch Agentschap

E. Selectie van variabelen vanuit de databank van FOD VVVL (MZG databank)⁷

Dataleverancier	: MZG
Aantal gevraagde variabelen (excl. variabelen voor koppeling en selectie)	: 26 (totaal 34 variabelen waarvan 4 gehercodeerde exacte datums)

Bestand(en)	Variabele	Bewijs van proportionaliteit	Selectieperiode			
			1	2	3	4
DIAGNOSE, PROCEDUR, PROCRIZI en TESTRESU	CODE_AGR	Deze variabele geeft het erkenningsnummer van het ziekenhuis weer waardoor het mogelijk wordt om te bepalen waar (ziekenhuis) de zorg verleend werd. Deze variabele wordt voorzien in een unieke gepseudonimiseerde code per voorziening.		X	X	X
STAYHOSP, STAYINDX, DIAGNOSE, PROCEDUR,	STAYNUM	Een nieuw verblijfsnummer wordt toegekend bij iedere ziekenhuisopname waardoor differentiatie tussen verleende		X	X	X

⁷ De kwaliteit van gekoppelde data vanuit de MZG databank kan niet gegarandeerd worden voor data voor 2012. Om de kwaliteit van analyses niet te compromitteren, wordt geopteerd om enkel gebruik te maken van MZG data voor dataleveringen 2 tot en met 4 (doelstellingen 2 en 3).

PROCRIZI en TESTRESU		zorg tijdens verschillende ziekenhuisopnames mogelijk is.				
STAYHOSP	A2_YEAR_HOSP_IN	Deze variabelen maken het mogelijk om de opnamedatum voor de gespecificeerde hospitalisatie (STAYNUM) te bepalen. De combinatie van deze drie variabelen wordt omgezet naar één gecodeerde variabele: ‘aantal dagen t.o.v. indexdatum’.		X	X	X
	A2_MONTH_HOSP_IN			X	X	X
	A2_DAY_HOSP_IN			X	X	X
STAYHOSP	A2_YEAR_HOSP_OUT	Deze variabelen maken het mogelijk om de ontslagdatum voor de gespecificeerde hospitalisatie (STAYNUM) te bepalen. De combinatie van deze drie variabelen wordt omgezet naar één gecodeerde variabele: ‘aantal dagen t.o.v. indexdatum’.		X	X	X
	A2_MONTH_HOSP_OUT			X	X	X
	A2_DAY_HOSP_OUT			X	X	X
STAYHOSP	A2_HOSP_TYPE_CAT	Deze variabele geeft het verblijfstype van de hospitalisatie weer met o.a. een onderscheid tussen “alle soorten verblijf” en “oncologisch daghospitaal”. Dit kan van belang zijn om de zorgen verleend aan vrouwen uit het cohort te bepalen.		X	X	X
STAYHOSP	A2_HOSP_TYPE_FAC	Deze variabele geeft het verblijfstype weer op basis van facturatiegegevens met o.a. een onderscheid tussen “klassieke hospitalisatie” en “niet-chirurgische daghospitalisatie”. Dit kan van belang zijn om de zorgen verleend aan vrouwen uit het cohort te bepalen.		X	X	X
STAYHOSP	A2_TOTAL_NUMBER_DAY_FAC	In deze variabele wordt de som van het totaal aantal gefactureerde ligdagen tijdens de registratieperiode weergegeven. Het aantal gefactureerde ligdagen is van belang om de kosten van de verleende zorg te kunnen berekenen.		X	X	X
STAYHOSP	A2_CODE_READMISSION	Deze variabele codeert heropnames binnen hetzelfde jaar, binnen hetzelfde ziekenhuis. Het is een indicator voor toegenomen zorggebruik.		X	X	X
STAYHOSP	A2_CODE_STAT_INSURANCE	Deze code geeft de verzekeringsstatus van de patiënt weer tijdens het verblijf. Deze variabele kan verklarend zijn voor het toekennen van verhoogde tegemoetkomingen.		X	X	X
STAYHOSP	A2_CODE_PLACE_BEFORE_ADM	Deze variabele geeft weer waar de patiënt zich bevond vlak voor de opname. Onder de codes die toegekend worden o.a. “Ander ziekenhuis” en “Ongepland		X	X	X

		klassiek verblijf onmiddellijk aansluitend op een verblijf in daghospitalisatie". Deze codes kunnen van belang zijn om de trajecten van de cohort in kaart te brengen.				
STAYHOSP	A2_CODE_ADM	Deze variabele geeft het type opname weer (o.a. "Geplande hospitalisatie" en "Dringende opname ..."). Deze codes kunnen van belang zijn om de urgentie van de behandelingen (en daaraan gerelateerde kosten) in te schatten.		X	X	X
STAYHOSP	A2_CODE_DESTINATE	Deze variabele geeft weer waarnaar de patiënt verwezen wordt na ontslag. Onder de codes die toegekend worden o.a. "Overleden". Deze code kan van belang zijn om te bepalen of de patiënt overleden is tijdens het verblijf in het ziekenhuis.		X	X	X
STAYHOSP	A2_CODE_DISCHARGE	Deze variabele geeft of de patiënt ontslagen is uit het ziekenhuis (evt. op verwijzing). Onder de codes die toegekend worden o.a. "Overleden". Deze code kan van belang zijn om te bepalen of de patiënt overleden is tijdens het verblijf in het ziekenhuis.		X	X	X
STAYINDX en PROCRIZI	ORDER_BEDINDEX	Deze variabele geeft in chronologische volgorde de bedindexen (dienstcodes/kostenplaatsen) weer waar de patiënt verbleven heeft tijdens de gespecificeerde hospitalisatie. Deze codes kunnen van belang zijn om de trajecten van de cohort in kaart te brengen.		X	X	X
STAYINDX	A4_CODE_BEDINDEX_FAC	Deze variabele geeft de code weer van de dienst waarop een bepaalde verstrekking gefactureerd werd. Deze codes kunnen van belang zijn om de trajecten van de cohort in kaart te brengen.		X	X	X
STAYINDX	A4_YEAR_BEDINDEX_IN	Deze variabelen maken het mogelijk om de opnamedag op een bepaalde bedindex tijdens een gespecificeerde hospitalisatie te bepalen. De combinatie van deze drie variabelen wordt omgezet naar één gecodeerde variabele: 'aantal dagen t.o.v. indexdatum'		X	X	X
	A4_MONTH_BEDINDEX_IN			X	X	X
	A4_DAY_BEDINDEX_IN			X	X	X
STAYINDX	A4_NUMBER_DAY_FAC	Deze variabele geeft het aantal aaneensluitende dagen op een bepaalde bedindex weer en is van belang om de trajecten van de cohort in kaart te brengen.		X	X	X

STAYINDEX	A4_NUMBER_DAY_FAC_PREV	Deze variabele laat toe om te corrigeren voor gefactureerde dagen in het voorgaande registratiejaar. Deze variabele dient samen geïnterpreteerd te worden met variabele "A4_NUMBER_DAY_FAC".		X	X	X
DIAGNOSE en PROCEDUR	ORDER_SPEC	Deze variabele geeft het volgnummer van de verschillende specialisten weer die verantwoordelijk zijn voor de hoofd- en nevendiaagnoses tijdens een gespecificeerde hospitalisaties.		X	X	X
DIAGNOSE en PROCEDUR	TYPE_DIAGNOSE ⁸	Deze variabele maakt het mogelijk om hoofd- en nevendiaagnoses (variabele "CODE_DIAGNOSE") tijdens een gespecificeerde hospitalisatie te differentiëren.		X	X	X
DIAGNOSE en PROCEDUR	CODE_SYS_DIAGNOSE ⁸	Deze variabele geeft het soort coderingssysteem (ICD-9-CM of ICD-10-BE) weer voor de diagnose. Deze variabele is noodzakelijk om variabele "CODE_DIAGNOSE" te interpreteren.		X	X	X
DIAGNOSE	CODE_DIAGNOSE ⁸	Deze variabele vermeldt de ICD code voor de hoofd- en nevendiaagnoses tijdens een gespecificeerde hospitalisatie. Deze variabele is noodzakelijk om het zorggebruik te linken aan ziekte.		X	X	X
PROCEDUR	M2_CODE_SYS_PROCEDURE ⁸	Deze variabele geeft het soort coderingssysteem (ICD-9-CM of ICD-10-BE) weer voor de uitgevoerde ingreep. Deze variabele is noodzakelijk om variabele "M2_CODE_PROCEDURE" te interpreteren.		X	X	X
PROCEDUR	M2_CODE_PROCEDURE	Deze variabele omschrijft de code van de uitgevoerde ingrepen tijdens een gespecificeerde hospitalisatie. Deze variabele is noodzakelijk om uitgevoerde ingrepen (verleende zorgen) te bepalen.		X	X	X
PROCEDUR	M2_YEAR_PROCEDURE	Deze variabelen laten toe om de datum van de uitgevoerde ingreep te bepalen.		X	X	X
	M2_MONTH_PROCEDURE			X	X	X
	M2_DAY_PROCEDURE	De combinatie van deze drie variabele wordt omgezet naar één		X	X	X

⁸ Dit project focust op cardiotoxiciteit en radiatie-geïnduceerde langetermijneffecten van borstkankerbehandeling. Opnames in kader van niet-gerelateerde diagnoses zijn niet relevant. Voorafgaand aan de datalevering zal, in samenspraak met FOD VVVL, een lijst opgesteld worden van relevante ICD codes. Enkel opnames als gevolg van deze ICD codes (hoofd- en/of nevendiaagnoses) worden geëxtraheerd.

		<p>gecodeerde variabele: 'aantal dagen t.o.v. indexdatum'.</p> <p>Deze procedures stellen ons in staat om de radiotherapeutische bestralingsprogramma's te reconstrueren voor de vrouwen in het cohort. Op deze manier kan nagegaan worden welke bestralingstechnieken en -schema's een impact hebben op de langetermijneffecten van de behandeling.</p>				
PROCEDUR	M2_NUMBER_PROC_1_DAY	<p>Deze variabele geeft aan of de ingreep meerdere keren uitgevoerd werd op dezelfde dag. Deze variabele is noodzakelijk om kosten die voortvloeien uit een bepaalde ingreep te kunnen verklaren.</p>		X	X	X

Tabel 4. Selectie van variabelen vanuit de databank van de Federale Overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu (Minimale Ziekenhuis Gegevens)

F. Selectie van variabelen vanuit de databank van KSZ

Dataleverancier	: IKSZ
Aantal gevraagde variabelen (excl. variabelen voor koppeling en selectie)	: 21

Bron	Variabele	Bewijs van proportionaliteit	Selectieperiode			
			1	2	3	4
DWH_BCSS_NomenclatureVarDer	Nomenclatuur van socio-economische positie (Nomenc)	<p>Deze variabele laat toe om de socio-economische positie (tewerkstellingsvorm) te bepalen. Via deze variabele kan een inschatting gemaakt worden van het percentage en de snelheid van werkhervatting na ziekte.</p> <p>Dit is noodzakelijk om een inschatting te maken van de indirecte kosten van zorg (in functie van doelstellingen 1 en 2). Deze variabele wordt geïnterpreteerd in combinatie met de variabelen van "bijkomende socio-economische positie" (zie hieronder).</p> <p>Om de heridentificatie van personen in de dataset te verhinderen, wordt het detail van de variabelen beginnend met 1 (werkend) beperkt tot dit eerste niveau.</p>	X	X	X	X

	Bijkomende positie van socio-economische positie (hercodering)	<p>Om de heridentificatie van personen in de dataset te verhinderen, werden een aantal variabelen gehercodeerd waarbij het plausibel geacht wordt dat deze personen administratief werkend zijn, doch feitelijk hun beroepsactiviteit (tijdelijk) stopgezet hebben. Concreet betreft het volgende variabelen:</p> <ul style="list-style-type: none"> - Werkend en gedeeltelijk in terbeschikkingstelling voorafgaand aan pensioen (Werkend_en_gedeeltelijk_TBS) - Werkend en gedeeltelijke loopbaanonderbreking/deel tijds tijds krediet (interr_carriere_temps_partiel) - Werkend en halftijds brugpensioen (prepension_mi_temps) - Werkend en halftijds in een stelstel van werkloosheid met bedrijfstoelag (prepension_mi_temps) - Werkend en volledig in terbeschikkingstellin voorafgaand aan pensioen (Werkend_en_volledig_TB S) - Werkend en volledige loopbaanonderbreking/voll edig tijds krediet (interr_carriere_complete) - Werkend en volledig brugpensioen (prepension_complete) - Werkend en volledig in stelstel van werkloosheid met bedrijfstoelag (prepension_complete) 	X	X	X	X
DWH_STATBEL_CENSUS2011	Gedetailleerd onderwijsniveau (CD_ISCED_DET)	Deze variabele geeft een momentopname van het hoogst behaalde opleidingsniveau op 01/01/2011. Om echter na te gaan of de personen in het cohort een hoger onderwijsniveau behaalden in de jaren volgend op deze momentopname, dienen onderstaande vier variabelen (Opleidingsniveau ISCED, cat_shu, cat_études, CO_DIPLOME) eveneens opgevraagd te worden.	X	X	X	X

DWH_VDAB_FOREM_ACTIRIS_ADG	Opleidingsniveau ISCED	Deze variabele geeft het opleidingsniveau weer (Vlaamse gemeenschap), en kan verklarend zijn voor de snelheid van werkhervatting na een periode van ziekte.	X	X	X	X
DWH_CREF_diplomes	Studie categorie (niet-universitair onderwijs) (cat_shu)	Deze variabele geeft het opleidingsniveau weer (Franstalige gemeenschap), en kan verklarend zijn voor de snelheid van werkhervatting na een periode van ziekte.	X	X	X	X
	Studie categorie (cat_études)					
	Code van diploma (CO_DIPLOME)					
DWH_ONSS_UniStatbaseDMFA	Prestatietype (T_Prest)	Deze variabele geeft het arbeidsregime (voltijds, deeltijds, speciaal, onbepaald, afwezigheid wegens ziekte) op de laatste dag van het kwartaal weer. Deze variabele laat toe om te bepalen of (en in welk regime) de patiënt werkzaam is.	X	X	X	X
	Voltijdsequivalent exclusief gelijkgestelde dagen (Eqtpsp)	Deze variabele geeft het tewerkstellingspercentage per arbeidsplaats weer gedurende het kwartaal.	X	X	X	X
	Voltijdsequivalent inclusief gelijkgestelde dagen (Tauxaa)	Deze variabele geeft het tewerkstellingspercentage per arbeidsplaats weer gedurende het kwartaal.	X	X	X	X
DWH_CIN	Begindatum arbeidsongeschiktheid (Begindatum)	Deze variabele geeft de begin- en einddatum van arbeidsongeschiktheid weer. Deze variabelen worden geconverteerd naar twee gecodeerde variabelen ('aantal dagen t.o.v. indexdatum'). Deze variabele laat toe om het productiviteitsverlies te bepalen. Deze variabele is noodzakelijk voor de berekening van de indirecte kosten. De exacte berekening van indirecte kosten zijn inherent bij het bepalen van de maatschappelijke kost.	X	X	X	X
	Einddatum arbeidsongeschiktheid (Einddatum)					
	Aantal dagen arbeidsongeschiktheid (Aantal_dagen_ao)	Deze variabele wordt beschouwd als een controlevariabele op de variabelen "Begindatum" en "Einddatum". Deze variabele maakt het mogelijk om op eenvoudige wijze berekeningen te maken omtrent productiviteitsverlies en indirecte kosten.	X	X	X	X

	Samenstelling gezin (Samenst_gezin)	Deze variabele geeft de samenstelling van het gezin weer en kan verklarend zijn voor het toekennen van verhoogde tegemoetkomingen.	X	X	X	X
DWH_INAMI_CMI	Type ziekte (Type_ziekte)	Deze variabele is noodzakelijk om variabele "Diag_ziekte" te interpreteren. Deze variabele geeft het coderingssysteem (ICD-9 of ICD-10) weer waarin de ziekte gecodeerd werd.	X	X	X	X
	Medische code (Gr_ziekte)	Deze variabele laat toe om de oorzaak van invaliditeit (erkend door Geneeskundige Raad voor Invaliditeit) te bepalen. Dit laat toe om de invaliditeit te linken aan de borstkankerbehandeling.	X	X	X	X
	Diag ziekte (Diag_ziekte)	Deze variabele vermeldt de ICD code op basis waarvan de invaliditeit (variabele "Gr_ziekte") toegekend werd.	X	X	X	X
	Begin (B_ziekte)	Deze variabele geeft de begindatum van de primaire arbeidsongeschiktheid weer. Deze variabele wordt omgezet naar een geconverteerde variabele ('aantal dagen t.o.v. indexdatum'). Deze variabele wordt beschouwd als een controlevariabele op de variabele "Begindatum".	X	X	X	X
	Begindatum [invaliditeit] (B_inval)	Deze variabele geeft de begin- en einddatum van invaliditeit weer. Deze variabelen worden geconverteerd naar twee gecodeerde variabelen ('aantal dagen t.o.v. indexdatum'). Deze variabele laat toe om het productiviteitsverlies te bepalen. Deze variabele is noodzakelijk voor de berekening van de indirecte kosten. De exacte berekening van indirecte kosten zijn inherent bij het bepalen van de maatschappelijke kost.	X	X	X	X
	Einddatum [invaliditeit] (E_inval)					
	R_exclus	Deze variabele is noodzakelijk om te corrigeren bij het samenvoegen van data uit verschillende kwartalen	X	X	X	X

Tabel 5. Selectie van variabelen vanuit de databank van de Kruispuntbank van de Sociale Zekerheid