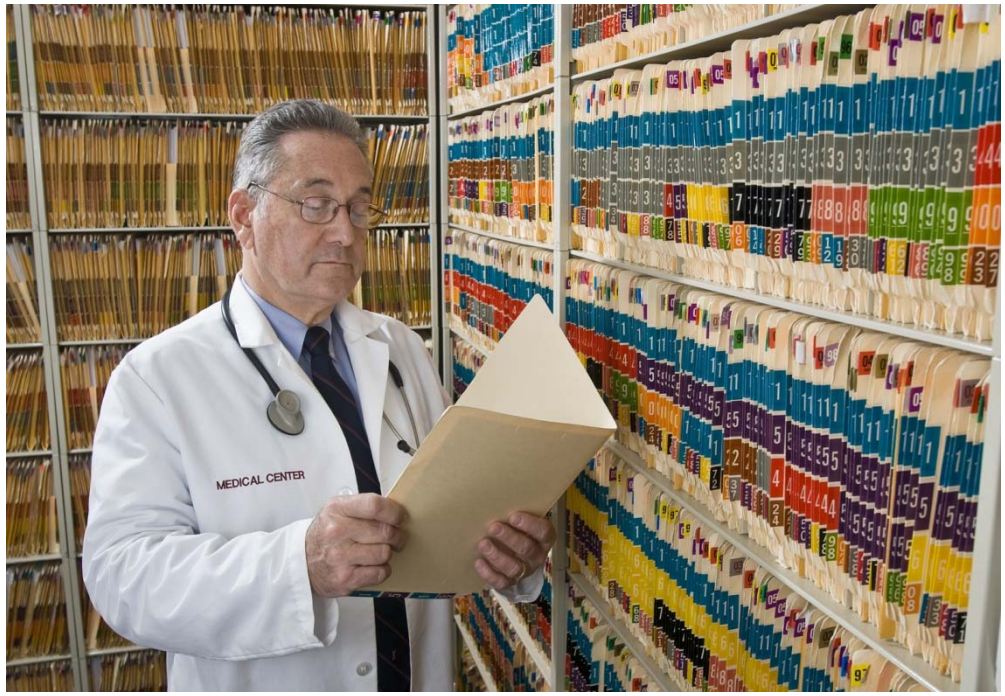
KCE REPORT 208

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

COUPLAGE DE L'ÉCHANTILLON PERMANENT AVEC LES DONNÉES HOSPITALIÈRES

ÉTUDE DE FAISABILITÉ ET DE REPRÉSENTATIVITÉ DES DONNÉES





KCE REPORT 208
HEALTH SERVICES RESEARCH

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

COUPLAGE DE L'ÉCHANTILLON PERMANENT AVEC LES DONNÉES HOSPITALIÈRES

ÉTUDE DE FAISABILITÉ ET DE REPRÉSENTATIVITÉ DES DONNÉES

CECILE CAMBERLIN, CECILE DUBOIS, TONIO DI ZINNO, RAGNA PREAL, JOERI GUILLAUME, REMI VAN DE REYD, CARINE VAN DE VOORDE, STEFAAN VAN DE SANDE, STEPHAN DEVRIESE

COLOPHON

Titre :	Couplage de l'échantillon permanent avec les données hospitalières Étude de faisabilité et de représentativité des données
Auteurs :	Cécile Camberlin (KCE), Cécile Dubois (KCE), Tonio Di Zinno (AIM), Ragna Préal (IMA), Joeri Guillaume (IMA), Remi Van de Reyd (IMA), Carine Van de Voorde (KCE), Stefaan Van De Sande (KCE), Stephan Devriese (KCE)
Relecture :	Nathalie Swartenbroekx (KCE), France Vrijens (KCE)
Stakeholders :	Nathalie Bossuyt (Wetenschappelijk Instituut Volksgezondheid – Institut Scientifique de Santé Publique), Joëlle Carton (Cellule Technique de traitement de données relatives aux hôpitaux - Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen), Lut De Prins (RIZIV - INAMI), Ri De Ridder (RIZIV - INAMI), Pascal Meeus (Cellule Technique de traitement de données relatives aux hôpitaux - Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen), Ingrid Mertens (Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen - Cellule Technique de traitement de données relatives aux hôpitaux), Yves Parmentier (Cellule Technique de traitement de données relatives aux hôpitaux - Technische Cel voor de verwerking van de gegevens met betrekking tot de ziekenhuizen), Alain Quewet (SPF Santé publique – FOD Volksgezondheid), Guy Van Camp (FOD Sociale Zekerheid – SPF Sécurité sociale), Johan Van Der Heyden (Wetenschappelijk Instituut Volksgezondheid – Institut Scientifique de Santé Publique), Peter Willemé (Federaal Planbureau – Bureau fédéral du Plan)
Validateurs externes :	Dirk De Wachter (Vlaams Agentschap Zorg en Gezondheid), Niel Hens (Universiteit Hasselt), Véronique Tellier (Service public de Wallonie - DG05 - DSE - OWS)
Autres intérêts déclarés :	Aucun intérêt déclaré
Layout :	Ine Verhulst, Sophie Vaes

Disclaimer :

- **Les experts externes ont été consultés sur une version (préliminaire) du rapport scientifique. Leurs remarques ont été discutées au cours des réunions. Ils ne sont pas co-auteurs du rapport scientifique et n'étaient pas nécessairement d'accord avec son contenu.**
- **Une version (finale) a ensuite été soumise aux validateurs. La validation du rapport résulte d'un consensus ou d'un vote majoritaire entre les validateurs. Les validateurs ne sont pas co-auteurs du rapport scientifique et ils n'étaient pas nécessairement tous les trois d'accord avec son contenu.**
- **Finalement, ce rapport a été approuvé à l'unanimité par le Conseil d'administration.**
- **Le KCE reste seul responsable des erreurs ou omissions qui pourraient subsister de même que des recommandations faites aux autorités publiques.**

Date de publication :	17 octobre 2013
Domaine :	Health Services Research (HSR)
MeSH :	Hospital records ; Medical record Linkage ; Sampling Studies ; Medical Informatics
Classification NLM :	W 26.55.I4
Langue :	Français
Format :	Adobe® PDF™ (A4)
Dépot légal :	D/2013/10.273/66

Copyright :

Les rapports KCE sont publiés sous Licence Creative Commons « by/nc/nd »
http ://kce.fgov.be/fr/content/a-propos-du-copyright-des-rapports-kce.



Comment citer ce rapport ?

Camberlin C, Dubois C, Di Zinno T, Préal R, Guillaume J, Van de Reyd R, Van de Voorde C, Van de Sande S, Devriese S. Couplage de l'échantillon permanent avec les données hospitalières. Health Services Research (HSR). Bruxelles : Centre Fédéral d'Expertise des Soins de Santé (KCE). 2013. KCE Reports 208B. D/2013/10.273/66.

Ce document est disponible en téléchargement sur le site Web du Centre Fédéral d'Expertise des Soins de Santé.

TABLE DES MATIÈRES

LISTE DES FIGURES **3**
LISTE DES TABLES **5**
LISTE D'ABRÉVIATIONS **7**
■ **RAPPORT SCIENTIFIQUE** **8**
1 INTRODUCTION **8**
1.1 CONTEXTE 8
1.2 QUESTIONS DE RECHERCHE 10
2 DESCRIPTION DES BASES DE DONNÉES **10**
2.1 L'ÉCHANTILLON PERMANENT 10
2.2 LES DONNÉES DE LA CELLULE TECHNIQUE 11
2.2.1 Données médicales hospitalières 11
2.2.2 Données financières hospitalières 11
2.2.3 Données hospitalières couplées 11
3 CADRE CONCEPTUEL ET NOTION DE SÉJOUR HOSPITALIER **13**
3.1 DÉFINITIONS RELATIVES AUX ENSEMBLES DE SOINS : SÉJOUR HOSPITALIER, ÉPISODE DE SOINS ET TRAJET DE SOINS 13
3.2 INVENTORISATION DES COÛTS LIES À UN ENSEMBLE DE SOINS 16
4 RECONSTITUTION DES SÉJOURS ET COUPLAGE **17**
4.1 MÉTHODOLOGIE 17
4.1.1 Reconstitution des séjours hospitaliers dans les données EPS 18
4.1.2 Calcul des remboursements 27
4.1.3 Couplage des séjours hospitaliers 27
4.2 RÉSULTATS 30
4.2.1 Reconstitution des séjours hospitaliers dans les données EPS 30
4.2.2 Couplage des séjours hospitaliers 32
5 TESTS DE REPRÉSENTATIVITÉ **37**
5.1 MÉTHODOLOGIE 37
5.1.1 Importance de la représentativité d'un échantillon 37

5.1.2 Méthodes pour tester la représentativité d'un échantillon 38
5.2 RÉSULTATS 40
5.2.1 Résultats globaux....40
5.2.2 Résultats au niveau APR-DRG....62
5.2.3 Exemple d'analyses par APR-DRG 68
5.3 CONCLUSION GÉNÉRALE SUR LES TESTS DE REPRÉSENTATIVITÉ....87
6 DISCUSSION ET CONCLUSIONS....88
■ **RÉFÉRENCES....92**

LISTE DES FIGURES

Figure 1 – Trajet de soins, épisode de soins et séjour hospitalier 14
Figure 2 – Épisode de soins à plusieurs admissions 15
Figure 3 – Séjours sélectionnés (vert) ou non (rouge) dans chacune des bases de données et après couplage EPS-TCT 19
Figure 4 – Schéma de boucle de l'algorithme de reconstitution des séjours dans les données EPS 24
Figure 5 – Technique du couplage 29
Figure 6 – Détail de la procédure de couplage EPS-TCT (les pourcentages sont toujours exprimés sur le nombre total de séjours EPS). 34
Figure 7 – Nombre de séjours EPS en fonction de la différence en jours (en valeur absolue) avec la date de début TCT, par année (séjours sans différence exclue) 35
Figure 8 – Couplage ou non des séjours en fonction du type d'hospitalisation (les quarts de cercle correspondent au pourcentage de séjours EPS)....36
Figure 9 – Pourcentage des séjours EPS non-couplés par hôpital selon le nombre de séjours EPS par hôpital. 37
Figure 10 – Poids AIM et Poids TCT par catégorie d'âge et par sexe - 2009....41
Figure 11 – Répartition des séjours par catégorie d'âge– 2009....44
Figure 12 – Fonction de Distribution empirique (Age) : juxtaposition des courbes de l'échantillon EPS-TCT (poids TCT) et de la population TCT de référence –2009 44
Figure 13 – Fréquences EPS-TCT pondérées par province du domicile du patient (poids AIM) par rapport aux fréquences observées sur la population TCT de référence - 2009....49
Figure 14 – Fréquences EPS-TCT pondérées par province du domicile du patient (poids TCT) par rapport aux fréquences observées sur la population TCT de référence - 2009....49
Figure 15 – Répartition des séjours dans l'échantillon EPS-TCT et dans la population TCT de référence par arrondissement du domicile du patient (poids TCT) – 2009 50
Figure 16 – Fréquences EPS-TCT pondérées par province de l'hôpital (poids AIM) par rapport aux fréquences observées sur la population TCT de référence - 2009 53
Figure 17 – Fréquences EPS-TCT pondérées par province de l'hôpital (poids TCT) par rapport aux fréquences observées sur la population TCT de référence – 2009....53
Figure 18 – Distribution empirique du montant total des remboursements : juxtaposition des courbes de l'échantillon EPS-TCT (poids TCT) et de la population TCT de référence – 2009....57
Figure 19 – Répartition des APR-DRG (APR-DRG représentant plus de 0,5% des séjours) dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - 2009 58

Figure 20 – Répartition des diagnostics principaux (diagnostics représentant plus de 0,5% des séjours) dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - 2009 59
Figure 21 – Nombre d'APR-DRG en fonction du nombre de séjours dans l'échantillon EPS-TCT 63
Figure 22 – Différence relative entre les répartitions par APR-DRG dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence 64
Figure 23 – Différence maximale entre les répartitions par sexe dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence 65
Figure 24 – Différence maximale entre les répartitions par région et par province du domicile du patient dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence 66
Figure 25 – Différence maximale entre les répartitions par statut social et par type d'hospitalisation dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence 66
Figure 26 – Nombre d'APR-DRG classés selon la différence maximale entre les répartitions (EPS-TCT versus population TCT de référence) pour toutes les variables confondues (<5% ou non), en fonction du nombre de séjours dans l'échantillon EPS-TCT 67
Figure 27 – Distribution du montant total de remboursement par niveau de sévérité (hospitalisations classiques) dans l'échantillon EPS-TCT (poids TCT) et dans la population de référence TCT - APR-DRG 560 Accouchement par voie vaginale 75
Figure 28 – Répartition des diagnostics principaux dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale - 2009 76
Figure 29 – Différence relative entre les répartitions par diagnostic principal dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale - 2009 77
Figure 30 – Répartition des diagnostics principaux dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 313 (Interventions des membres inférieurs et genoux, excepté le pied) - 2009 81
Figure 31 – Différence relative entre les répartitions par diagnostic principal dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 313 (Interventions des membres inférieurs et genoux, excepté le pied) - 2009 82
Figure 32 – Distribution du montant total de remboursement par niveau de sévérité dans l'échantillon EPS-TCT (poids TCT) et dans la population de référence TCT – Hospitalisations Classiques - APR-DRG 313 Interventions des membres inférieurs et genoux, excepté le pied 83
Figure 33 – Distribution du montant total de remboursement par niveau de sévérité dans l'échantillon EPS-TCT (poids TCT) et dans la population de référence TCT – Hospitalisations de jour - APR-DRG 313 Interventions des membres inférieurs et genoux, excepté le pied 84

LISTE DES TABLES

Table 1 – Nombres possibles de journées non tarifiables par type de week-end / jour férié 25
Table 2 – Nombre et type d'hospitalisations reconnues dans l'EPS pour les années 2008 et 2009. 31
Table 3 – Taux de couplage des tables de correspondance (HJA- ou SHA-EPS), par organisme assureur, année 2008 et 2009 (pourcentage de séjours) 32
Table 4 – Rejet de séjours EPS dû à l'absence d'identifiant patient. 33
Table 5 – Proportion des séjours de l'EPS-TCT par rapport aux séjours TCT [IC 95%] par catégorie d'âge 40
Table 6 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par catégorie d'âge 42
Table 7 – Répartition des séjours par sexe dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux 43
Table 8 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par sexe 43
Table 9 – Paramètres de distribution de l'âge dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence 45
Table 10 – Répartition des séjours par statut social dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux 46
Table 11 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par statut social 46
Table 12 – Répartition des séjours par région du domicile du patient dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux 47
Table 13 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par région du domicile du patient 48
Table 14 – Répartition des séjours par région de l'hôpital dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux 51
Table 15 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par région de l'hôpital 52
Table 16 – Répartition des séjours par type d'hospitalisation dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux 54
Table 17 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par type d'hospitalisation 55
Table 18 – Paramètres de distribution du montant total des remboursements (euros) dans l'échantillon EPS-TCT et dans la population TCT de référence 56
Table 19 – Table récapitulative des tests de représentativité par variable – au niveau global 60

Table 20 – Répartition des séjours par catégorie dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux d'âge - APR-DRG 560 Accouchement par voie vaginale.....69
Table 21 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par catégorie d'âge - APR-DRG 560 Accouchement par voie vaginale 70
Table 22 – Répartition des séjours par niveau de sévérité dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux - APR-DRG 560 Accouchement par voie vaginale.....71
Table 23 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par niveau de sévérité - APR-DRG 560 Accouchement par voie vaginale.....72
Table 24 – Paramètres de distribution du montant total des remboursements (euros) dans l'échantillon EPS-TCT et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale 73
Table 25 – Paramètres de distribution du montant total des remboursements (euros) par niveau de sévérité dans l'échantillon EPS-TCT et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale.....73
Table 26 – Tableau récapitulatif des tests de représentativité par variable - APR-DRG 560 Accouchement par voie vaginale.....78
Table 27 – Tableau récapitulatif des tests de représentativité par variable – APR-DRG 313 <w Interventions des membres inférieurs et genoux, excepté le pied 85

LISTE D'ABRÉVIATIONS

ABRÉVIATION	DÉFINITION
AIM	Agence InterMutualiste
AMI	Assurance maladie-invalidité
APR-DRG	All Patients Refined Diagnosis Related Group
CSSS	Comité sectoriel de la sécurité sociale et de la santé
CPAS	Centre Public d'Aide Sociale
EPS	Echantillon Permanent
HJA	Hospitalisations de Jour Anonymisées
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
INAMI	Institut national d'assurance maladie-invalidité
LOS	Length of stay = durée de séjour
MRPA-MRS	Maison de Repos pour Personnes Agées-Maison de Repos et de Soins
NISS	Numéro d'identification à la sécurité sociale
OA	Organisme assureur
RCM	Résumé Clinique Minimum
RHM	Résumé Hospitalier Minimum
SHA	Séjours Hospitaliers Anonymisés
SOI	Severity of illness = niveau de sévérité
SPF-SP	Service public fédéral Santé publique, Sécurité de la Chaîne Alimentaire et Environnement
TCT	Cellule Technique

RAPPORT SCIENTIFIQUE

1 INTRODUCTION

1.1 Contexte

La mission du KCE consiste notamment à soutenir la décision politique en matière de soins de santé grâce à l'exploitation des données de santé, qu'il s'agisse de données de littérature ou de données statistiques. Dans le passé, le KCE a déjà fait appel au cours de ses études à des bases de données nationales telles que les données médicales hospitalières hébergées au Service Public Fédéral Santé publique, Sécurité de la chaîne alimentaire et Environnement (SPF-SP), les données couplées des séjours hospitaliers hébergées par la Cellule Technique, les données du Registre du Cancer ou les données de remboursement dans le cadre de l'assurance maladie-invalidité rassemblées par l'Agence InterMutualiste (AIM).

Les données collectées au sein de bases de données administratives permettent une analyse rétrospective, longitudinale, à peu de frais.[1-3] Elles sont souvent collectées de façon régulière à grande échelle avec une structure de banque de données relativement stable, contrairement par exemple aux données prospectives ad hoc taillées sur mesure pour chaque étude clinique ou encore à l'extraction de données à partir du dossier médical. En Belgique, le Résumé Hospitalier Minimum (RHM) contient des informations sur chaque séjour passé dans un hôpital général en Belgique. Du côté des remboursements, les organismes assureurs disposent de tous les remboursements effectués par l'assurance maladie-invalidité au nom de leurs affiliés, en milieu hospitalier comme en milieu ambulatoire.

Si les grandes bases de données administratives exhaustives ont leurs avantages, elles présentent des défauts bien connus.[4,5] Elles sont conçues à des desseins particuliers, souvent éloignés des questions de recherche des études utilisant ces bases. Elles recourent à des systèmes de codage et de classification parfois imprécis ou aspécifiques. Enfin, les procédures pour y avoir accès sont généralement lourdes, en particulier lorsqu'il s'agit de coupler différentes bases. A cet égard, la Belgique possède sa propre législation sur la protection des données à caractère personnel, créée en 1992 et mise en conformité avec la législation européenne en 1998.[6]

Cette loi impose des restrictions plus sévères lorsqu'il s'agit de données à caractère personnel concernant la santé individuelle. Ainsi, toute transmission de données de l'AIM requiert une autorisation de principe du Comité sectoriel de la Sécurité Sociale et de la Santé. De même, l'analyse de ces données, leur stockage et leur dissémination dans le cadre d'une étude du KCE est strictement réglementée.[7] La procédure préalable à la mise à disposition des chercheurs des données à analyser pourrait paraître relativement lourde de prime abord. Toutefois, elle présente l'avantage d'inciter l'équipe de recherche à réfléchir fondamentalement à l'architecture des bases de données, à la méthodologie à suivre dans l'analyse et aux données strictement nécessaires à cette analyse. Conformément à la Directive Européenne, la législation belge sur la protection de la vie privée à l'égard des traitements de données à caractère personnel impose le respect des principes de finalité et de proportionnalité. En d'autres termes, les données demandées ne doivent pas excéder les nécessités de l'analyse et doivent être nécessaires pour répondre à la finalité de celle-ci. Enfin, les risques d'identification indirecte des patients et des prestataires de soins doivent être réduits à un minimum acceptable.

En 2005, le législateur a prévu la création par l'AIM de l'Echantillon Permanent (EPS), qui reprend les données de remboursement dans le cadre de l'assurance maladie-invalidité d'une fraction de la population assurée en Belgique, représentative au niveau sexe et âge de la population nationale.[8] L'objectif de cet échantillon était d'offrir un outil maniable, rapidement accessible à certaines institutions contrairement à l'ensemble des données de l'AIM qui ne sont accessibles qu'après une procédure plus longue. Cet outil permet des analyses scientifiques dans l'optique d'une politique de santé. Outre les données de remboursement des prestations ou de médicaments, l'EPS contient aussi certaines caractéristiques démographiques et socio-économiques des patients, mais aucune information clinique ou pathologique. En ce qui concerne les séjours hospitaliers, ces informations sont disponibles au sein du RHM. Ces données médicales hospitalières permettent notamment l'analyse de groupes de patients souffrant de certaines pathologies ou encore l'ajustement de résultats d'analyse aux risques individuels des patients.

C'est pourquoi des couplages entre les données de l'AIM et les données médicales hospitalières ont déjà eu lieu de manière ad hoc pour certaines études scientifiques menées, entre autres, par le KCE. Le rapport concernant le système des montants de référence en offre un exemple parmi d'autres.[9] Le processus administratif d'autorisation et de mise à disposition des données est donc parcouru à nouveau lors de chaque nouvelle étude.

Il existe un exemple de couplage similaire pérenne chez nos voisins. La France a également constitué un échantillon de 1/97ème de la population couverte par l'Assurance maladie Française, sondée sur base du numéro de sécurité sociale.[10,11] Cet Echantillon Généraliste des Bénéficiaires contient les informations anonymes sur les caractéristiques sociodémographiques et médicales des bénéficiaires ainsi que (certaines de) leurs prestations reçues en hospitalier comme en ambulatoire. Les informations portent sur la présence d'une maladie chronique (affection de longue durée), les diagnostics hospitaliers, les médicaments prescrits en ambulatoire, la biologie clinique, les dispositifs médicaux et la plupart de actes médicaux. Ainsi, en dehors des diagnostics posés en hospitaliers comme en Belgique, les affections de longue durée ou polypathologies du patient sont renseignées ainsi que la codification des diagnostics y afférent en ICD-10.

Mettant à profit leur expérience en la matière, l'AIM et le KCE ont explicité dans un premier volet de ce rapport le processus de couplage RHM-EPS et ses implications techniques (informatiques et organisationnelles) en vue des institutions ayant déjà directement accès à l'EPS. Mais si l'EPS est représentatif pour l'âge et le sexe au niveau de la population des assurés, la représentativité des variables cliniques, rendues disponibles grâce au couplage, reste inconnue au niveau des hospitalisations. Le deuxième volet du présent rapport vise donc à vérifier cette représentativité, c'est-à-dire à vérifier que les résultats observés lors d'analyses de l'échantillon couplé pourront être généralisés à l'ensemble des séjours hospitaliers belges. Ce test de représentativité a également été élargi au montant des remboursements pour les séjours hospitaliers par l'assurance maladie-invalidité.

1.2 Questions de recherche

1. Quelles sont les conditions techniques nécessaires pour coupler les données de l'EPS à celles du RHM ? Quelles sont les étapes nécessaires à un tel couplage ? Quelles sont les variables indispensables au couplage ?
2. Quelle est la représentativité des séjours RHM ainsi couplés à l'EPS par rapport à la totalité des séjours RHM, pour certaines variables du RHM et pour les remboursements ? A quel niveau la représentativité se vérifie-t-elle (global, groupe de pathologie : All patients refined Diagnosis related group (APR-DRG), APR-DRG x niveau de sévérité) ?

Les données des organismes assureurs s'étendant au milieu hospitalier mais également au milieu ambulatoire, la première question implique donc l'établissement préalable d'un algorithme d'extraction des prestations relatives aux différents séjours hospitaliers et de reconstitution des séjours au sein des enregistrements de l'EPS. Cette extraction et cette reconstitution doivent se faire avant le couplage par séjour.

2 DESCRIPTION DES BASES DE DONNÉES

Le présent chapitre introduit brièvement les bases de données nécessaires dans la cadre de cette étude. Pour plus d'informations, le lecteur peut également consulter le rapport KCE de 2006 qui dresse l'inventaire des bases de données de soins de santé disponibles en Belgique, et dont une mise à jour devrait paraître fin 2013.[5]

2.1 L'Échantillon Permanent

L'Agence InterMutualiste est une association représentant tous les organismes assureurs (OA – fédérations des mutualités), créée par la loi-programme (I) du 24 décembre 2002 (art.278).[12] C'est elle qui constitue les fichiers de l'EPS sur la base, d'une part, des données de consommation des soins des ayants droit de tous les OA et, d'autre part, des données administratives de ces ayants droit (données démographiques, socio-économiques et statut d'assuré).

L'EPS est composé de 1/40 des assurés sociaux affiliés ou inscrits auprès d'un organisme assureur, tirés de manière stratifiée sur base de la date de naissance et du sexe. Un tirage supplémentaire de 1/39 des assurés âgés de 65 ans ou plus assure un sur-échantillonnage permettant de disposer d'une personne sur 20 dans cette tranche. Toutes les données d'échantillonnage sont mises à jour le 31 décembre de chaque année calendrier. Cette mise à jour comporte non seulement la mise à jour des données des ayants droit présents dans l'échantillon mais également l'entrée de nouveaux ayants droits (nouveau-nés, immigrés et nouveaux ayants droits de plus de 65 ans) ainsi que la sortie d'autres ayants droit (émigrés et décédés). L'individu et ses prestations détaillées constituent le niveau d'observation de l'échantillon (au sens large, y compris les traitements administratifs tels que les corrections comptables).

Conformément à la loi-programme du 27 décembre 2005, l'AIM donne accès permanent via une connexion sécurisée à l'échantillon ainsi constitué au KCE, à l'INAMI, au SPF Santé publique, au SPF Sécurité sociale ainsi qu'au Bureau fédéral du Plan.[8] Les institutions ayant accès à cet échantillon ne peuvent utiliser ces données que dans le cadre de leurs missions légales de gestion et de recherche ou des tâches de contrôle et d'évaluation qui leur sont conférées par la Loi.

2.2 Les données de la Cellule Technique

2.2.1 Données médicales hospitalières

Depuis le début des années nonante, l'enregistrement des données médicales hospitalières est obligatoire pour tous les hôpitaux généraux non psychiatriques en Belgique. Cet enregistrement continu est envoyé tous les semestres au SPF Santé Publique et porte aussi bien sur l'hospitalisation de minimum une nuit (dite hospitalisation classique) que sur l'hospitalisation de jour. Les données relatives au séjour hospitalier comprennent les données du patient, telles que la date de naissance, le sexe, le code postal du domicile. D'autres informations comme la durée de séjour, les passages dans les différents services, le séjour en soins intensifs sont enregistrées ainsi que la codification ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) des diagnostics du patient et des procédures diagnostiques ou thérapeutiques qu'il a subies. Ces diagnostics et procédures sont collectés pour chaque service de l'hôpital qui doit coder le diagnostic principal (la cause de l'admission dans le service) ainsi que les diagnostics secondaires. Après réception des données codées au niveau du patient de manière à empêcher toute identification directe de celui-ci, le SPF-SP agrège les enregistrements et attribue le diagnostic principal de tout le séjour pour que le logiciel APR-DRG Grouper[a] puisse, sur base des données cliniques, calculer l'APR-DRG et le niveau de sévérité du séjour. Jusque fin 2007, la base de données relationnelle ainsi constituée formait le RCM (Résumé Clinique Minimum). A partir des données 2008, plusieurs bases de données administratives, auparavant collectées séparément, ont été intégrées pour former le **RHM** (Résumé Hospitalier Minimum), qui englobe les données administratives sur les hôpitaux, les données médicales (RHM-M), les données infirmières, les données relatives au personnel et les données de la fonction « service mobile d'urgence ». Pour le présent projet qui porte sur les années d'enregistrement 2008-2009, il s'agit donc d'utiliser le RHM-M.

[a] La version APR-DRG 15 s'applique aux données 2008 et 2009

2.2.2 Données financières hospitalières

L'INAMI dispose d'une base de données appelée **SHA/HJA** (Séjours Hospitaliers Anonymisés-Hospitalisations de Jour Anonymisées). Comme leur nom l'indique, ces données portent sur les hospitalisations classiques et l'hospitalisation de jour. Cette base de données regroupe les données de remboursement liées à l'hospitalisation de leurs affiliés que les organismes assureurs transmettent annuellement à l'Institut. Parmi ces remboursements se trouvent notamment les prestations des différents prestataires de soins au cours du séjour hospitalier (actes médicaux et chirurgicaux, examens diagnostics, tests de biologie clinique…), le matériel implantable et les spécialités pharmaceutiques, et autres (radio-isotopes, produits sanguins...).

2.2.3 Données hospitalières couplées

Chaque année depuis 1997, les données médicales hospitalières et les données financières sont transmises à la **Cellule Technique** (TCT) par le SPF Santé Publique (données médicales) et par l'INAMI (données financières). Les données financières auront préalablement été structurées par séjour par les organismes assureurs. La TCT couple les deux bases de données grâce aux deux tableaux de correspondance que cette Cellule reçoit d'une part des hôpitaux et des organismes assureurs d'autre part. Ces tableaux comprennent un numéro de patient unique résultant d'une double procédure de **hashing**, qui transforme le numéro d'identification du patient à la sécurité sociale belge (numéro NISS)[b] en une clef unique codée. Un premier hashing est exécuté séparément au sein des OA et au sein des hôpitaux. Un second hashing est exécuté séparément par le conseiller en sécurité du SPF-SP. L'algorithme est irréversible de sorte que l'anonymisation du patient est garantie, bien qu'en termes légaux les données soient qualifiées de données à caractère personnel codées (et non pas données anonymes). Les délais légaux de transmission des tables de correspondance et des données entre les différents acteurs ainsi que le processus de validation permettent après deux ans d'aboutir à un couplage consolidé des deux sources de données.[13-18]

[b] Ou le numéro d'affilié auprès de l'organisme assureur pour les quelques cas sans numéro NISS.

Le pourcentage de séjours couplés a constamment augmenté et est supérieur à 95% les dernières années. Enfin, les hospitalisations classiques sont couplées depuis 1997, tandis que les hospitalisations de jour ne le sont que depuis 2006. Ce décalage n'a pas d'implication pour cette étude-ci puisque les données analysées sont celles de 2008-2009.

La **Cellule Technique (TCT)**, légalement créée en 1996 au sein du SPF-SP et de l'INAMI, a pour mission de collecter, relier, valider et anonymiser les données relatives aux hôpitaux. En outre, elle rend les données disponibles au SPF-SP et à l'INAMI pour l'exécution de leurs missions.[19]

Le **hashing**, en informatique, transforme une donnée de n'importe quelle longueur en une série unique de longueur fixe et généralement plus courte. Dans le domaine spécifique du couplage de données personnelles, cette fonction est utilisée pour transformer irréversiblement le code NISS du patient en une série unique de caractères dépourvue de signification particulière. Ainsi, l'identité physique est masquée mais les données restent personnalisées, c'est-à-dire qu'elles ont toujours trait à une seule et même personne (patient).[20,21]

En définitive, ce couplage permet d'établir un lien entre une pathologie donnée et les remboursements y relatifs par l'assurance maladie-invalidité. La base de données ainsi obtenue offre l'avantage d'intégrer les données de tous les hôpitaux aigus non psychiatriques du pays ainsi que celles de tous les organismes assureurs.

Il existe un exemple similaire d'un tel couplage chez nos voisins. La France a également constitué un échantillon de 1/97ème de la population couverte par l'Assurance maladie Française, sondée sur base du numéro de sécurité sociale.[10,11] Cet Echantillon Généraliste des Bénéficiaires contient les informations anonymes sur les caractéristiques sociodémographiques et médicales des bénéficiaires ainsi que (certaines de) leurs prestations reçues en hospitalier comme en ambulatoire. Les informations portent sur la présence d'une maladie chronique (affection de longue durée), les diagnostics hospitaliers, les médicaments prescrits en ambulatoire, la biologie clinique, les dispositifs médicaux et la plupart de actes médicaux.

Ainsi, en dehors des diagnostics posés en hospitaliers comme en Belgique, les affections de longue durée ou polypathologies du patient sont renseignées ainsi que la codification des diagnostics y afférent en ICD-10.

Les données analysées dans le présent rapport ont trait aux années de prestation **2008-2009**.

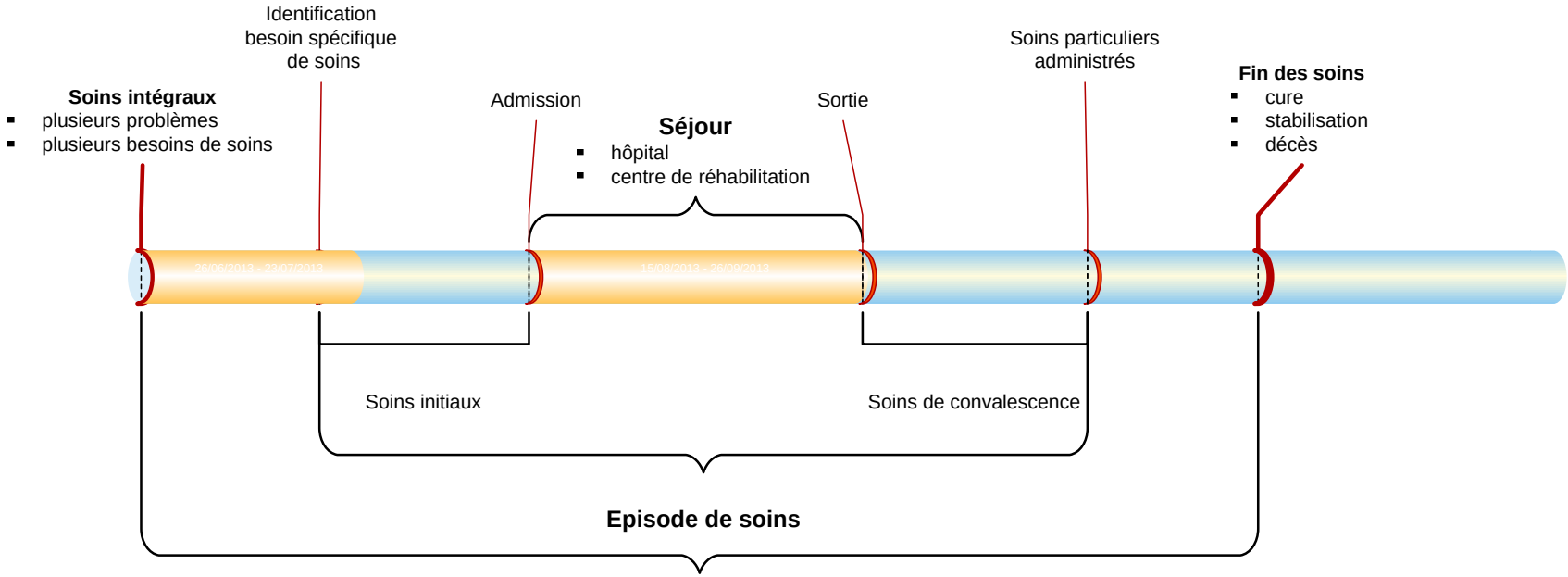
3 CADRE CONCEPTUEL ET NOTION DE SÉJOUR HOSPITALIER

La présente section tâche de définir ce que la notion de séjour hospitalier recouvre, de l'inclure dans la notion plus large d'épisode de soins ou plus généralement encore du trajet de soins suivi par le patient. Des questions légitimes se posent telles que : quand débute officiellement un séjour et quand finit-il ou encore quand peut-on considérer qu'un besoin de soins a été satisfait ? Ces notions sont importantes pour définir l'unité sur laquelle porte l'étude.

3.1 Définitions relatives aux ensembles de soins : séjour hospitalier, épisode de soins et trajet de soins

La notion de *séjour hospitalier* peut être définie sur la base d'une logique marchande, dans laquelle un ensemble de services et de biens (prestations de soins) est livré par un fournisseur déterminé qui doit en retour être rémunéré par le débiteur de la facture. Le plus souvent, la date de début et la date de fin de l'ensemble peuvent être clairement délimitées : il s'agit de la date d'admission et la date de sortie de l'hôpital respectivement. La facture reprend la liste précise de tous les services et de tous les biens fournis ou dont la fourniture a été « commandée » pendant le séjour (fournitures différées) pour lesquels un remboursement est réclamé. L'enregistrement des données du RHM et des données SHA/HDA et par extension leur couplage ultérieur par la cellule technique (TCT) s'inscrit dans le cadre de cette définition.

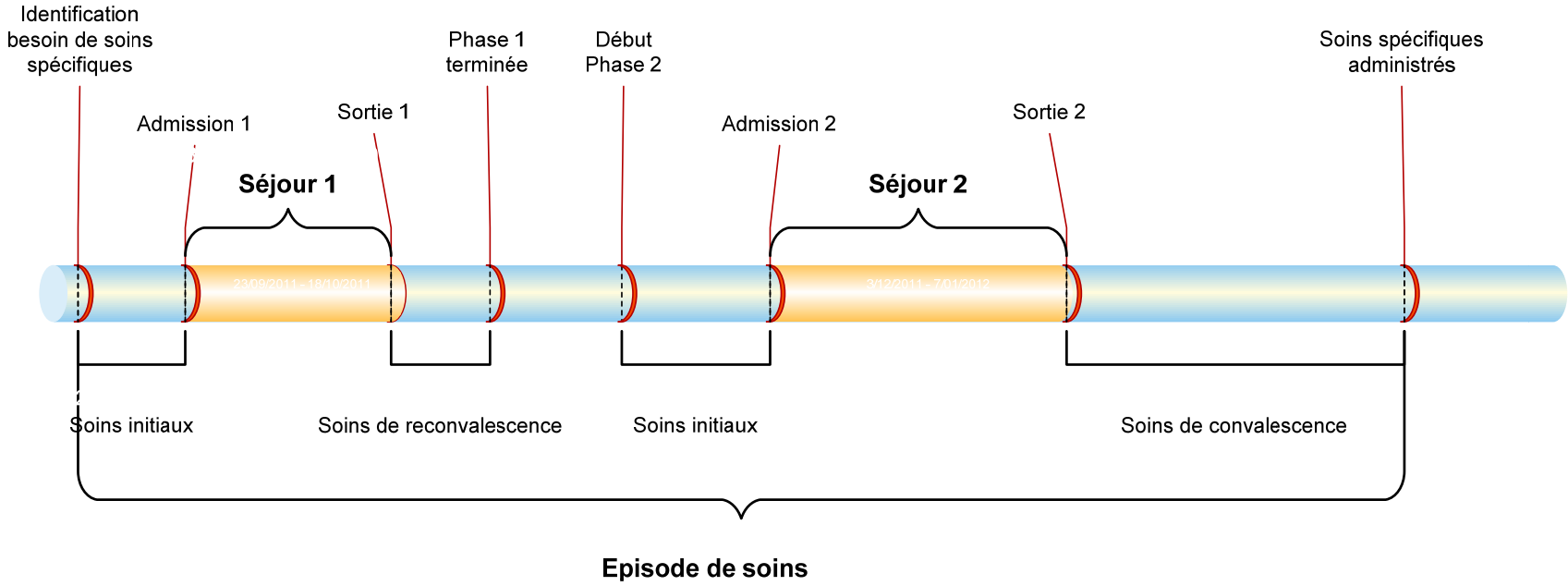
Figure 1 – Trajet de soins, épisode de soins et séjour hospitalier



Trajectoire de soins

Dans la majorité des cas, le caractère d'entité isolée du séjour hospitalier est beaucoup moins pertinent d'un point de vue clinique (Figure 1). En effet, chaque hospitalisation est généralement précédée d'une phase de soins préalables (diagnostic, prétraitements..., en principe fournis par les soins de première ligne) et d'une phase de soins de suivi (soins le plus souvent aussi fournis par les soins de première ligne). Cet ensemble s'inscrit dans le concept d'*épisode de soins* : un problème de santé particulier peut être circonscrit dans le temps par une date de début – la date d'officialisation du besoin de soins – et une date de fin – la date à laquelle il a été répondu à ce besoin de soins. Cet épisode peut comprendre une ou plusieurs hospitalisation(s), selon le problème de santé dont il s'agit C'est ainsi, par exemple, que l'épisode de soins cholécystolithiase comprend une phase diagnostique et une phase chirurgicale ainsi que – le cas échéant – une ou plusieurs phases de soins de suivi comprenant des (ré)hospitalisations, aussi bien en hospitalisation de jour qu'en hospitalisation classique (Figure 2).

Figure 2 – Épisode de soins à plusieurs admissions



Dans la perspective de la politique en matière de soins de santé, une perspective encore plus large est souvent utilisée : celle du *trajet de soins* (Figure 1). Pour les programmes de soins qui requièrent généralement une approche multidisciplinaire, le trajet de soins regroupe l'ensemble de tous les épisodes de soins relatifs à une même problématique de soins. Un même trajet de soins comprend donc plusieurs épisodes de soins, chacun ayant trait à différents éléments du programme de soins complet. Le début d'un tel trajet de soins coïncide avec le début de premier épisode de soins ayant donné lieu à l'inscription du patient dans le programme de soins. Théoriquement, la fin d'un programme de soins doit se situer au moment où les objectifs ont été réalisés ; mais dans la pratique, la détermination précise de ce moment est souvent très difficile. A l'exception du décès, la problématique revient à se poser la question : « quand un patient est-il guéri ? ».

En outre, les trajets de soins individuels d'un même patient peuvent se croiser ou même partiellement se chevaucher chez les patients présentant plusieurs comorbidités. Ce cas de figure illustre la problématique de l'imputation des dépenses en soins à un trajet de soins en particulier dans le cadre d'études en économie de la santé.

Limites posées dans le cadre de l'étude

Dans la présente étude, nous nous sommes nécessairement limités au concept restreint de « séjour hospitalier ». Cette restriction résulte de la nécessité de faire coïncider les données de l'EPS à la définition restreinte du séjour hospitalier dans les données de la TCT. Ce séjour peut mêler des soins et des ressources consacrées à différents problèmes de santé et exclure des soins dispensés avant ou après lui. Ainsi, par exemple, la représentativité que nous allons tester ne dira rien de la représentativité du coût d'une grossesse menée à terme, mais uniquement du coût engendré par le séjour d'accouchement.

3.2 Inventorisation des coûts liés à un ensemble de soins

Après avoir défini l'ensemble de soins en fonction du thème spécifique de l'étude, se pose la question de l'inventorisation correcte des coûts liés à cet ensemble de soins.

Les coûts liés à l'ensemble de soins peuvent être définis et inventoriés de manières très diverses. L'objectif central de cette étude est d'étudier la représentativité des séjours hospitaliers dans l'EPS sur la base d'une comparaison avec les données nationales fournies par la TCT. Le coût moyen total par séjour constitue une des variables dont la représentativité va être vérifiée. La TCT a déjà élaboré un algorithme pour le compte de la Banque Nationale de Belgique pour le calcul du coût total par séjour hospitalier pour tous les séjours classiques de 2008 et 2009. Pour que le choix de la définition et la délimitation des coûts des séjours dans l'EPS n'influencent pas les résultats de l'étude au niveau de la représentativité, nous avons aussi choisi d'adopter cette méthodologie pour calculer le coût total moyen par séjour. Le point 4.1.2 reprend une description détaillée de l'algorithme mis au point par la TCT.

4 RECONSTITUTION DES SÉJOURS ET COUPLAGE

4.1 Méthodologie

La présente section ne décrit pas la procédure de couplage primaire des deux bases de données au niveau du patient (voir demande d'autorisation au Comité sectoriel en annexe 1) mais bien le couplage secondaire au niveau des séjours.

Il convient néanmoins de préciser que ce couplage primaire revient essentiellement à la mise en correspondance des pseudonymes des patients des deux bases de données de départ. En effet, conformément au principe restrictif imposé par la loi relative à la protection des données à caractère personnel (loi vie privée[6]), ces deux bases de données ont leur propre pseudonyme, non interchangeable et créé par une procédure de hachage cryptographique (« hashing »), chacune avec ces propres clés (principe de la spécificité de domaine[5]). La mise en correspondance nécessite le retour aux identifiants primaires, plus exactement le numéro d'identification à la sécurité sociale (NISS) du patient connu des OA, et se fait par l'intermédiaire d'un tiers de confiance, la plateforme eHealth. Elle collecte les tables de conversion primaires auprès des OA et procède à un recodage aboutissant à un pseudonyme commun et spécifique au projet, ci-après nommé pseudonyme de couplage.

La notion « **pseudonymisation** » signifie une désidentification des données de sorte que les informations appartenant à la même personne gardent leur appartenance à la même personne (patient) dans les données après pseudonymisation. Bien que l'identité physique ait été dissimulée, les données restent donc personnalisées.[22 23]

Ainsi, il est important de souligner qu'un couplage des données administratives de l'EPS aux données cliniques du RHM revient en fait à coupler l'EPS aux séjours hospitaliers des données RHM-SHA/HJA, enregistrements indépendants et annuellement couplés par la Cellule Technique. En effet, tout couplage avec ces données nécessite un recours inéluctable au pseudonyme patient issu des tables de correspondance séparément livrées à la TCT et donc absent dans les enregistrements RHM et SHA/HJA primaires. En outre, le couplage n'est possible que dans le cas de séjours pour lesquels des données sont enregistrées aussi bien dans le RHM que dans les SHA/HJA. Il en est de même pour le couplage à l'EPS, puisque les données SHA/HJA proviennent de la même source que l'EPS, à savoir des données administratives des OA, issues des remboursements effectués dans le cadre de l'assurance maladie obligatoire (AMI). Les autres séjours RHM, issus de prises en charge par les Centres Publics d'Aide Sociale (CPAS), les assurances accidents du travail ou les assurances privées, ne sont donc pas concernés par le couplage. Ces séjours seront donc absents de l'Echantillon couplé, risquant de mener ainsi à des sous-estimations du nombre total d'hospitalisations ayant lieu en Belgique. Ils représentent environ 3 % des séjours RHM.

Dès lors, lorsque nous parlerons de données RHM dans le contexte du projet présent, il faudra lire données couplées RHM-SHA/HJA ou données des séjours hospitaliers de la TCT (en bref : « données TCT » versus « données EPS »).

Dans ce rapport, les données RHM-SHA/HJA sont appelées **données TCT** (de la Cellule Technique). Le couplage final avec l'Echantillon Permanent donnant lieu aux **données EPS-TCT**.

Ces séjours hospitaliers sont structurés dans une base de données dite relationnelle : le modèle de données comprend plusieurs tables, interconnectées par un nombre de variables bien spécifiques qui composent ensemble une clé unique pour chaque séjour individuel. Autrement dit : l'unité de base des données TCT est le séjour hospitalier et non le patient[c]. En effet, un même patient peut avoir eu plusieurs séjours dans une même année d'enregistrement.

Les données de l'EPS, par contre, ne présentent pas une telle structure. Elles sont le résultat d'une simple succession d'enregistrements, chacun

[c] Bien que logiquement un séjour égale un patient (sauf nouveau né sans problèmes médicaux graves, qui reste dans la chambre de la mère).

d'eux représentant un seul acte de comptabilisation[d] (inscription ou suppression) au sein des 7 OA nationaux. Centralisées au sein de l'AIM, elles composent les données nationales de facturation de l'AIM. Chacun de ces enregistrements comprend – entre autres – l'identification du bénéficiaire (codée).

En conclusion, avant de pouvoir procéder au couplage secondaire, c'est-à-dire au niveau des séjours, il faudra d'abord passer par un travail de reconstruction des séjours hospitaliers dans les données EPS.

4.1.1 Reconstitution des séjours hospitaliers dans les données EPS

4.1.1.1 Définition des séjours hospitaliers

La méthodologie pour la reconstitution des séjours hospitaliers à partir des données de facturation de l'AMI dépend fortement des questions de recherche du projet.

Pour le présent projet, il faut entendre par la notion « séjour » la définition telle qu'elle est reprise dans les séjours hospitaliers de la Cellule Technique (TCT). Pour la partie RHM de ces données, chaque séjour dans un hôpital général non psychiatrique fait l'objet d'un enregistrement le semestre au cours duquel a eu lieu la sortie du séjour et ceci qu'elle que soit de la partie payante (AMI, CPAS, assurance privée comme pour les accidents de travail...). Du côté des données EPS, par contre, seules les données des séjours à charge de l'AMI sont enregistrées.

D'autre part, les séjours résidentiels comme dans les maisons de repos pour personnes âgées ou les maisons de repos et de soins (MRPA-MRS), ainsi que les séjours en hôpital psychiatrique, ne sont pas repris dans les données TCT.

Conformément à l'article 4 de l'arrêté royal du 27 avril 2007 déterminant les règles suivant lesquelles certaines données hospitalières doivent être communiquées au Ministre qui a la Santé publique dans ses attributions,[24] les directives RHM (applicables aux années 2008-2009) font la distinction entre les différents types de séjour :

1. les hospitalisations classiques donnant lieu à la facturation d'un prix de journée pour au moins une journée complète (ou incomplète en cas de décès) ;
2. les hospitalisations de jour pour lesquelles le patient sort le même jour que son admission et pour lesquelles sont dus un maxi-forfait ou un forfait d'hospitalisation de jour chirurgicale ou des groupes 1 à 7[e], tel que visé à l'art.4 §4 (les maxi forfaits) et §5 (les forfaits pour la fonction d'hôpital de jour[f]) de la convention nationale de l'année concernée conclue entre les établissements hospitaliers et les organismes assureurs.[25]

Les séjours de longue durée (>6 mois) constituent un cas particulier d'hospitalisation classique. Dans ces cas-là, les données du séjour sont enregistrées dans le RHM à la fin de chaque semestre. Le premier enregistrement a lieu à la fin du semestre de l'admission. Tant que le patient n'est pas sorti de l'hôpital, les semestres suivants donnent lieu à des enregistrements RHM dits intermédiaires. Enfin, le dernier enregistrement reprend toutes les données, de l'admission jusqu'à la sortie. Ce n'est que ce dernier enregistrement (complet) qui est apte au couplage aux données SHA. Les données RHM 2008 comprennent des séjours avec admission en 2007 qu'il faudra mettre en correspondance avec des données EPS relatives à la période précédant le premier janvier 2008. La transmission de ces données a été explicitement prévue dans la demande d'autorisation auprès du Comité sectoriel de la sécurité sociale et de la santé (CSSS), section santé et approuvée par ce dernier.

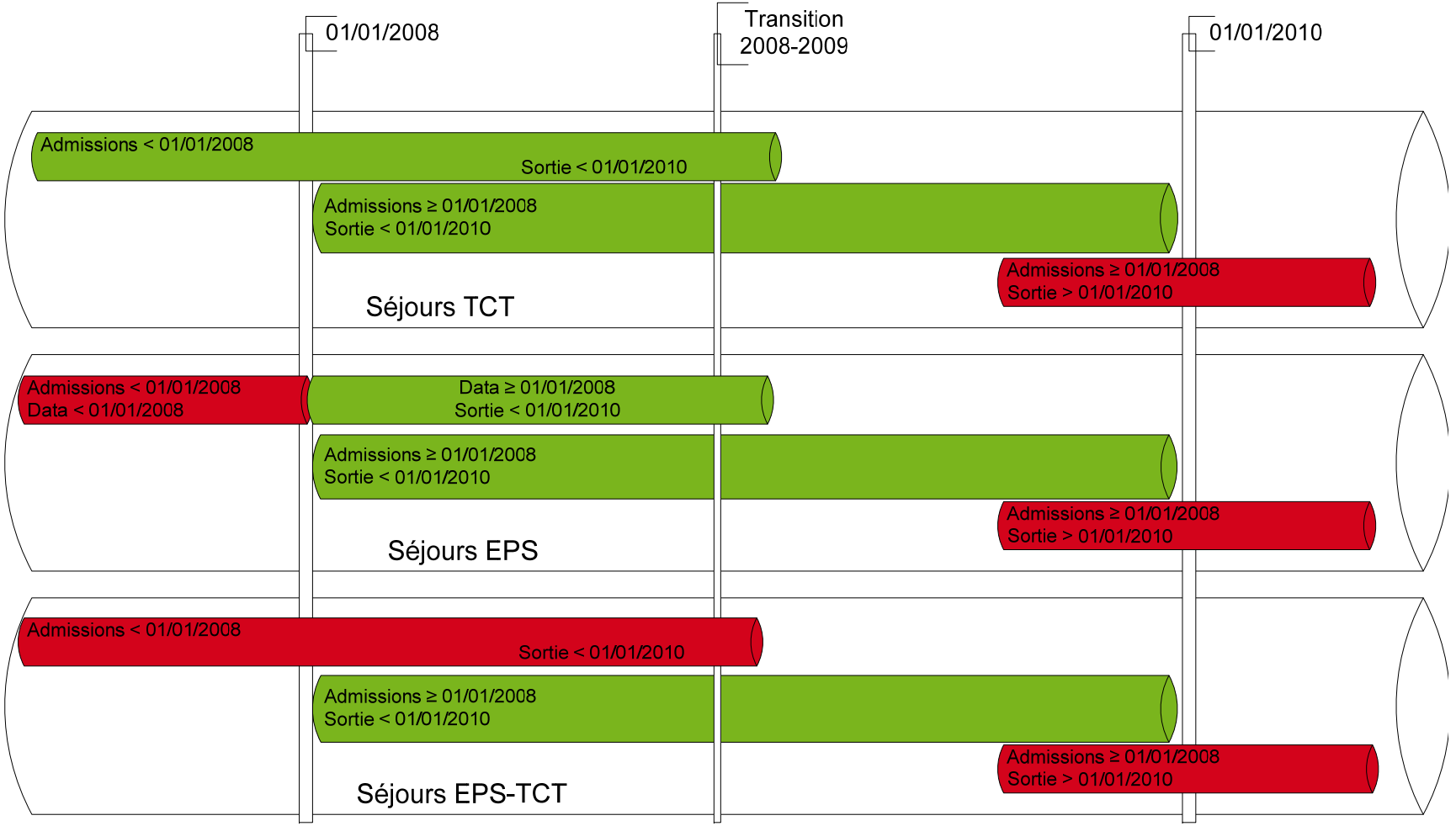
La Figure 3 illustre les séjours inclus ou exclus de la sélection dans chacune des deux bases de données et les séjours communs issus du couplage EPS-TCT.

[d] A souligner que l'acte de comptabilisation peut comprendre un enregistrement 'en bloc' d'un seul code, mais pour plusieurs jours consécutifs.

[e] Ainsi que les forfaits douleur chronique, mini-forfait et forfait salle de plâtre.

[f] Forfaits A, B, C et D avant et forfaits groupes 1 à 7 ainsi que les groupes douleur chronique 1 à 3 douleur chronique dès le 01/07/2007.

Figure 3 – Séjours sélectionnés (vert) ou non (rouge) dans chacune des bases de données et après couplage EPS-TCT



4.1.1.2 *Données de base nécessaires à la reconstitution des séjours*

La délimitation univoque d'un séjour hospitalier nécessite de disposer de l'identification du patient et de l'hôpital, ainsi que des dates complètes de début et de fin du séjour. Du côté des données EPS ces dates ne sont pas enregistrées pour les hospitalisations de jour et sont parfois manquantes pour les séjours en hospitalisation classique. Bref les dates d'admission et de sortie ne sont pas assez fiables. Par contre, les dates des prestations tarifiées sous leur code de nomenclature INAMI sont intégrales. Ceci nous permet de déduire les dates complètes de début et de fin de séjour par l'intermédiaire des facturations des forfaits de l'hôpital de jour, chirurgical ou non chirurgical et de l'hospitalisation classique. Ces derniers forfaits, ainsi que ceux de l'hôpital de jour chirurgical ne couvrent que la partie variable du financement des journées d'entretien.

Le **budget des moyens financiers (BMF)** se compose :

• d'une part fixe payée directement aux hôpitaux sur une base mensuelle, sans facturation par patient ;

• d'une part variable payée au prorata de deux **forfaits par patient (par admission** et **par journée d'entretien**).

Ces paiements couvrent les investissements médicaux et non médicaux, les installations et l'équipement, l'administration, le personnel paramédical et autre personnel non médical, une partie des frais d'admission et de séjour...

Le service public fédéral Santé publique, Sécurité de la chaine alimentaire et Environnement est responsable du calcul des paiements individuels aux hôpitaux. Les deux parties – fixe et variable – sont, pour une grande part, payées par l'Institut national d'assurance maladie invalidité (INAMI).

Les forfaits par admission

Un forfait par admission est dû pour chaque première journée d'hospitalisation comprenant au moins une nuit, c'est-à-dire une admission qui commence avant minuit et se termine après huit heures le lendemain.

Un montant par admission est également dû en cas d'admission en hospitalisation chirurgicale de jour lors de laquelle est effectuée une prestation figurant dans la liste A de l'arrêté royal du 25 avril 2002 relatif à la fixation et la liquidation du budget des moyens financiers des hôpitaux (annexe 3, 6) ou une prestation visée à l'article 101bis de ce même arrêté.[26]

Ce montant par admission n'est pas dû en cas d'admission dans un service G isolé (hôpitaux dits gériatriques), dans des unités pour le traitement des grands brûlés ou dans un service Sp. Ce type de service rassemble les services spécialisés pour le traitement et la réadaptation (soins palliatifs, soins des affections particulières cardio-pulmonaires, locomotrices, neurologiques ou psycho-gériatriques ou soins des poly-pathologies chroniques).

Les forfaits par journée d'entretien

Le forfait par journée d'entretien est dû pour chaque journée d'hospitalisation obéissant aux mêmes règles précitées pour les forfaits par admission (paragraphe précédent). Au cas où, le même jour et dans le même hôpital, une admission en hospitalisation chirurgicale de jour est suivie, immédiatement ou non, d'une hospitalisation, le montant par admission et le montant par journée d'entretien ne sont dus qu'une seule fois pour cette journée.

Le transfert du bénéficiaire vers un autre service du même hôpital, situé ou non sur le même campus, est considéré comme une continuation de l'hospitalisation et ne donne droit à aucun montant par admission sauf s'il concerne un patient qui, initialement admis à l'hôpital dans un service Sp par exemple, fait ensuite l'objet d'un transfert interne vers un service qui n'est pas un service Sp. Toutefois, en cas de deuxième transfert (ou suivant) entre les services précités, aucun nouveau montant par admission ne peut être porté en compte.

Comptabilisation des forfaits pour les hospitalisations classiques

Le forfait par admission est toujours comptabilisé à la date d'admission du bénéficiaire (indépendamment de l'heure de l'entrée, avant ou après 12 heures).

Les jours d'hospitalisation du bénéficiaire sont comptabilisés comme suit :

1. en cas d'admission avant 12 heures le jour de son entrée conjointement avec un départ après 14 heures le jour de sa sortie : comptabilisation du jour d'entrée et du jour de la sortie ;
2. en cas d'admission avant 12 heures le jour de son entrée et départ avant 14 heures le jour de sa sortie : comptabilisation du jour d'entrée ;
3. en cas d'admission après 12 heures le jour de son entrée, quelle que soit l'heure de départ le jour de sa sortie : comptabilisation du jour de sortie.

Calcul de la durée de séjour

Il en résulte que, si l'on utilise les forfaits par admission pour définir le début de séjour, le calcul de la durée de séjour pourrait être biaisé : les séjours aigus à forfait par admission avec une admission après 12 heures auraient comme début la date avant la première date de comptabilisation des forfaits pour journée d'entretien.

Par conséquent, on n'utilisera que les comptabilisations des forfaits pour journée d'entretien pour déterminer les dates de début et de fin de séjour. Ces forfaits sont habituellement comptabilisés « en bloc » avec une première date et une dernière date. Pour les longs séjours, il arrive que plusieurs de ces blocs soient comptabilisés.

La date de début de séjour égale la toute première date des forfaits par journée d'entretien pour un même séjour dans les données EPS (il en est de même pour les données TCT).

En théorie, la dernière date d'une série ininterrompue de facturation de ces forfaits pour un même patient dans un même hôpital nous fournit la date de fin de séjour (dernière date de début de prestation + nombre de journées facturées à cette date – 1). Il existe pourtant des exceptions à cette règle : pour des séjours généralement de durée prolongée, il arrive qu'un patient puisse passer la fin de semaine hors de l'hôpital. En cas de jour férié cette permission peut s'élargir à plusieurs jours – week-end prolongé – voire à d'autre(s) jour(s) de la semaine. L'enregistrement comptable du séjour se poursuit dès le retour du patient à l'hôpital. La délimitation des séjours hospitaliers devra donc tenir compte des dates des week-ends et des jours fériés de la période analysée (nous avons conventionnellement considéré les dates belges du 11 juillet et du 27 septembre comme jours fériés).

Enfin, pour les hospitalisations de jour, les données EPS ne fournissent pas systématiquement la date d'admission (ni de sortie, égale à la date d'admission). Le recours aux dates de facturation des forfaits de l'hôpital de jour, chirurgical ou non chirurgical, est donc indispensable.

4.1.1.3 Variables nécessaires à la reconstitution des séjours

Les variables nécessaires à la délimitation des séjours hospitaliers se résument aux éléments suivants (les noms des variables concernées dans les données soins de santé de l'EPS sont repris entre parenthèses) :

1. l'identification du bénéficiaire codé (SS00010) ;
2. l'identification de l'institution de soins (SS00075, SS00085, SS00105) ;
3. les codes nomenclature des prestations concernées (SS00020 - voir la section 4.1.1.4 suivante) ;
4. les dates de début et de fin de ces prestations (SS00015, SS00125) ;
5. les dates d'admission et de sortie du bénéficiaire (SS00110, SS00115) ;
6. le type d'hospitalisation (déduit à partir du code nomenclature) ;
7. les informations sur les dates des week-ends et jours fériés de la période analysée.

4.1.1.4 Sélection des codes nomenclature nécessaires

Les prestations nécessaires à la reconstitution des séjours sont les prestations facturées pour les montants par jour de l'hôpital chirurgical de jour et de la partie variable de l'hospitalisation classique, ainsi que les forfaits pour hospitalisation (non chirurgicale) de jour (voir annexe 4 – liste à jour pour les années 2008-2009). Dorénavant l'ensemble de ces codes sera nommé « codes forfaits ».

Les mini-forfaits (dus lors de soins urgents ou pour administration d'un médicament ou de sang ou produits sanguins labiles par perfusion intraveineuse à un patient non hospitalisé) et le forfait dû lors l'utilisation de la salle de plâtre[25] n'ont pas été considérés comme donnant lieu à un séjour proprement dit dans la présente étude.

4.1.1.5 Préparation des données

Pour les hospitalisations classiques les définitions des dates de début et de fin du séjour ont déjà été reprises à la section 0 (calcul de la durée de séjour).

Il faut également procéder à l'identification de l'hôpital de séjour. Par souci de préserver l'intégrité des données de départ, la variable-clé SS0075bis sera créée, qui correspond au champ <stay_place_id> des données TCT (sachant que la TCT aura déjà géré les fusions d'hôpitaux dans ses données).

Pour l'identification de l'hôpital de séjour, le champ SS00075bis est rempli avec les 8 premières positions du numéro d'institution (SS00075) si celui-ci correspond à un hôpital général (trois premières positions égalent 710). Dans le cas contraire, le champ SS00075bis égale les 8 premières positions du lieu de prestation (SS00085) si celui-ci est bien un hôpital général. Si aucun des deux champs précédents ne renseigne un hôpital général mais que le numéro de tiers (SS00105) en est un, ce sont les 8 premières positions de ce dernier numéro qui sont reprises dans la variable SS00075bis. Enfin, si aucun des trois champs en question ne répond à la condition, le champ SS00075bis reste manquant (« missing »).

Concernant les numéros d'agrément INAMI des institutions de soins il importe de préciser que ces numéros se composent de différentes zones :[27]

1. les trois premières positions représentent le type d'hôpital : 710 pour hôpital général (aigu ou chronique), 720 pour hôpital (purement) psychiatrique (aucun enregistrement dans les données TCT, donc non couplable), 715 pour services spécialisés et gérés en commun par deux ou plusieurs hôpitaux (p.ex. service de cathétérisme cardiaque), 716 pour associations d'hôpitaux dans le cadre de bassins de soins et 719 pour l'hôpital militaire de Neder-Over-Heembeek[g] ;
2. un numéro d'ordre unique par hôpital : positions de 4 à 6 et ayant une valeur entre 001 et 999 ;
3. un check-digit (CD) en deux positions obtenues par la formule : CD = 97- reste entier de la division des 6 premiers chiffres du numéro d'agrément par 97 (modulo 97) ;
4. parfois un code de compétence : du 9ième au 11ième chiffre.

Il en résulte que la préparation des données devra inclure pour le champ SS00075 :

1. un contrôle des check-digits des numéros d'agrément des institutions de soins et ;
2. une troncation de ces numéros à leurs 6 premières positions.

Il faut également procéder aux régularisations en agrégeant les enregistrements par bénéficiaire / hôpital / code forfait et date de prestation et en traitant (acceptant ou non) les valeurs obtenues (absence ou nombres négatifs de jours facturés, absence de forfait d'admission ou de journée forfaitaires…).

4.1.1.6 Tri des données

Après insertion dans les données du numéro de l'hôpital de séjour normalisé (SS00075bis), accompagné d'un « flag » indiquant un code pour hospitalisation classique ou de jour, toutes les données sont triées.

Tout dépend alors de l'objectif : si l'on vise à cerner toutes les prestations d'un même séjour, en dehors de la chronologie des séjours, on trie par 1) bénéficiaire, 2) hôpital du séjour, 3) type de prestation (hospitalisations de jour et puis classiques) et 4) dates des prestations avec un lieu de prestation codé pour un hôpital ou service affilié.

---

[g] Les activités de centre pour grands brûlés de l'hôpital militaire de Neder-Over-Heembeek ne font pas l'objet d'un enregistrement RHM.

Si, par contre, on veut établir la chronologie des séjours, l'ordre de tri doit être : 1) bénéficiaire, 2) date de sortie, 3) hôpital. Dans ce cas la date d'admission est superflue étant donné l'impossibilité de deux sorties du même hôpital le même jour pour un même bénéficiaire (contrairement aux dates d'admission, suite à un transfert vers un autre hôpital le jour d'admission).

A noter que ces trois dernières variables constituent ce que l'on appelle la clé primaire du séjour, qui l'identifie de manière univoque. Voir 4.1.1.10 pour la composition de la clé primaire.

4.1.1.7 Définition d'un nouveau séjour dans les données

A l'issue de l'algorithme de reconstitution des séjours hospitaliers, deux nouvelles variables seront créées :

1. une variable SS00015bis reprenant le début du séjour comptable : à remplir avec la première date de début de prestation des codes forfaits et correspondant à une variable similaire à créer dans les données TCT (à nommer p.ex. <date_first>) ;
2. une variable SS00125bis reprenant la fin du séjour comptable : à remplir avec la date de fin calculée sur base des dates de début de prestation des codes forfaits du séjour correspondant (à calculer dans les données TCT et à nommer p.ex. <date_last>)[h].

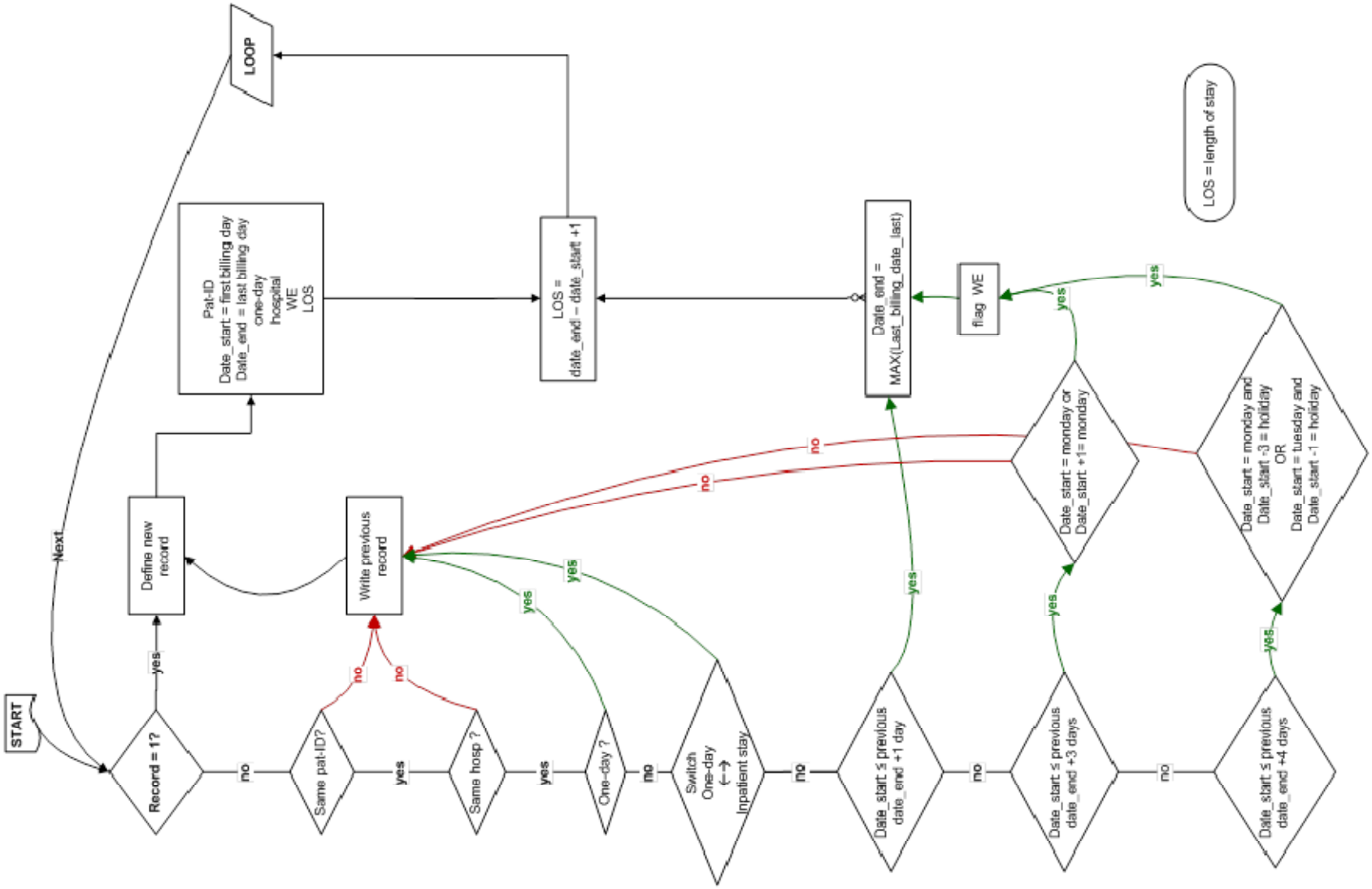
Un nouveau séjour est créé lorsqu'un des événements suivants est détecté :

1. changement de bénéficiaire ;
2. changement d'hôpital de séjour ;
3. succession de séjours en hôpital de jour (optionnel en fonction des questions de recherche) ;
4. passage de l'hôpital de jour à un séjour classique et vice-versa (optionnel en fonction des questions de recherche) ;
5. écart entre deux séjours consécutifs plus grand que 1 jour, sauf week-ends et jours fériés (voir Table 1 ci-dessous) ;
6. passage d'un séjour classique à une revalidation (optionnel en fonction des questions de recherche - non pertinent pour la présente étude tant que le patient ne quitte pas l'hôpital).

L'algorithme est décrit à la Figure 4 (LOS : length of stay – durée de séjour, Pat-ID : variable identifiant le patient).

[h] La date de sortie sera égale à la dernière date de début de prestation + nombre de journées facturées à cette date – 1

Figure 4 – Schéma de boucle de l'algorithme de reconstitution des séjours dans les données EPS



4.1.1.8 Le problème des interruptions temporaires de l'hospitalisation

Le dernier alinéa de l'art. 3 de la Convention nationale entre les établissements hospitaliers et les organismes assureurs stipule qu'en cas d'octroi de congé le jour de départ ne peut être porté en compte que lorsque le bénéficiaire quitte l'établissement hospitalier après 14 heures et le jour de retour ne peut être porté en compte que lorsque le bénéficiaire revient à l'établissement hospitalier avant 12 heures.

Il arrive donc qu'un patient puisse passer la fin de semaine voire même un autre jour férié hors de l'hôpital. La Table 1 énumère les différentes configurations et le nombre de jours non-tarifiables correspondants, c'est-à-dire la longueur d'interruption de facturation des forfaits pour journée d'entretien à tolérer. Nous n'avons pas tenu compte des ponts impliquant plus de 3 jours non tarifiables (à titre d'exemple, pour les deux années, nous avons vérifié qu'à peine 10 séjours étaient concernés par les ponts de l'ascension).

Table 1 – Nombres possibles de journées non tarifiables par type de week-end / jour férié

Catégorie	Départ de l'hôpital	Retour à l'hôpital	Nombre de journées 'extra muros' non tarifiables	Journée(s) non tarifiées	Première journée tarifiée après le retour à l'hôpital
Week-end normal	vendredi soir / samedi matin	dimanche soir / lundi matin	2	samedi + dimanche	lundi
	samedi après-midi	dimanche soir / lundi matin	1	dimanche	lundi
Week-end prolongé Jour férié le lundi	vendredi soir / samedi matin	lundi soir / mardi matin	3	samedi + dimanche + lundi	mardi
	samedi après-midi	lundi soir / mardi matin	2	dimanche + lundi	mardi
Week-end prolongé Jour férié le vendredi	jeudi soir / vendredi matin	dimanche soir / lundi matin	3	vendredi + samedi + dimanche	lundi
	vendredi après-midi	dimanche soir / lundi matin	2	samedi + dimanche	lundi
Autres jours fériés (p.ex. le mercredi)	mardi soir / mercredi matin	mercredi soir / jeudi matin	1	mercredi pour l'exemple	autre jour de la semaine que lundi ou mardi ; un jour après le jour férié

Néanmoins, il faut signaler qu'il existe des cas particuliers. Le cas typique est celui d'un patient hospitalisé dans un établissement de soins aigus qui nécessite une rééducation fonctionnelle consécutive et prolongée. Supposons qu'il quitte l'hôpital aigu le samedi pour entrer à l'hôpital de rééducation chronique le lundi. Logiquement, ceci constituerait un nouveau séjour hospitalier, puisqu'il s'agit d'un changement d'hôpital (donc d'un autre numéro d'agrément INAMI).

4.1.1.9 Le problème des actes différés

Il arrive que certains actes de diagnostic en laboratoire de biologie clinique, de génétique humaine ou d'anatomopathologie soient prestés à une date ultérieure à la sortie du patient : tests spécialisés de toxicologie, médecine tropicale, analyse ADN (acide désoxyribonucléique), examens de cyto- ou histodiagnostic avec ou sans biochimie moléculaire... sur prélèvement humoral ou tissulaire. Les expériences en traitements de données hospitalières dans le passé nous ont appris que ces tests, typiquement effectués en sous-traitance hors de l'hôpital de séjour (extra muros), sont facturés bien au nom du patient (son numéro NISS) mais très souvent sans mention de l'hôpital où il/elle séjourne. Il en résulte que le lien entre ces prestations différées et le séjour correspondant est difficile à établir.

Si les questions de recherche le requièrent, et en admettant que ces actes soient reliés au dernier séjour précédent – ce qui n'est donc qu'une hypothèse, bien que hautement probable – on peut attribuer à ces enregistrements la clé primaire du séjour immédiatement précédent[i] : concrètement on complète le champ « hôpital de séjour » avec le numéro d'agrément de l'hôpital du dernier séjour précédent, ainsi que le champ « date de fin » (SS00125bis) avec la date de fin calculée sur base des dates de début de prestation des forfaits pour journée d'entretien, ce qui rétablit le lien entre ces prestations différées et le séjour précédent.

[i] En cas d'hospitalisations de jour répétitives – typiquement le cas pour les traitements de suivi en cancérologie – une telle attribution par convention pourrait être erronée.

4.1.1.10 Attribution d'une clé primaire univariée aux séjours et à leurs prestations

Dès le séjour délimité, il est souvent d'usage d'attribuer une clé primaire univariée à tout enregistrement (record) dans la base de données indiquant une tarification d'acte (ou actes[j]) effectué(s) durant le même séjour. Bien que l'on parle à cet égard souvent de numéro de série il convient de souligner qu'en matière d'indexation les numéros de série sont fréquemment de caractère alphanumérique. Stricto sensu de telles clés univariées ne sont pas vraiment indispensables, mais dans des bases de données complexes, à multiples tables, elles sont extrêmement utiles.

Une méthode facile et en même temps fiable consiste à créer dans la base de données un nouveau champ, communément appelé clé primaire – à nommer p.ex. stay_pk_EPS – où l'on inscrit la concaténation des variables de l'index du séjour en question, notamment (dans l'ordre indiqué, ce qui nous apporte l'avantage d'une chronologie quasi automatique) : 1) date de sortie, 2) numéro d'agrément INAMI de l'hôpital et 3) pseudonyme du bénéficiaire.

4.1.1.11 Contrôles

Les principaux contrôles se font sur la séparation correcte des séjours classiques et de jour, sur les dates de début et de fin de prestation et sur la durée de séjour. Un contrôle important à faire concerne une vérification de l'équivalence de la date de dernière prestation pour les codes forfaits, date rapportée dans le champ SS00125 avec la même date calculée sur base de la combinaison de la date de début (champ SS00015) et le nombre de cas rapportés au champ SS00050.

[j] Certains actes – notamment les forfaits de jour – sont tarifiés 'en bloc', avec une date de début et une de fin.

4.1.2 Calcul des remboursements

Le montant total des remboursements par séjour, tels que calculés par la TCT, comprend les totaux des montants remboursés par l'AMI pour les honoraires médicaux, les produits pharmaceutiques, les fournitures telles que le sang, le plasma, le lait maternel et les radio-isotopes, les implants et les prestations de biologie clinique. A ces cinq totaux s'ajoute encore le montant calculé pour le prix de journée.

En ce qui concerne les prestations de biologie clinique, d'imagerie médicale et les spécialités pharmaceutiques, les forfaits à l'admission[k] ou par jour[l] sont comptabilisés dans le calcul ainsi que la rémunération (partielle) à l'acte (dans d'autres études,[28] les forfaits peuvent être écartés et les montants à l'acte extrapolés).

En ce qui concerne le prix de journée, le nouveau système de financement introduit en juillet 2002 octroie aux hôpitaux 80% de leur budget des moyens financiers sous forme d'avance mensuelle (absente des données couplées de la TCT). Les 20% restants sont facturés à l'acte sous forme d'un forfait par jour et d'un forfait par admission, enregistrés dans les données hospitalières couplées de la TCT. La manière appliquée par la TCT pour évaluer les montants remboursés par l'assurance maladie-invalidité pour un séjour hospitalier consiste à multiplier le nombre de journées d'hospitalisation facturées au cours du séjour hospitalier par le prix de journée à 100%. Ce prix de journée à 100% est calculé par hôpital par type de journée (forfait par jour) par le Service Public Fédéral de la Santé publique sur base du budget des moyens financiers de l'année et du nombre de journées d'hospitalisation de l'année précédente de l'hôpital. Ce prix de journée à 100% inclut la quote-part personnelle des patients (quelques pourcents). Les montants des forfaits par admission n'entrent pas en compte dans le calcul. La méthode est la même pour l'hospitalisation classique comme pour l'hospitalisation chirurgicale de jour. Pour l'hospitalisation de jour non-chirurgicale, il n'y a que des forfaits par jour et donc aucun montant par admission à écarter du calcul.

Deux caractéristiques spécifiques au financement actuel des hôpitaux doivent être prises en compte lors de l'analyse de représentativité des remboursements des données TCT-EPS.

Les forfaits et le prix de journée à 100% constituent la première caractéristique. Concrètement, ils impliquent qu'un même APR-DRG peut faire l'objet d'un remboursement différent selon l'hôpital où le patient a été admis. Vu que l'EPS n'est pas représentatif au niveau de l'hôpital, il n'est pas exclu que dans les données couplées TCT-EPS, certains APR-DRG se retrouvent plus concentrés dans des hôpitaux avec un haut (ou un bas) prix de journée.

La seconde caractéristique a trait au remboursement qui varie selon que le patient a droit ou non à une intervention majorée. Le prix de journée à 100% inclut le ticket modérateur que le patient paye par journée. Dans ce cas, le statut du patient n'a pas d'influence. Mais les remboursements à l'acte (par ex. les honoraires et une partie des actes de biologie clinique et d'imagerie médicale), le montant du remboursement dépend du statut d'assurabilité du patient.

4.1.3 Couplage des séjours hospitaliers

4.1.3.1 Choix des variables d'appariement

Pour pouvoir procéder au couplage des séjours il nous faut les délimiter d'une manière univoque dans les deux bases de données de départ. Autrement dit : quels sont les éléments présents dans ces deux bases de données qui permettent l'appariement des séjours? Pour identifier le séjour hospitalier à partir des données, 3 variables suffisent : 1) le (pseudonyme du) patient, 2) l'identification de l'hôpital de séjour et 3) la date intégrale de sortie ou, comme proposé ici, la dernière date de la prestation des forfaits pour journée d'entretien. Prises ensemble ces variables constituent la clé primaire, qui apporte son unicité au séjour (voir section 4.1.1.10).

k Forfaits à l'admission : Biologie clinique, imagerie médicale, produits antibioprophylactiques et certaines autres spécialités pharmaceutiques.

l Forfaits par jour : Biologie clinique et imagerie médicale.

Variables primaires d'appariement

Les trois variables précitées formeront les variables primaires d'appariement (matching variables dans la littérature anglo-saxonne).

Bien que les données TCT, plus précisément le Recordtype 3 (RT3 – fichier JR_ENTR) des SHA[29] qui groupe entre autres les forfaits pour les journées d'entretien, nous livrent les dates précises des prestations, la **date de fin** de séjour devra très souvent être calculée, étant donné que ces forfaits sont généralement facturés en bloc (p.ex. date de début de prestation = 01/01/2011 et nombre de forfaits par jour facturés = 4 nous donne le 04/01/2000 comme date de fin de séjour comptable[m]).

Quant à la date d'admission, elle est accessoire puisque la date de sortie en soi est univoque, vu l'impossibilité de deux sorties le même jour pour un même patient. En outre, dans certains cas, il existe une imprécision concernant la date d'admission, notamment pour les nouveau-nés. Normalement, c'est-à-dire pour un nouveau-né sans problèmes médicaux graves, il est de coutume de le laisser à la chambre de la maternité auprès de sa mère et tout acte médical concernant le bébé – p.ex. examen pédiatrique de routine – est facturé au nom de la mère. Si sa condition reste satisfaisante, le nouveau-né quitte la maternité avec sa mère et un seul séjour hospitalier est comptabilisé (au nom de la mère).

Si, par contre, le bébé présente des complications sévères, il est séparé de sa mère pour admission dans un service de soins intensifs néonatals (néonatal intensive care – NIC) voire même de chirurgie pédiatrique. Du point de vue de la comptabilité, un nouveau séjour hospitalier est créé, différent de celui de la mère. C'est à cet instant que la possibilité de confusion émerge, certains hôpitaux enregistrant comme date d'admission la date de naissance et d'autres la date de début de la prise en charge du bébé en détresse.

En ce qui concerne l'identifiant de l'hôpital, les données TCT ont déjà fait l'objet d'une uniformisation prenant en compte notamment les fusions d'hôpitaux ayant eu lieu dans l'année d'enregistrement. L'univocité dans les données EPS nécessite un travail de préparation considérable (voir section 4.1.1.5 ci-dessus), y compris la vérification des fusions des hôpitaux, le cas échéant suivi par une adaptation (dans le champ parallèle SS00075bis).

La date de sortie mérite également quelques considérations. En effet, dans certains cas un décalage d'un jour entre la date de sortie et la dernière date de prestation des forfaits pour journée d'entretien est possible. Suite aux règles de tarification imposées par l'INAMI, c'est le cas des séjours où le patient quitte l'hôpital avant 14.00 h (perte du droit au forfait pour journée d'entretien). Ainsi on utilisera la dernière date de prestation comme « matching variable » partant de l'expérience que les données comptables de tarification font preuve d'une robustesse plus fiable que les autres enregistrements administratifs.

Quant à la date réelle d'admission des séjours classiques, elle est plus facilement identifiable dans les hôpitaux à soins aigus (à l'exclusion des centres pour grands brûlés, les services Sp d'hôpitaux mixtes (lits aigus et chroniques) ou G isolés), pour lesquels s'applique le forfait par admission, facturé à la date même de l'admission, indépendamment de l'heure de l'entrée, avant ou après midi.

Variables secondaires d'appariement

En ce qui concerne le contrôle du couplage au-delà des variables qui ont servi à coupler, la concordance du type d'hospitalisation sera également examinée (hospitalisation classique ou de jour).

Données de base nécessaires au couplage et sa validation

En résumé, il nous faut les composantes des clés primaires des séjours hospitaliers des deux bases de données de départ (TCT et EPS), comprenant :

1. le pseudonyme du patient issu des tables de correspondance des hôpitaux et des OA lors du couplage RHM-SHA-HJA effectué par la TCT et recodé par la plateforme eHealth en appliquant un algorithme à clé spécifique au projet. Ce pseudonyme à été inséré dans les deux bases de données : les champs SS00010 du côté EPS et <patient_anonymous_id> du côté TCT,
2. le numéro d'agrément normalisé de l'hôpital de séjour : les champs SS00075bis, créé du côté EPS et <stay_place_id> du côté TCT,

[m] La formule étant : date de début de prestation + nombre de prestations - 1

3. la dernière date de la prestation des forfaits pour journée d'entretien : les champs SS00115bis du côté EPS et <date_last> du côté TCT. Du côté données TCT ce champ devra être créé à partir des données du fichier Recordtype 3 (RT3) ;
4. la première date de la prestation des forfaits pour journée d'entretien : les champs SS00015bis du côté EPS et <date_first> du côté TCT. Du côté données TCT ce champ devra être créé à partir des données du fichier Recordtype 3 (RT3) ;
5. Le type d'hospitalisation : sur base de la présence dans le fichier SHA ou HJA dans les données TCT et du champ SS00020 reprenant le code de nomenclature spécifique (voir plus haut) dans l'EPS.

4.1.3.2 Préparation au couplage

Deux tables reprenant les variables d'appariement primaires et secondaires seront créées (voir Figure 5), l'une extraite des données EPS et l'autre des données TCT.

Avant d'effectuer le couplage, certains séjours sans données correspondantes dans l'EPS doivent être éliminés des données TCT :

1. élimination des mini-forfaits et forfaits pour utilisation de la salle de plâtre non retenus dans l'algorithme de reconstitution des séjours du côté EPS (cf. section 4.1.1.4) ;
2. élimination des séjours sans codes forfaits (automatiquement éliminés par la Cellule Technique en 2008 mais pas en 2009) ;
3. élimination des séjours débutant avant 2008 puisque seuls les séjours avec prestation en 2008 ou 2009 sont retenus dans l'EPS ;
4. élimination des séjours dont le montant des remboursements est négatif ou nul, pour des raisons de validation ;
5. élimination des séjours des patients dont l'âge est manquant, inutilisables dans notre analyse ;
6. élimination des séjours d'assujettis en Belgique ou de patients étrangers dont les remboursements de soins sont couverts par des conventions internationales conclues entre la Belgique et les pays étrangers (deuxième chiffre du code titulaire 1 = 8). Aucune donnée relative à ces patients n'est disponible dans les données de l'AIM donc a fortiori dans l'EPS.

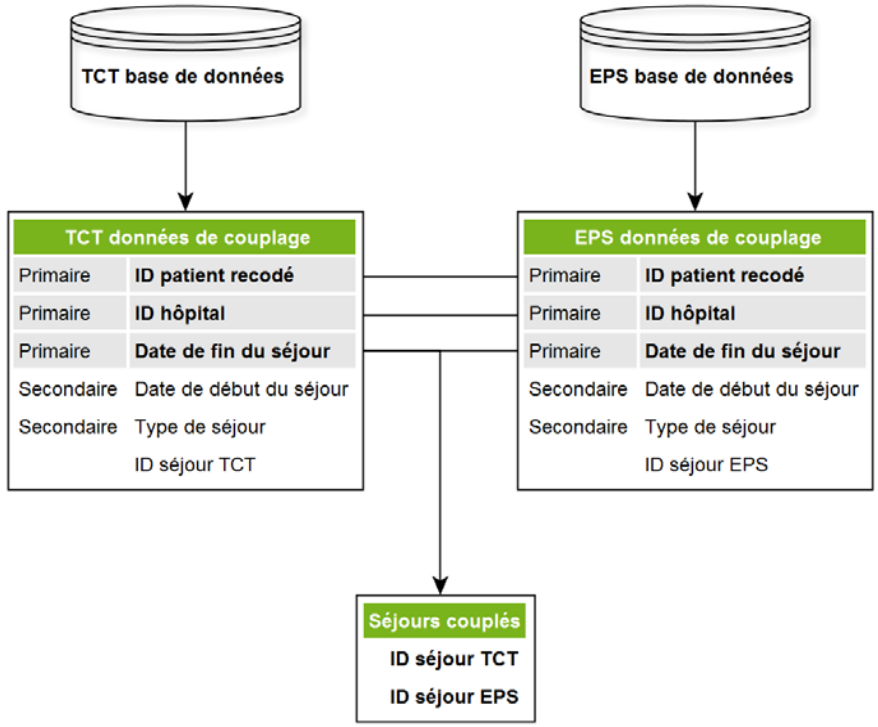
Ensuite ces deux tables sont mises en relation grâce aux trois variables qui composent leur clé primaire.

Dans la table de sortie on reprend les clés primaires univariées (voir section 4.1.1.10) des séjours TCT et des séjours EPS.

4.1.3.3 Couplage

Le couplage s'effectue automatiquement par l'exécution de la requête (voir Figure 5), qui ne sélectionnera que les séjours à valeurs équivalentes dans les champs-clé des deux tables citées et nous permettra aisément de les isoler dans la table de sortie (« output table ») c'est-à-dire la table des clés primaires couplées (ci-après nommée « Link_table »).

Figure 5 – Technique du couplage



4.1.3.4 *Validation du couplage*

Validation sur base des variables secondaires

Il est de bonne pratique d'envisager quelques contrôles sur la validité du couplage. Comme détaillé à la section 4.1.3.1 (variables secondaires d'appariement), on utilisera la première date de prestation des codes forfaits pour journée d'entretien et le type d'hospitalisation. La validation réside dans la comparaison des deux variables entre les données TCT et EPS, couplées sur base de l'identification patient, l'identification hôpital et la date de fin du séjour.

Le taux de couplage

Finalement, on calcule le taux global de couplage. Ce taux exprimé en pourcentage est le résultat de la division dans laquelle le numérateur représente le nombre de séjours EPS couplés et validés, et le dénominateur représente le nombre total de séjours dans les données EPS.

Ce taux de couplage nous fournit un test ultime de validation. En effet, vu que le couplage réalisé peut être qualifié de type « appariement ou couplage déterministe »[n], vu la sélection et la préparation méticuleuse des variables d'appariement ainsi que l'utilisation des dates comptables de prestation, on devrait s'attendre à un taux de couplage très élevé, d'environ 95%. Ce taux à priori se base sur les taux de couplage réalisés par la TCT lors des couplages RHM-SHA, depuis déjà plus de 10 ans.

En revanche, un taux de couplage faible indiquerait la présence d'importantes discordances dans les données administratives de départ.

Le traitement des données a été réalisé en SAS 9.3.[31]

[n] Certaines publications opposent le terme « appariement exact » à « appariement probabiliste ». L'appariement probabiliste se fonde sur la théorie du rapport de vraisemblance pour attribuer des résultats de comparaison à la décision la plus correcte ou la plus plausible. Cette méthode s'appuie sur la théorie statistique et l'analyse de données. Comparée à l'approche déterministe, elle permet d'établir un lien plus étroit entre les enregistrements lorsqu'ils comportent des erreurs typographiques et des schémas d'erreurs plus complexes.[30]

4.2 Résultats

4.2.1 *Reconstitution des séjours hospitaliers dans les données EPS*

La méthodologie décrite ci-dessus nous a permis de reconstituer 218 438 séjours dans les données soins de santé de l'EPS pour les années 2008 et 2009 sans rencontrer de problème particulier une fois la liste des codes nomenclature concernant les journée d'entretien et les forfaits d'admission correctement définie (voir section 4.1.1.4 et liste en annexe 4).

La Table 2 montre la répartition entre les différents types d'hospitalisations par région et pour la Belgique sur base du lieu d'hospitalisation. 53,3% correspondent à un séjour classique (dont 0,47% débutant avant la période d'analyse 01/01/2008), 14,6 % sont des séjours chirurgicaux de jour et 29,5% des hospitalisations de jour. Le reste (2,6%) est principalement composé de séjours en service SP et d'hospitalisation de jour en service A (sans admission). Sans les inconnus, ces pourcentages sont respectivement de 54,7%, 15,0% et 30,3%. En comparaison, les pourcentages observés sur les deux années dans les données nationales TCT étaient de 54,3%, 16,4% et de 29,4%.

Table 2 – Nombre et type d'hospitalisations reconnues dans l'EPS pour les années 2008 et 2009

Type d'hospitalisation	Belgique Nombres	Pourcentage des types d'hospitalisation pour chaque région				Pourcentage des régions pour chaque type d'hospitalisation			
		Bruxelles	Flandres	Wallonie	Belgique	Bruxelles	Flandres	Wallonie	Belgique
Classique	116 448	54,4%	51,6%	56,6%	53,3%	12, 5%	58,2%	29,3%	100%
Hospitalisation de jour	64 472	27,5%	30,9%	27,5%	29,5%	11,4%	62,9%	25,7%	100%
Chirurgie de jour	31 893	14,7%	15,4%	12,9%	14,6%	12,3%	63,3%	24,4%	100%
Inconnu	5 625	3,4%	2,2%	3,1%	2,6%	16,0%	50,6%	33,4%	100%
Total	218 438	100%	100%	100%	100%	12,2%	60,2%	27,6%	100%

L'algorithme de reconstitution des séjours a permis de reconstituer **218 438 séjours** sur l'EPS 2008-2009, prêts à être couples aux données TCT.

4.2.2 Couplage des séjours hospitaliers

4.2.2.1 Préparation au couplage

Ci-dessous la Table 3 reprend les taux de couplage des tables de correspondance des différents organismes assureurs (reçues dans le cadre du présent projet) entre les identifiants patients issus des documents ADH (hospitalisations de jour) et SHA (hospitalisations classiques) et les identifiants patients issus de l'EPS des années 2008 et 2009. Nous avons contacté la Cellule Technique qui ne dispose actuellement d'aucune explication pour les taux plus faibles des OA 2, 5 et dans une moindre mesure 6.

Table 3 – Taux de couplage des tables de correspondance (HJA - ou SHA-EPS), par organisme assureur, année 2008 et 2009 (pourcentage de séjours)

	Pourcentage couplage 2008		Pourcentage couplage 2009	
Organisme assureur	HJA	SHA	HJA	SHA
1	99,65%	99,31%	99,70%	99,49%
2	81,60%	80,80%	81,13%	80,34%
3	99,90%	99,76%	99,86%	99,71%
4	99,97%	99,93%	99,98%	99,95%
5	89,87%	89,22%	89,34%	89,47%
6	93,02%	91,53%	95,40%	92,96%
9	100%	100%	100%	99,87%
TOTAL	97,47%	97,12%	97,34%	97,17%

Globalement, pour 2,72 % des séjours SHA ou ADH, la clé d'identification du patient n'a pas été retrouvée dans l'EPS.

Le nombre de séjours transmis par la Cellule Technique se montait à 7 558 507 séjours pour les deux années.

Les données TCT nationales ont été agencées en une seule table reprenant, par séjour, les informations suivantes :

- caractéristiques démographique du patient : sexe, âge, région, province, arrondissement, tarif préférentiel ou non ;
- caractéristique de l'hôpital de séjour : numéro d'agrément, région, province, arrondissement ;
- caractéristique du séjour : date de début et date de fin, MDC, APR-DRG, diagnostic principal, niveau de sévérité, montant total des remboursements.

Avant d'effectuer le couplage, certaines vérifications et/ou éliminations de certains séjours ont été exécutées :

1. élimination des mini-forfaits et forfaits pour utilisation de la salle de plâtre non retenus dans l'algorithme de reconstitution des séjours du côté EPS (1 281 250 forfaits) ;
2. élimination des séjours sans codes forfaits (automatiquement éliminés par la Cellule Technique en 2008 mais pas en 2009) (5 190 séjours) ;
3. élimination des séjours débutant avant 2008 (34 998 séjours) ;
4. élimination des séjours dont le montant total des remboursements est négatif ou nul (3 312 séjours) ;
5. élimination des séjours des patients dont l'âge est manquant (22 séjours) ;
6. élimination des séjours de patients dont les remboursements de soins sont couverts par des conventions internationales conclues entre la Belgique et les pays étrangers (deuxième chiffre du code titulaire 1 = 8) (93 540 séjours).

Après nettoyage suivant les différents critères précités, les données TCT nationales qui seront utilisées pour effectuer et le couplage avec les données EPS et qui serviront comme référence nationale dans les analyses comprennent **6 163 767 séjours** pour les deux années.

4.2.2.2 Couplage

Le couplage des séjours implique deux fichiers, celui des séjours TCT et celui des séjours EPS, comprenant tous deux les variables de couplage. Un certain nombre de patients présents dans les séjours EPS sont absents des séjours TCT. Les séjours de ces patients ne sont a priori pas couplables et ne sont donc pas repris dans la description du couplage (Table 4).

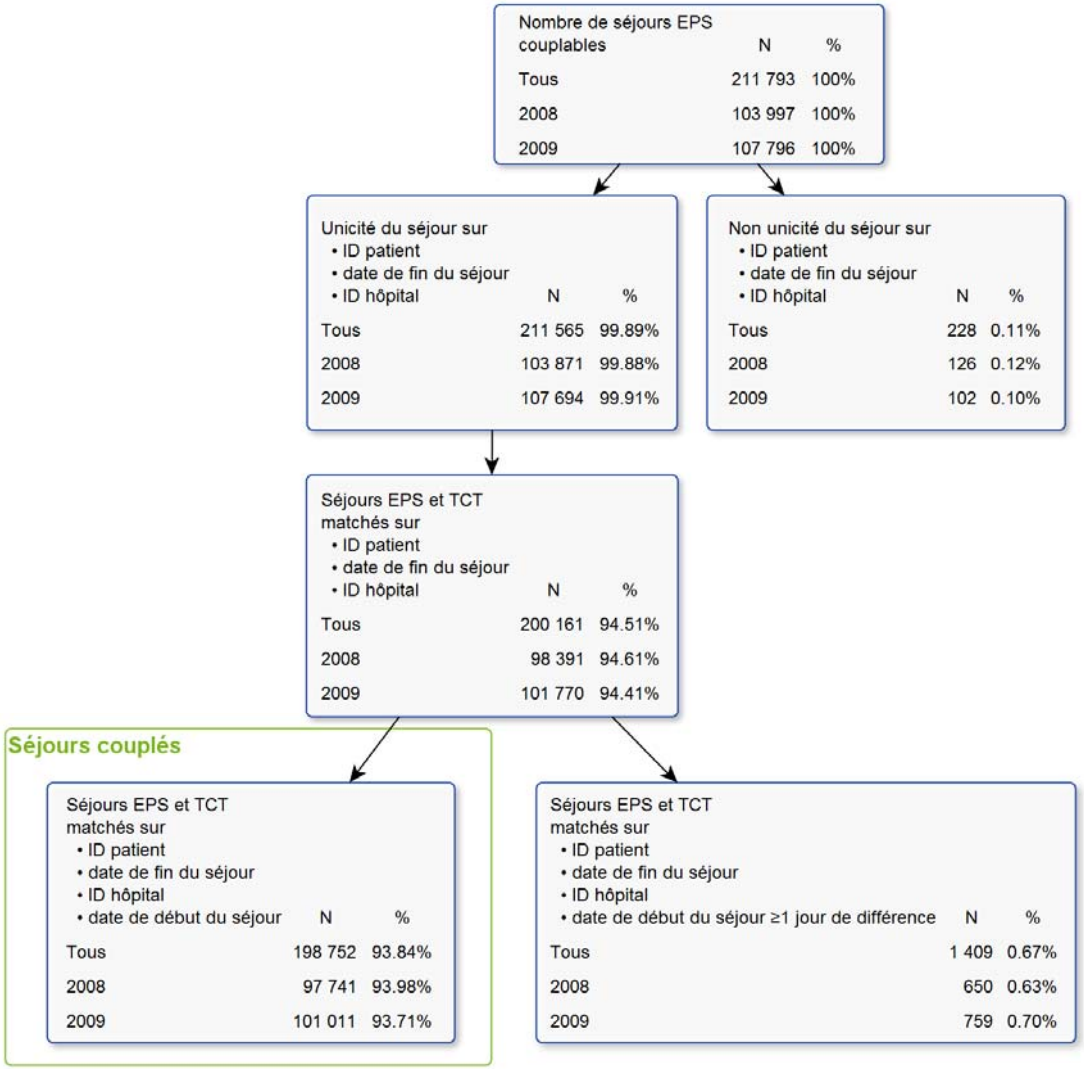
Lorsque, dans la présente section, il est fait référence aux « séjours EPS », il s'agira uniquement des séjours potentiellement couplables (colonne « N restants » dans la Table 4).

Table 4 – Rejet de séjours EPS dû à l'absence d'identifiant patient

	N	%	N restants	% restants
Patients	4 238	4,14%	98 037	95,86%
Séjours	6 645	3,04%	211 793	96,96%
2008	3 291	1,59%	103 997	96,93%
2009	3 354	1,62%	107 796	96,98%

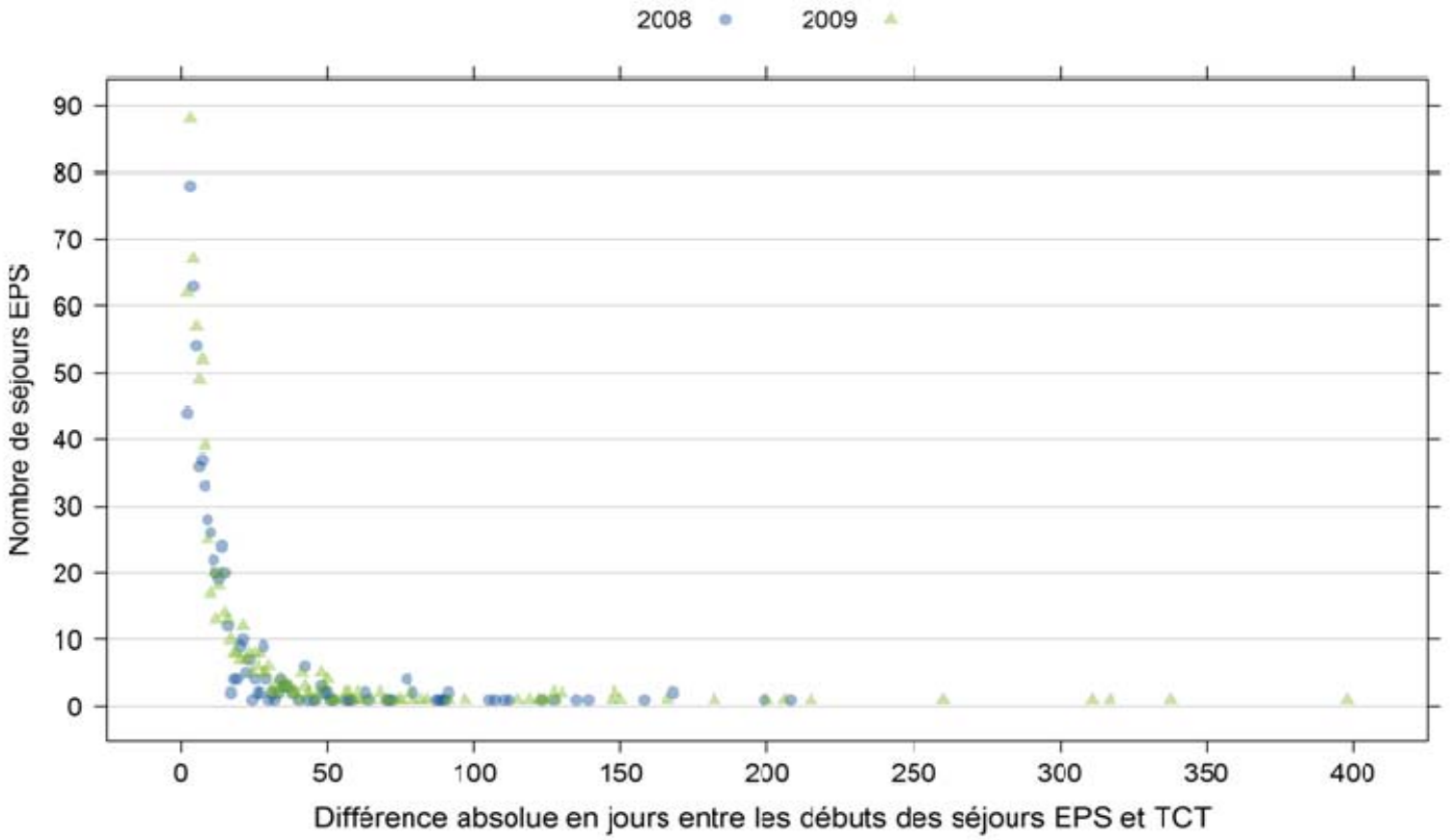
La Figure 6 détaille les résultats de la procédure de couplage.

Figure 6 – Détail de la procédure de couplage EPS-TCT (les pourcentages sont toujours exprimés sur le nombre total de séjours EPS).



Un petit nombre de séjours EPS et TCT (0,67%) ont les mêmes identifiant patient, date de fin et identifiant hôpital mais ont des dates de début qui diffèrent d'un 1 jour ou plus. Ces différences sont présentées à la Figure 7. Nous n'avons pas trouvé d'explications pour les quelques séjours avec une très grande différence.

Figure 7 – Nombre de séjours EPS en fonction de la différence en jours (en valeur absolue) avec la date de début TCT, par année (séjours sans différence exclue)



4.2.2.3 *Validation du couplage*

Contrôle des séjours couplés

Le type d'hospitalisation (classique ou de jour) des séjours couplés concordent entre les sources EPS et TCT excepté pour 2 séjours (<0,01%).

Séjours non couplés

Relativement, Il y a plus d'hospitalisations de jour couplées EPS que d'hospitalisations classiques couplées : rapport de cotes (odds ratio) = 2,28 [IC 95% 2,19 – 2,37], χ^2 (dl :1,n : 211 793) = 1 725,02, p<0,001. En d'autres termes, les séjours classiques ont plus de chance de ne pas être couplés que les séjours en hospitalisations de jour.

Le pourcentage de séjours non couplés varie considérablement par hôpital (Figure 9). Le pourcentage de non-couplage est élevé, surtout dans les hôpitaux avec peu de séjours EPS. Mais, compte tenu de leur faible nombre de séjours, la variabilité se situe dans la fourchette attendue. Le graphique montre que certains hôpitaux sont en dehors de la variabilité attendue, mais les données n'indiquent pas la raison.

Figure 8 – Couplage ou non des séjours en fonction du type d'hospitalisation (les quarts de cercle correspondent au pourcentage de séjours EPS)

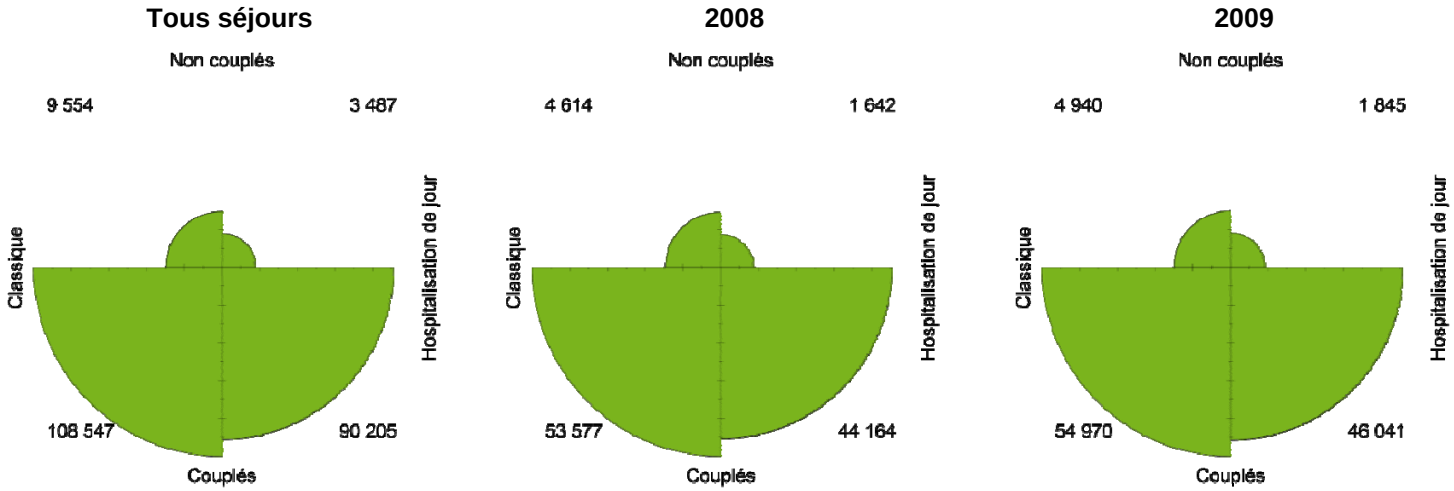
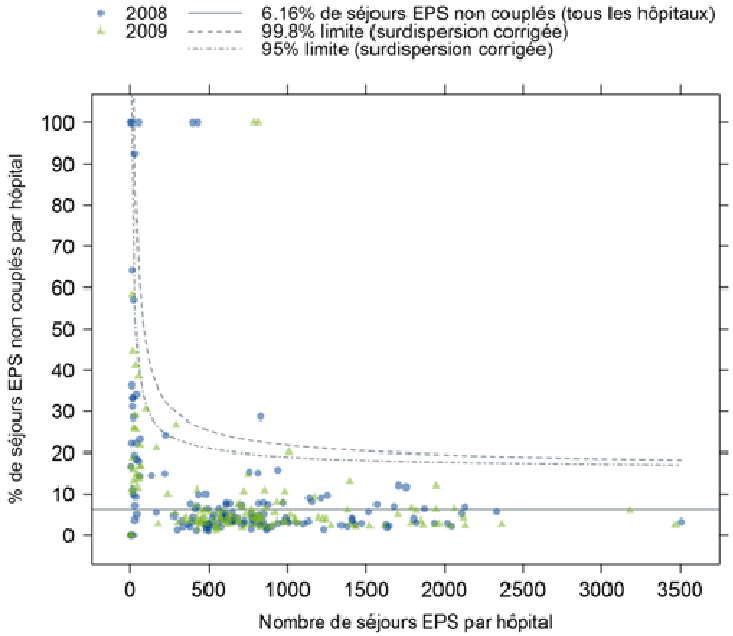


Figure 9 – Pourcentage des séjours EPS non-couplés par hôpital selon le nombre de séjours EPS par hôpital[o]



Les données couplées EPS-TCT représentent un total de **198 752 séjours** pour les deux années 2008 et 2009.

[o] Limites de variabilités : méthode Agresti-Coull,[32] corrigées pour surdispersion par paramètre de surdispersion, calculé sur des données 20% winsorized pour des raisons de robustesse.[33] Le winsorizing revient à remplacer les valeurs les plus extrêmes par la valeur d'un percentile prédéterminé (dans la présente etude, les percentiles 20 et 80).

5 TESTS DE REPRÉSENTATIVITÉ

5.1 Méthodologie

5.1.1 Importance de la représentativité d'un échantillon

Le but d'un échantillon EPS couplé aux données TCT consiste à pouvoir utiliser les données cliniques dans des analyses de sous-groupe ou d'ajustement par rapport à des facteurs de risque.

Technique d'échantillonnage

L'objectif de tout échantillonnage est de pouvoir inférer des conclusions concernant la population de référence sur base des données de l'échantillon, c'est-à-dire de pouvoir tirer des conclusions généralisables à la population.[34]

Par définition, si l'échantillon est sélectionné par une méthode probabiliste (sélection aléatoire), alors les estimations de résultats issues de cet échantillon sont fiables et l'échantillon est dit représentatif de la population dont il est tiré, pour peu que la taille de l'échantillon soit suffisante. La précision (marge d'erreur) des estimations est en effet liée à la taille de l'échantillon.

Par définition, l'EPS est composé de 1/40 des assurés sociaux affiliés ou inscrits auprès d'un organisme assureur, tirés de manière stratifiée sur base de la date de naissance et du sexe (voir section 2.1). Un tirage supplémentaire de 1/39 des assurés âgés de 65 ans ou plus assure un sur-échantillonnage permettant de disposer d'une personne sur 20 dans cette tranche.

Pour l'EPS néanmoins il est important de tester sa représentativité afin de pouvoir extrapoler les résultats obtenus lors d'analyses de cet échantillon à la population hospitalière nationale et de conclure à des résultats non-biaisés. La représentativité de l'échantillon doit être testée du point de vue global ainsi qu'au niveau des différents APR-DRG et qu'au niveau APR-DRG-SOI, puisque ces niveaux de groupement des séjours sont communément utilisés lorsqu'on étudie les données RHM.

Il est à noter que si la représentativité n'est pas vérifiée pour certaines catégories (sur- ou sous-représentativité de la catégorie dans l'échantillon), elle permet d'accompagner les résultats de toutes les précautions et limitations d'usage. Il se peut aussi qu'en connaissant la grandeur de la sur- (ou sous-) représentativité de l'échantillon pour un facteur, on puisse ajuster ou décider de stratifier les analyses propres à chaque étude en fonction de ce facteur.

Il est aussi important de noter les différents points d'attention/limitations de la question de recherche :

1. un échantillon n'est représentatif que sur les variables sur lesquels les tests de représentativité ont portés,
2. les analyses de sous-groupe d'un échantillon peuvent ne pas être représentatives pour une variable alors que l'échantillon est globalement représentatif par rapport à cette variable,
3. en règle générale, le sujet d'étude détermine les facteurs sur lesquels la représentativité doit être testée afin de pouvoir présenter les résultats dans un cadre adéquat (extrapolation ou stratification des analyses selon le résultat du test de représentativité de l'échantillon).

5.1.2 Méthodes pour tester la représentativité d'un échantillon

Méthodes

Pour les analyses, nous avons utilisé des techniques de sondage (« Survey methods ») propres aux données d'échantillons. Ces techniques prennent en compte le plan d'échantillonnage (notamment les facteurs de stratification) et permettent d'obtenir un estimateur pour le paramètre étudié (moyenne, médiane, taux de répartition...) ainsi qu'une précision de cet estimateur (erreur d'échantillonnage).

Ces méthodes prennent en considération le poids d'échantillonnage. Ce poids représente le nombre que chaque unité de l'échantillon représente par rapport à la population. Chaque unité de l'échantillon est classifiée de manière univoque dans une strate. A chaque strate est attribué un poids calculé en divisant le nombre d'unités issues de la population de référence appartenant à cette strate par le nombre d'unités incluses dans l'échantillon et appartenant à cette même strate. Dans notre étude, nous avons analysé les données selon deux types de poids différents.

Pour rappel, l'EPS est un échantillon stratifié de la population sur base de l'âge et du sexe, comprenant les données de 1/40ème des assurés sociaux (population AIM nationale) plus les données de 1/39ème supplémentaire des assurés sociaux âgés de 65 ans ou plus. Ceci revient à une représentation de 1/40ème pour les personnes âgées de 0-64 ans et de 1/20ème pour les personnes de 65 ans ou plus. Les poids utilisés dans la première méthode (dénommés « Poids AIM » dans la suite du rapport) sont calculés par rapport aux données AIM nationales (voir annexe 5.1.1. Table 1).

Lors du couplage de l'échantillon EPS-TCT, certains séjours n'ont toutefois pas pu être couplés. Le poids que chaque unité d'échantillon représente varie donc par rapport au poids théoriquement calculé par rapport aux données nationales AIM. Par conséquent, nous avons ajusté ces poids sur base des informations connues, c'est-à-dire le nombre total de séjours par strate dans les données nationales TCT et le nombre de séjours par strate dans l'échantillon (EPS-TCT). Cette pondération (« post-stratification ») permettra un redressement de l'échantillon plus en adéquation avec les données hospitalières. Dans la suite du rapport nous appellerons ces poids « Poids TCT » (voir annexe 5.1.1. Table 1).

Les analyses s'effectueront en 3 étapes :

- globalement sur tous les séjours (tous les APR-DRG confondus) ;
- par APR-DRG ;
- et par APR-DRG et niveau de sévérité (SOI[p]) .

Dans un premier temps, nous nous intéresserons à la répartition de tous les séjours et nous répondrons aux questions suivantes :

1. Les séjours couplés (données TCT-EPS) représentent-ils 2,5% et 5% des séjours TCT pour les classes d'âge <65 et >= 65 ans, respectivement ?
2. La distribution de l'âge et du sexe dans les données TCT-EPS sont-elles les mêmes que dans les données TCT ?
3. La répartition du nombre de séjours par APR-DRG dans les données TCT-EPS est-elle identique à celle que l'on retrouve dans les données nationales TCT ? Si non, quels sont les extrêmes ?

[p] SOI : Severity of Illness – Niveau de sévérité

4. Quels sont les APR-DRG à retenir pour effectuer le test par APR-DRG en fonction du nombre de séjours ?

Les analyses par APR-DRG n'ont été effectuées que sur les APR-DRG dont le nombre de séjours dans l'échantillon est d'au moins 30 séjours sur l'année afin d'exclure les trop petits nombres. Le détail du nombre de séjours par APR-DRG dans l'échantillon EPS-TCT et dans la population TCT est présenté en annexe 5.1.1 Table 11.

Dans un second temps, nous analyserons la représentativité de l'échantillon par APR-DRG. Et finalement les analyses seront effectuées par APR-DRG-SOI.

Les tests de la représentativité porteront sur les variables suivantes :

1. Age
2. Sexe
3. Niveau de sévérité (uniquement dans le cas des analyses par APR-DRG)
4. Type d'hospitalisation (hospitalisation de jour versus hospitalisation classique)
5. Données géographiques (dans cet ordre : Région, province et arrondissement – on s'arrêtera au niveau où la représentativité n'est plus respectée)
6. Remboursements totaux par séjour (selon l'algorithme utilisé par la Cellule Technique)
7. Répartition du diagnostic principal (avec près d'un millier de possibilités de diagnostics principaux à 3 positions[q], les niveaux à plus de 3 positions n'ont pas été analysés pour des raisons évidentes de nombre de séjours).
8. statut préférentiel ou non[r] : cette variable n'est pas disponible dans les données RHM mais bien dans les données SHA/HJA. Elle a été retenue en raison de son importance dès qu'il s'agit de traiter le montant total des remboursements

[q] 4 positions dans le cas des codes E : causes extérieures de traumatismes et d'empoisonnement

[r] Pour cette variable également, la référence nationale sera l'ensemble des données nationales TCT

Tests statistiques utilisés

Les tests pour valider la représentativité des données TCT-EPS dépendront des caractéristiques des variables (continue, dichotomique, catégorielle) sur lesquelles porte l'analyse. Le seuil de significativité (α) est fixé à 0,05.

De manière générale pour toutes les variables étudiées, un estimateur et un intervalle de confiance à 95% seront calculés sur base des données EPS-TCT en prenant en considération les spécificités de l'échantillonnage (facteurs de stratification, poids de chaque observation) et les caractéristiques de la variable étudiée (continue, dichotomique, catégorielle) via des procédures propres aux enquêtes (« SURVEY methods »). Les paramètres calculés sur l'entièreté des données TCT seront comparés à ces intervalles de confiance.[35]

1. Soit l'intervalle de confiance contient le paramètre calculé sur l'entièreté des données TCT et dans ce cas, on pourra affirmer que l'échantillon EPS-RHM est représentatif des données TCT.
2. Soit l'intervalle de confiance ne contient pas ce paramètre et dans ce cas, on ne pourra pas dire que l'échantillon EPS-RHM est représentatif (en tous cas au niveau étudié). Cela n'implique pas que l'échantillon n'est pas représentatif pour d'autres niveaux (analyse de sous-groupe).

Afin de montrer la représentativité des données de l'échantillon par rapport à la population de référence pour un facteur donné, nous avons comparé la répartition de ce facteur dans les deux bases de données (échantillon et population) mais nous nous sommes aussi intéressés à l'extrapolation du nombre de séjours – fréquences pondérées (nécessaires pour les analyses épidémiologiques). Les différents paramètres ont été calculés en utilisant d'une part une pondération dérivée des données AIM et d'autre part la pondération dérivée des données TCT. Nous avons ensuite comparé les estimations selon les deux pondérations. Les tests statistiques utilisés pour comparer les données de l'échantillon et celle de la population de référence étaient le test de Rao-Scott [36] pour les variables catégorielles et le test one-sample de Kolmogorov-Smirnov pour les variables continues (âge et montant total des remboursements) (références).

L'interprétation des résultats doit se faire selon un certain bon sens. En effet, la taille de l'échantillon ou du sous-groupe est déterminante pour les résultats statistiques (variance et p-value). Il est donc important d'aussi prendre en compte les différences absolues entre l'estimation et la valeur réelle de la population et considérer la taille de l'échantillon. Plus la taille de l'échantillon (ou sous-groupe) est grande, plus la probabilité de détection d'une petite différence sera grande. Il faudra donc décider d'un seuil d'acceptation de différence entre les résultats de l'échantillon et de la population (1%, 5% …) pour chaque étude.

Ces analyses seront supportées par des représentations graphiques permettant de visualiser les distributions des variables dans l'échantillon EPS-TCT et dans les données nationales TCT. Un EDF plot (Empirical Distribution Function) permettra de vérifier si les distributions des variables continues sont les mêmes pour l'EPS-TCT et les données TCT, des histogrammes (« bar charts ») et des « violin plots » permettront de vérifier les distributions pour les variables catégorielles. Les violin plots ont été tronqués : 0.5% des observations ne sont pas montrées (valeurs les plus hautes).

Les analyses ont été effectuées en SAS 9.3 et R 3.0.1. [31, 37]

5.2 Résultats

5.2.1 Résultats globaux

Pour rappel, les données nationales TCT disponibles servant de référence aux tests de représentativité contiennent 6 163 767 séjours pour les deux années 2008 et 2009. Les données couplées EPS-TCT représentent un total de 198 752 séjours.

Pour chacune des années, la proportion des séjours couplés EPS-TCT représente aux alentours de 2,3% et 4,7% des séjours des données nationales TCT pour les assurés de moins de 65 ans et ceux de 65 ans ou plus, respectivement (Table 5). Le fait que ces taux soient un peu en-dessous de ceux que représente l'EPS vis-à-vis de l'AIM (2,5% et 5% respectivement) est dû aux séjours non-couplés. Les résultats sont semblables au niveau patient. La Table 2 de l'annexe présente la proportion des séjours EPS-TCT par rapport aux séjours TCT selon la combinaison âge (< ou ≥ 65 ans) et sexe.

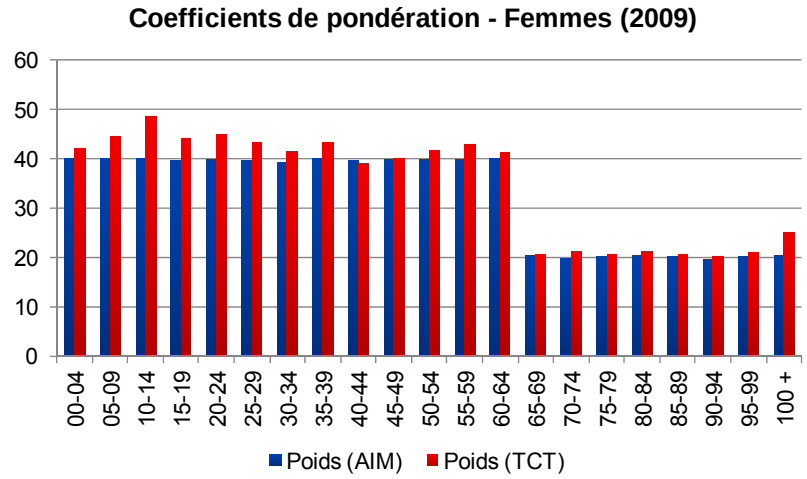
Table 5 – Proportion des séjours de l'EPS-TCT par rapport aux séjours TCT [IC 95%] par catégorie d'âge

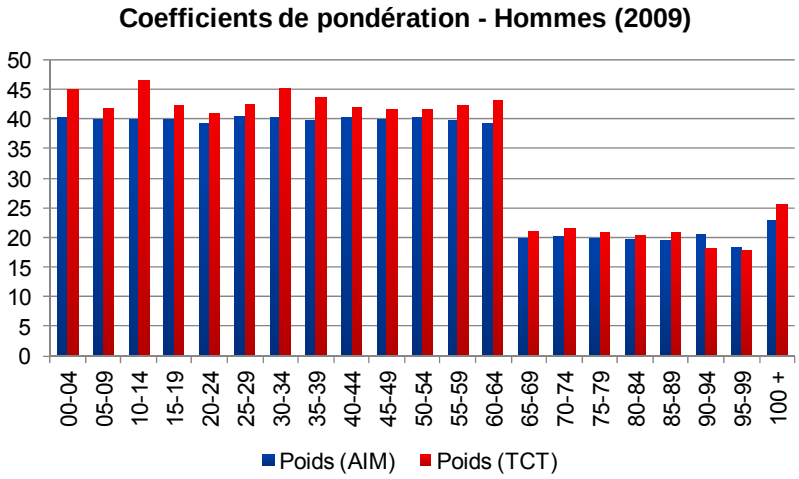
Catégorie d'âge	2008	2009
<65 ans	2,36% [2,33%; 2,38%]	2,34% [2,32%; 2,36%]
65 ans et +	4,77% [4,74%; 4,81%]	4,75% [4,71%; 4,79%]

Comme décrit dans la méthodologie, par strate, chaque observation de l'échantillon représente plusieurs observations de la population de référence dont elle est issue. Nous devons donc donner un coefficient de pondération à chaque observation de l'échantillon correspondant à l'inverse du taux d'échantillonnage dans cette strate. Nous appelons ces valeurs les « poids ». Les poids AIM sont dérivés directement des taux d'échantillonnage des patients de l'EPS parmi les patients dans les données AIM, tandis que les poids TCT sont dérivés du taux d'échantillonnage des séjours de l'EPS-TCT parmi les séjours dans les données nationales TCT couplées.

La Figure 10 représente les poids AIM et les poids TCT par catégorie d'âge et par sexe (détails en annexe). Les fluctuations observées dans les données EPS-TCT sont dues au fait que le calcul est effectué post-stratification, sur un sous-ensemble d'individus de l'EPS (les individus ayant été hospitalisés durant l'année) et que le taux de couplage avec les données TCT n'est pas de 100%.

Figure 10 – Poids AIM et Poids TCT par catégorie d'âge et par sexe - 2009





La Table 6 reprend les résultats de l'extrapolation du nombre de séjours par catégorie d'âge (< 65 et 65+) en prenant en compte les 2 types de poids (poids AIM et poids TCT) dans les analyses.

Les pourcentages relatifs sont exprimés sur base du nombre de séjours de la référence (données nationales TCT). En terme de différence relative (différence entre le nombre de séjours extrapolés et le nombre de séjours dans les données nationales TCT), nous observons que cette différence passe d'une sous-estimation de 6% (poids AIM – range : 4,8 – 6,9 ; par année et par catégorie d'âge) à une sous-estimation de 1% (poids TCT – range : 1,0 – 1,3 ; par année et par catégorie d'âge) pour chaque année. En utilisant les poids TCT pour calculer notre estimation du nombre de séjours, nous obtenons donc un nombre de séjours équivalent à 99% du nombre de séjours dans les données TCT. Dans le cadre d'analyses épidémiologiques, il est très important d'approcher le plus possible ce nombre de séjours.

Table 6 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par catégorie d'âge

		Référence (= TCT)	Echantillon	Echantillon (Poids AIM)	Echantillon (Poids TCT)
Année	Catégorie d'âge	Nb. Séjours TCT	Nb. Séjours EPS-TCT	Nb. Séjours extrapolés + Diff. Relative* (%)	Nb. Séjours extrapolés + Diff. Relative* (%)
2008	< 65 ans	1 932 791	45 523	1 812 201 (-6,2%)	1 910 872 (-1,1%)
	65 ans et +	1 093 622	52 210	1 041 373 (-4,8%)	1 079 106 (-1,3%)
	Total	3 026 413	97 733	2 853 574 (-5,7%)	2 989 978 (-1,2%)
2009	< 65 ans	1 989 748	46 531	1 851 938 (-6,9%)	1 970 256 (-1,0%)
	65 ans et +	1 147 606	54 488	1 087 485 (-5,2%)	1 134 302 (-1,2%)
	Total	3 137 354	101 019	2 939 423 (-6,3%)	3 104 558 (-1,0%)

**Différence relative par rapport aux données de référence.*

Répartition des séjours par sexe

La Table 7 montre que les différences entre la répartition (homme/femme) estimée des séjours et la répartition nationale sont de l'ordre de 0,02% et 0,01% pour 2008 et 2009, respectivement, en utilisant les poids TCT. Statistiquement parlant, on ne peut pas dire qu'il y ait des différences dans la répartition hommes/femmes entre l'échantillon et les données TCT (valeur p ~ 0,9). Ceci n'est pas surprenant puisque les poids TCT ont été calculés post-stratification sur base des facteurs sexe et âge (réparti en catégorie de 5 ans). La Table 8 reprend le nombre de séjours extrapolés (fréquences pondérées) par sexe. En utilisant les poids AIM, nous observons qu'il y a une sous-estimation pouvant varier de 5% à 7% tandis qu'en utilisant les pois TCT, cette sous-estimation est de 1% du nombre total de séjours observés dans les données nationales TCT.

Table 7 – Répartition des séjours par sexe dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux

Année	Sexe	Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
		% séjours	% séjours [IC 95%]	Différence avec la Référence*	% séjours [IC 95%]	Différence avec la Référence*
2008	Femmes	54,5%	54,7% [54,4%;55,0%]	+0,2%	54,5% [54,1%;54,8%]	-0,02%
	Hommes	45,5%	45,3% [45,0%;45,6%]	-0,2%	45,5% [45,2%;45,9%]	+0,02%
	Total	100%	100%	p=0,25	100%	p=0,91
2009	Femmes	54,4%	54,7% [54,4%;55,0%]	+0,3%	54,4% [54,1%;54,7%]	-0,01%
	Hommes	45,6%	45,3% [45,0%;45,6%]	-0,3%	45,6% [45,3%;45,9%]	+0,01%
	Total	100%	100%	p=0,07	100%	p=0,94

**La valeur de p compare la distribution des séjours dans les catégories entre l'échantillon et la population de référence (sur base du Rao-Scott Chi square).*

Table 8 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par sexe

Année	Sexe	Référence (= TCT)	Echantillon	Echantillon (Poids AIM)	Echantillon (Poids TCT)
		Nb. Séjours TCT	Nb. Séjours EPS-TCT	Nb. Séjours extrapolés + Diff. Relative* (%)	Nb. Séjours extrapolés + Diff. Relative* (%)
2008	Femmes	1 648 914	52 851	1 560 309 (-5,4%)	1 628 479 (-1,2%)
	Hommes	1 377 499	44 882	1 293 265 (-6,1%)	1 361 498 (-1,2%)
	Total	3 026 413	97 733	2 853 574 (-5,7%)	2 989 978 (-1,2%)
2009	Femmes	1 706 371	54 692	1 607 625 (-5,8%)	1 688 159 (-1,1%)
	Hommes	1 430 983	46 327	1 331 798 (-6,9%)	1 416 399 (-1,0%)
	Total	3 137 354	101 019	2 939 423 (-6,3%)	3 104 558 (-1,0%)

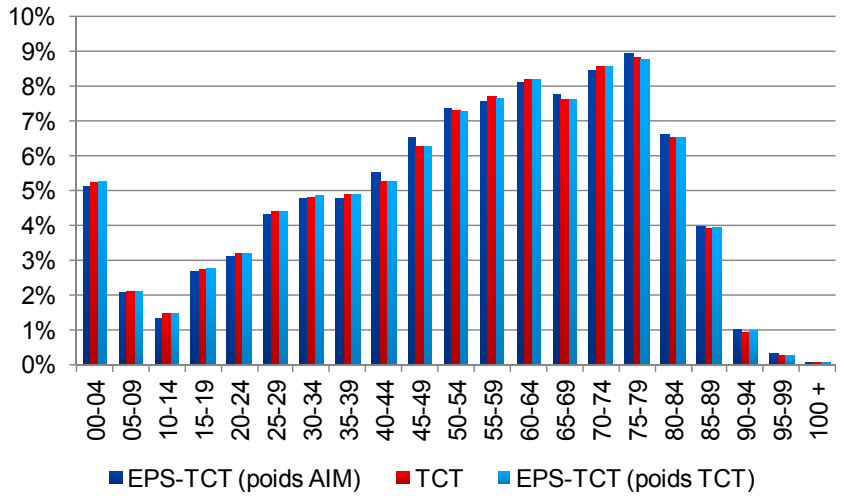
**Différence relative par rapport aux données de référence.*

L'échantillon est donc représentatif de la population pour le facteur sexe.

Répartition des séjours par âge

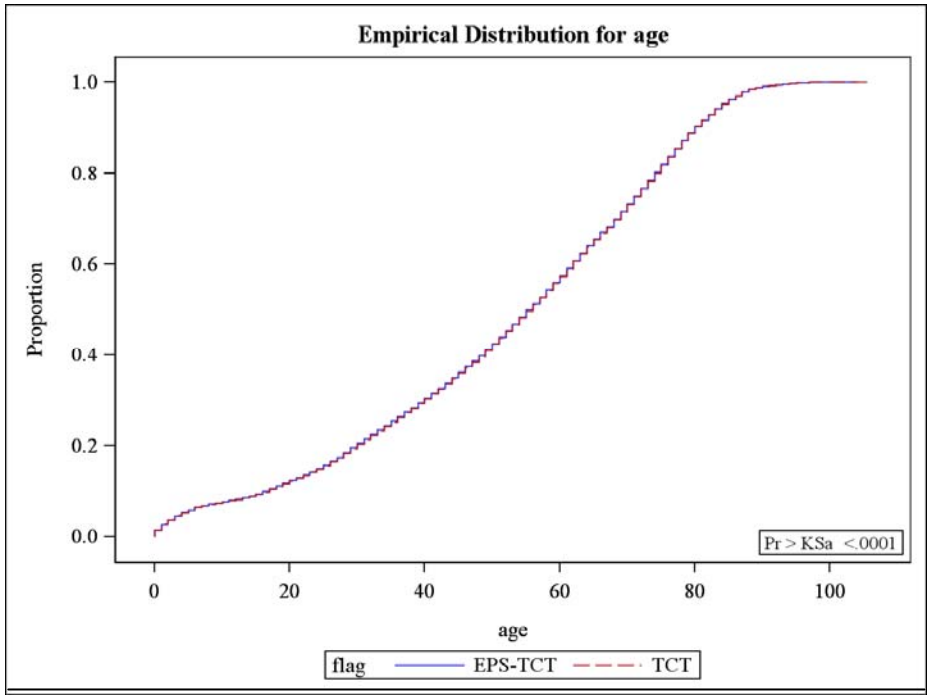
La répartition des séjours dans les différentes catégories d'âge est plus ou moins semblable à celle des données nationales, quelques soient les poids appliqués (différences maximale de 0,2%). Lorsqu'on analyse les différences entre les résultats obtenus avec les poids AIM et avec les poids TCT, on observe que la méthode utilisant les poids TCT colle parfaitement aux données TCT. A nouveau, comme pour le sexe, ceci est logique puisque les poids TCT ont été calculés sur base des répartitions dans les différentes classes d'âge (voir annexe 5.1.1. Table 3).

Figure 11 – Répartition des séjours par catégorie d'âge - 2009



La Figure 12 présente la fonction de répartition de l'âge dans les deux populations (échantillon et nationale). Les deux courbes se juxtaposent presque parfaitement. Cependant, le test de Kolmogorov qui compare les deux distributions est statistiquement significatif (p<0,0001). Sur un grand échantillon, la probabilité de détecter de petites différences est grande.

Figure 12 – Fonction de Distribution empirique (Age) : juxtaposition des courbes de l'échantillon EPS-TCT (poids TCT) et de la population TCT de référence - 2009



La Table 9 nous montre les différents paramètres de distribution pour l'âge. Ceux-ci sont très proches les uns des autres (moyenne, médiane, Q1-Q3, min-max). La différence entre l'estimation de l'âge moyen des patients ayant eu une hospitalisation est de 52 ans dans l'échantillon et dans les données nationales (différence de 0,04 an) et l'estimation de la médiane de 55 ans (différence de 0,1 an).

Table 9 – Paramètres de distribution de l'âge dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence

Année	Données	Moyenne (erreur standard)	Médiane [IC 95%]	Q1-Q3	Min-max	Valeur p*
2008	EPS-TCT	52,05 (0,005)	55,1 [55,1;55,2]	34,7-71,2	0-104	
	TCT	52,09 (0,001)	55,2	34,8-71,1	0-106	
	Différence	0,04	0,1			p<0,001
2009	EPS-TCT	52,34 (0,005)	55,5 [55,5;55,6]	35,1-71,3	0-105	
	TCT	52,38 (0,001)	55,6	35,2-71,3	0-106	
	Différence	0,04	0,1			p<0,001

**Kolmogorov Smirnov one-sample test.*

Répartition des séjours par statut social

Pour le statut social, nous observons que les estimations de répartition faites avec les données EPS-TCT sont très proches des résultats observés dans la population TCT (Table 10). Les différences entre estimations et valeurs observées dans les données nationales peuvent varier d'au plus 0,15%. D'un point de vue statistique, on ne peut pas dire qu'il y ait des différences de répartition entre les 2 populations (EPS-TCT et TCT). Le nombre de séjours en hôpital par statut social (Table 11) est sous-estimé lorsqu'on utilise les données de l'échantillon. Cette sous-estimation varie d'environ 5% si on utilise les poids AIM et de 1% en utilisant les poids TCT. Cela nous conforte dans le fait que les données de l'échantillon sont représentatives des données nationales par rapport au statut social.

Table 10 – Répartition des séjours par statut social dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux

Année	Statut social	Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
		Répartition	Répartition [IC 95%]	Différence avec la Référence*	Répartition [IC 95%]	Différence avec la Référence*
2008	Tarif préférentiel	22,6%	22,8% [22,5%;23,0%]	+0,15%	22,7% [22,4%;22,9%]	+0,05%
	Tarif non préférentiel	77,4%	77,2% [77,0%;77,5%]	-0,15%	77,3% [77,1%;77,6%]	-0,05%
	Manquant	<0,1%	<0,1% [<0,1%; <0,1%]	<0,01%	<0,1% [<0,1%; <0,1%]	<0,01%
	Total	100%	100%	p=0,56	100%	p=0,84
2009	Tarif préférentiel	23,1%	23,2% [23,0%;23,5%]	+0,10%	23,1% [22,9%;23,4%]	0,01%
	Tarif non préférentiel	76,9%	76,8% [76,5%;77,0%]	-0,10%	76,9% [76,6%;77,1%]	+0,01%
	Manquant	<0,1%	<0,1% [<0,1%; <0,1%]	<0,01%	<0,1% [<0,1%; <0,1%]	<0,01%
	Total	100%	100%	p=0,38	100%	p=0,50

**La valeur de p compare la distribution des séjours dans les catégories entre l'échantillon et la population de référence (sur base du Rao-Scott Chi square).*

Table 11 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par statut social

Année	Statut social	Référence (= TCT)	Echantillon	Echantillon (Poids AIM)	Echantillon (Poids TCT)
		Nb. Séjours TCT	Nb. Séjours EPS-TCT	Nb. Séjours extrapolés + Diff. Relative* (%)	Nb. Séjours extrapolés + Diff. Relative* (%)
2008	Tarif préférentiel	684 532	25 408	649 700 (-5,1%)	677 856 (-1,0%)
	Tarif non préférentiel	2 341 765	72 323	2 203 795 (-5,9%)	2 312 038 (-1,3%)
	Manquant	116	2	80 (-31,3%)	84 (-27,9%)
	Total	**3 026 413**	**97 733**	**2 853 574 (-5,7%)**	**2 989 978 (-1,2%)**
2009	Tarif préférentiel	726 160	26 611	683 263 (-5,9%)	718 298 (-1,1%)
	Tarif non préférentiel	2 411 072	74 407	2 256 119 (-6,4%)	2 386 215 (-1,0%)
	Manquant	122	1	40 (-67,1%)	45 (-63,1%)
	Total	**3 137 354**	**101 019**	**2 939 423 (-6,3%)**	**3 104 558 (-1,0%)**

**Différence relative par rapport aux données de référence.*

Répartition des séjours par localisation de domicile du patient

Lorsque nous nous intéressons à la répartition des séjours selon la région de domicile du patient, nous observons que les estimations faites en utilisant les données de l'EPS-TCT sont très proches des valeurs observées dans la population TCT (Table 12) avec des intervalles de confiance très étroits. Les différences entre les estimations de répartition et les répartitions observées dans les données nationales sont de moins de 0,5% pour 2008 et 2009. Statistiquement, les différences de répartition pour 2008 sont significatives au seuil de 5% (p=0,04). Donc, pour résumer pour 2008, si nous utilisons les données de l'échantillon, nous risquons de faire une erreur d'estimation de répartition de 0,30% pour l'une des régions et de <0,30% pour les autres. Cette différence est considérée statistiquement significative au seuil de 5% (la taille de l'échantillon est de 97 622 séjours et donc la variation très petite). On peut donc raisonnablement estimer que les estimations sur base des données de l'échantillon s'approchent fortement des valeurs de la population (TCT).

Le nombre de séjours en hôpital par région de domicile du patient (Table 13) est sous-estimé et la sous-estimation est la plus grande pour la région Bruxelloise (8% et 4% en utilisant respectivement les poids AIM et TCT).

Les résultats par province entre les répartitions des deux bases de données (EPS-TCT vs TCT) (voir annexe 5.1.1. Tables 5) sont semblables, le plus grand écart étant inférieur à 1%.

Table 12 – Répartition des séjours par région du domicile du patient dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux

		Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
Année	Région (patient)	Répartition	Répartition [IC 95%]	Différence avec la Référence*	Répartition [IC 95%]	Différence avec la Référence*
2008	Région Bruxelloise	7,6%	7,3% [7,2%; 7,5%]	-0,21%	7,3% [7,2%; 7,5%]	-0,21%
	Région Flamande	61,8%	62,1% [61,8%;62,4%]	+0,30%	62,1% [61,8%;62,5%]	+0,32%
	Région Wallonne	30,6%	30,5% [30,2%;30,8%]	-0,09%	30,5% [30,2%;30,8%]	-0,11%
	Total	100%	100%	p=0,04	100%	p=0,04
2009	Région Bruxelloise	7,8%	7,8% [7,6%; 7,9%]	-0,03%	7,8% [7,6%; 8,0%]	-0,02%
	Région Flamande	61,8%	61,7% [61,4%;62,0%]	-0,15%	61,7% [61,3%;62,0%]	-0,17%
	Région Wallonne	30,4%	30,5% [30,2%;30,8%]	+0,18%	30,6% [30,3%;30,9%]	+0,19%
	Total	100%	100%	p=0,51	100%	p=0,47

**La valeur de p compare la distribution des séjours dans les catégories entre l'échantillon et la population de référence (sur base du Rao-Scott Chi square).*

Table 13 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par région du domicile du patient

Année	Région (patient)	Référence (= TCT) Nb. Séjours TCT	Echantillon Nb. Séjours EPS-TCT	Echantillon (Poids AIM) Nb. Séjours extrapolés + Diff. Relative* (%)	Echantillon (Poids TCT) Nb. Séjours extrapolés + Diff. Relative* (%)
2008	Région Bruxelloise	228 179	7 161	209 235 (-8,3%)	219 250 (-3,9%)
	Région Flamande	1 867 968	60 684	1 770 029 (-5,2%)	1 855 265 (-0,7%)
	Région Wallonne	925 631	29 777	870 404 (-6,0%)	911 319 (-1,5%)
	Total	3 021 778	97 622	2 849 668 (-5,7%)	2 985 835 (-1,2%)
2009	Région Bruxelloise	244 423	7 825	228 091 (-6,7%)	241 343 (-1,3%)
	Région Flamande	1 936 627	62 370	1 810 094 (-6,5%)	1 910 947 (-1,3%)
	Région Wallonne	950 966	30 684	896 177 (-5,8%)	946 894 (-0,4%)
	Total	3 132 016	100 879	2 934 361 (-6,3%)	3 099 184 (-1,0%)

**Différence relative par rapport aux données de référence.*

Les fréquences pondérées sont présentées en Figure 13 et Figure 14, utilisant respectivement les poids AIM et les poids TCT (voir annexe 5.1.1. Table 6 pour plus de détails). On observe une différence relative (sous-estimation) pouvant aller jusqu'à 10% (province du Luxembourg) lorsqu'on utilise les poids AIM. En utilisant les poids TCT, il y a surestimation du nombre de séjours pour certaines provinces (Brabant Wallon, Flandre occidentale, Flandre orientale et Limbourg) de maximum 3,5% et sous-estimation pour les autres provinces de maximum 6%.

Les résultats par arrondissement sont présentés en Figure 15 pour l'année 2009 et plus en détail en annexe (5.1.1. Table 9). Les différences entre les répartitions des séjours des arrondissements dans les deux bases de données (EPS-TCT vs TCT) sont inférieures à 1% pour les 2 années. Par contre, pour les fréquences pondérées (poids TCT), on observe des différences relatives pouvant aller jusqu'à une sous-estimation de 27% (Ostende) par rapport aux données nationales. Ces données sont donc sensibles pour évaluer les fréquences de séjours par arrondissement. Il est à noter que la répartition des séjours par arrondissement varie d'une année à l'autre et que la différence observée à Turnhout en 2009 ne l'est pas en 2008.

Figure 13 – Fréquences EPS-TCT pondérées par province du domicile du patient (poids AIM) par rapport aux fréquences observées sur la population TCT de référence - 2009

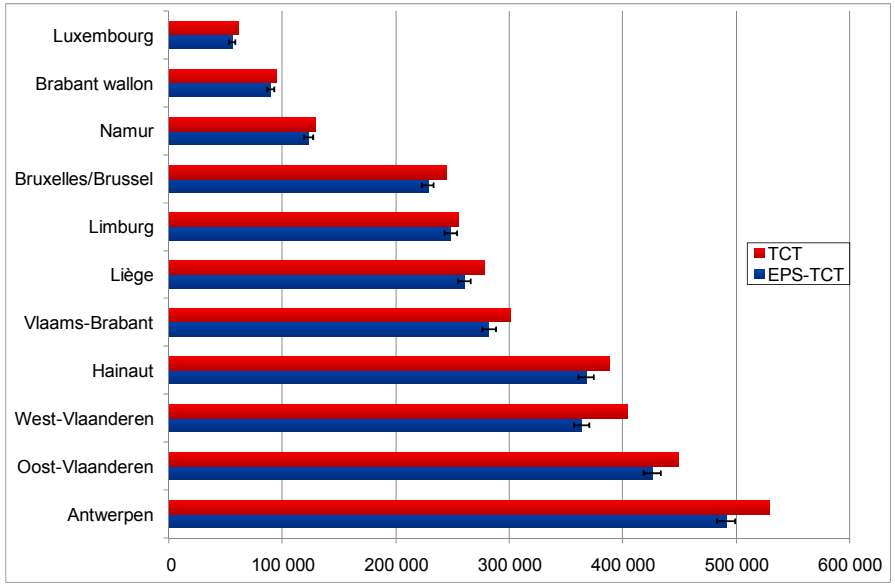


Figure 14 – Fréquences EPS-TCT pondérées par province du domicile du patient (poids TCT) par rapport aux fréquences observées sur la population TCT de référence - 2009

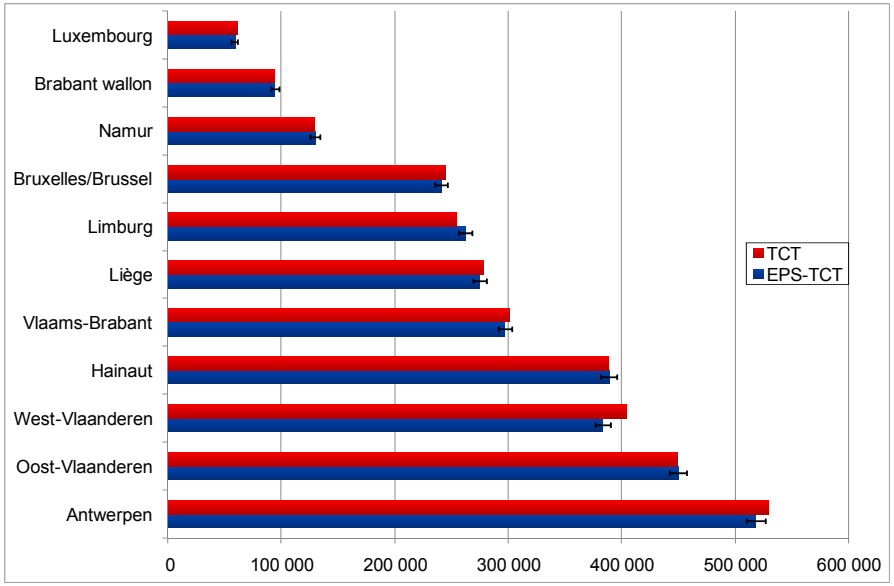
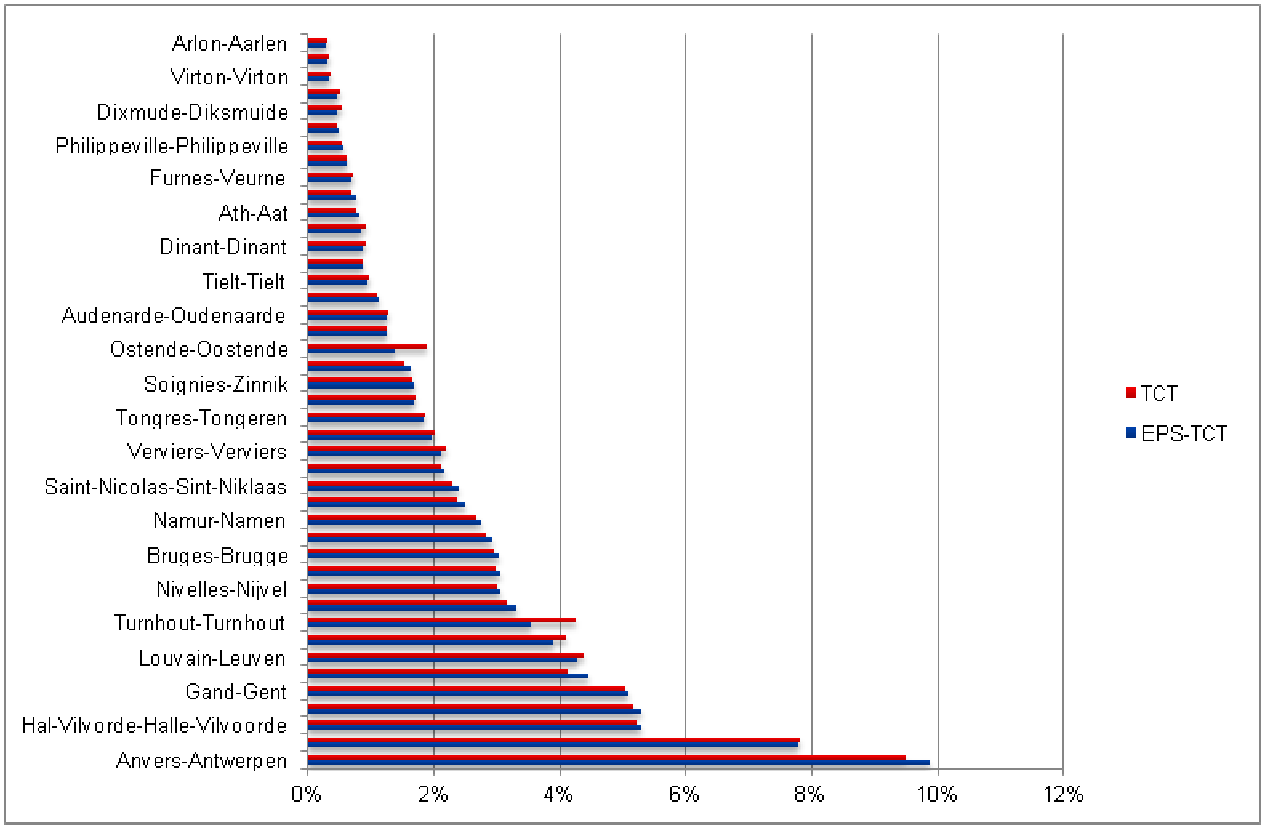


Figure 15 – Répartition des séjours dans l'échantillon EPS-TCT et dans la population TCT de référence par arrondissement du domicile du patient (poids TCT) – 2009



Répartition de la région et de la province de l'hôpital

Pour la répartition des séjours selon la région de l'hôpital, les résultats sont très semblables à ceux observés pour les région et province du domicile du patient. Les différences de répartition sont statistiquement non-significatives.

Table 14 – Répartition des séjours par région de l'hôpital dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux

		Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
Année	Région (hôpital)	Répartition	Répartition [IC 95%]	Différence avec la Référence*	Répartition [IC 95%]	Différence avec la Référence*
2008	Région Bruxelloise	12,2%	12,0% [11,8%;12,2%]	-0,18%	12,0% [11,8%;12,2%]	-0,17%
	Région Flamande	60,3%	60,5% [60,2%;60,9%]	+0,28%	60,6% [60,2%;60,9%]	+0,29%
	Région Wallonne	27,6%	27,5% [27,2%;27,8%]	-0,10%	27,4% [27,1%;27,7%]	-0,13%
	Total	100%	100%	p=0,17	100%	p=0,17
2009	Région Bruxelloise	12,5%	12,5% [12,3%;12,7%]	+0,01%	12,5% [12,3%;12,7%]	+0,03%
	Région Flamande	60,2%	60,0% [59,7%;60,4%]	-0,18%	60,0% [59,7%;60,3%]	-0,21%
	Région Wallonne	27,3%	27,5% [27,2%;27,8%]	+0,17%	27,5% [27,2%;27,8%]	+0,18%
	Total	100%	100%	p=0,49	100%	p=0,43

**La valeur de p compare la distribution des séjours dans les catégories entre l'échantillon et la population de référence (sur base du Rao-Scott Chi square).*

Table 15 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par région de l'hôpital

Année	Région (hôpital)	Référence (= TCT) Nb. Séjours TCT	Echantillon Nb. Séjours EPS-TCT	Echantillon (Poids AIM) Nb. Séjours extrapolés + Diff. Relative* (%)	Echantillon (Poids TCT) Nb. Séjours extrapolés + Diff. Relative* (%)
2008	Région Bruxelloise	368 507	11 461	342 375 (-7,1%)	359 119 (-2,5%)
	Région Flamande	1 823 688	59 261	1 727 445 (-5,3%)	1 810 458 (-0,7%)
	Région Wallonne	834 218	27 011	783 754 (-6,0%)	820 400 (-1,7%)
	Total	3 026 413	97 733	2 853 574 (-5,7%)	2 989 978 (-1,2%)
2009	Région Bruxelloise	391 295	12 341	366 798 (-6,3%)	388 163 (-0,8%)
	Région Flamande	1 889 154	60 873	1 764 727 (-6,6%)	1 862 939 (-1,4%)
	Région Wallonne	856 905	27 805	807 898 (-5,7%)	853 456 (-0,4%)
	Total	3 137 354	101 019	2 939 423 (-6,3%)	3 104 558 (-1,0%)

**Différence relative par rapport aux données de référence.*

Figure 16 – Fréquences EPS-TCT pondérées par province de l'hôpital (poids AIM) par rapport aux fréquences observées sur la population TCT de référence - 2009

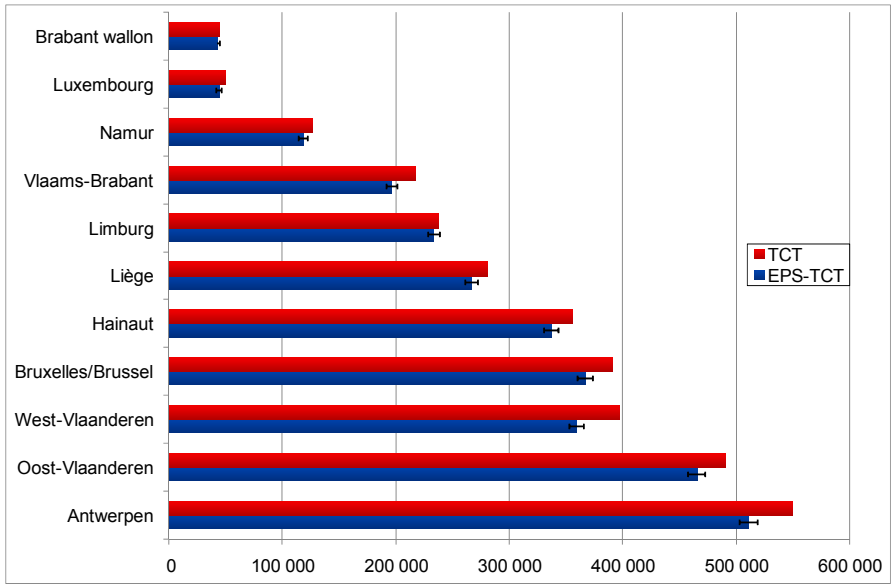
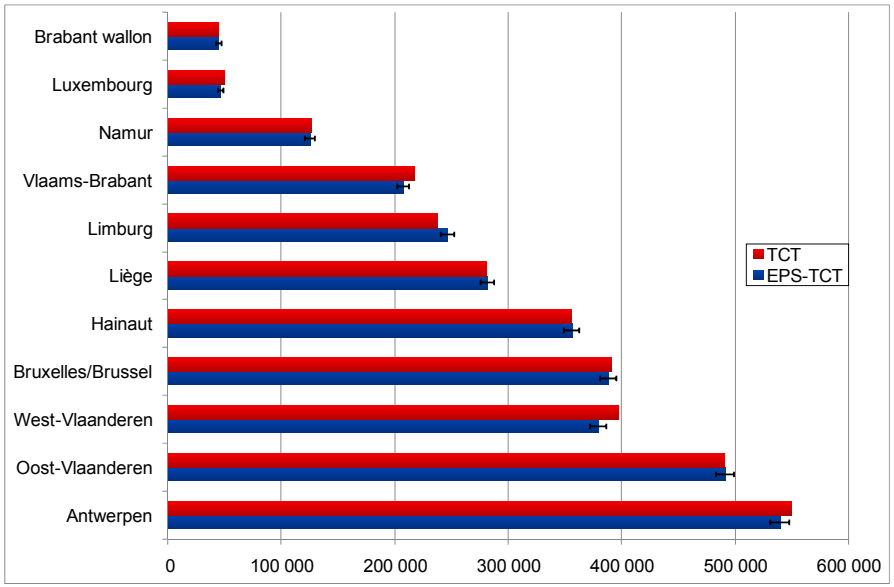


Figure 17 – Fréquences EPS-TCT pondérées par province de l'hôpital (poids TCT) par rapport aux fréquences observées sur la population TCT de référence - 2009



Répartition par type d'hospitalisation

La Table 16 montre que les différences entre la répartition estimée des séjours par type d'hospitalisation et la répartition nationale sont de l'ordre de 0,74% et 0,65% pour 2008 et 2009, respectivement, en utilisant les poids TCT. Comme décrit à la section 4.2.2.3, les hospitalisations de jour ont plus de chance d'être couplées que les séjours classique. Ceci implique un biais dans les analyses lorsqu'elles portent sur des paramètres pouvant différer d'un type d'hospitalisation à l'autre (par exemple les remboursements). Dans ce cas, il est recommandé d'effectuer les analyses par type d'hospitalisation. Cette différence de couplage amène donc des différences de répartition de séjours entre les différents types d'hospitalisation. Globalement, cette différence n'est pas importante mais pourrait l'être au niveau de sous-groupes.

La Table 17 reprend les fréquences pondérées. En utilisant les poids AIM, nous observons qu'il ya a une sous-estimation dans le nombre de séjours pouvant varier 5% à 7% tandis qu'en utilisant les pois TCT, cette sous-estimation est de 3% du nombre total de séjours observés dans les données nationales.

Table 16 – Répartition des séjours par type d'hospitalisation dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux

		Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
Année	Type d'hospitalisation	Répartition	Répartition [IC 95%]	Différence avec la Référence*	Répartition [IC 95%]	Différence avec la Référence*
2008	Classique	54,5%	53,7% [53,4%;54,0%]	-0,79%	53,8% [53,4%;54,1%]	-0,74%
	Hospitalisation de jour	45,5%	46,3% [46,0%;46,6%]	+0,79%	46,2% [45,9%;46,6%]	+0,74%
	Total	100%	100%	p<0,0001	100%	p<0,0001
2009	Classique	54,1%	53,4% [53,1%;53,7%]	-0,68%	53,4% [53,1%;53,7%]	-0,65%
	Hospitalisation de jour	45,9%	46,6% [46,3%;46,9%]	+0,68%	46,6% [46,3%;46,9%]	+0,65%
	Total	100%	100%	p<0,0001	100%	P=0,0001

**Valeur de p compare la distribution des séjours dans les catégories entre l'échantillon et la population de référence (Basé sur Rao-Scott Chi square).*

Table 17 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par type d'hospitalisation

		Référence (= TCT)	Echantillon	Echantillon (Poids AIM)	Echantillon (Poids TCT)
Année	Type d'hospitalisation	Nb. Séjours TCT	Nb. Séjours EPS-TCT	Nb. Séjours extrapolés + Diff. Relative* (%)	Nb. Séjours extrapolés + Diff. Relative* (%)
2008	Classique	1 649 268	53 569	1 532 587 (-7,1%)	1 607 173 (-2,6%)
	Hospitalisation de jour	1 377 145	44 164	1 320 987 (-4,1%)	1 382 805 (0,4%)
	Total	3 026 413	97 733	2 853 574 (-5,7%)	2 989 978 (-1,2%)
2009	Classique	1 696 098	54 980	1 569 062 (-7,5%)	1 658 246 (-2,2%)
	Hospitalisation de jour	1 441 256	46 039	1 370 360 (-4,9%)	1 446 312 (0,4%)
	Total	3 137 354	101 019	2 939 423 (-6.3%)	3 104 558 (-1,0%)

**Différence relative par rapport aux données de référence.*

Répartition par montant total des remboursements

Les différents paramètres de distribution pour le montant de remboursement par séjour sont présentés dans la Table 18. Quelques petites variations sont observables. Il est à noter que les données TCT comprenaient des montants très infimes (minimum : 0,37). Seuls les montants nuls ou négatifs ont été éliminés avant analyse. Donc, nous avons dans les minima, des montants de 0€ (en arrondi).

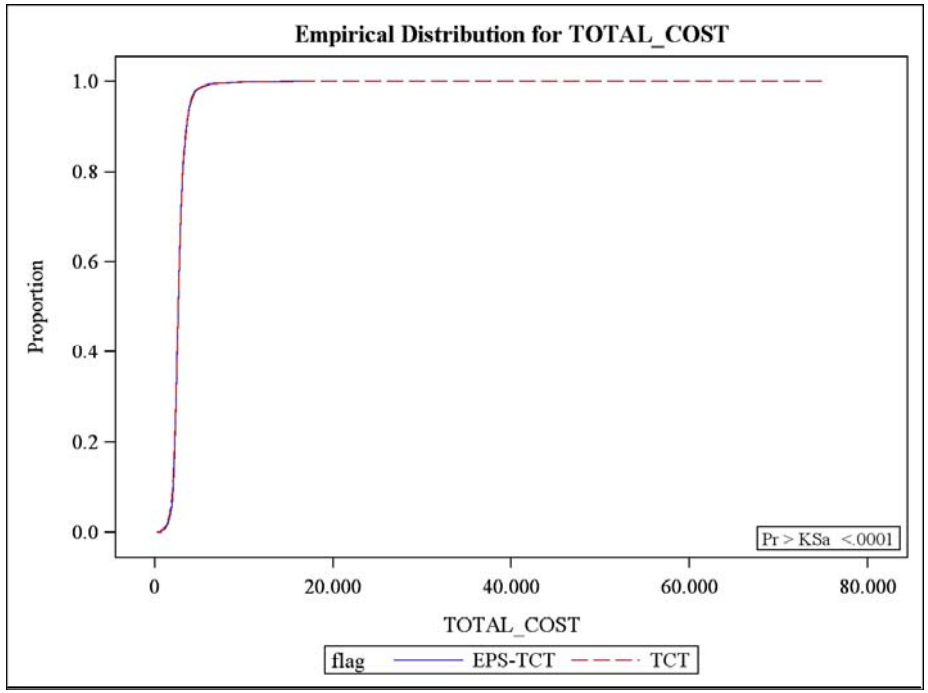
Pour les maxima, nous observons de grandes différences entre l'échantillon et les données nationales. En effet, la probabilité qu'un séjour outlier ayant un grand montant total soit sélectionné dans l'échantillon est infime. Il sera donc important de prendre cette observation en considération lors des analyses effectuées sur cette variable. La Figure 18 présente la distribution empirique des montants dans l'échantillon et dans les données nationales. Les deux courbes se juxtaposent presque parfaitement. Le test de Kolmogorov qui compare les deux distributions est statistiquement significatif ($p<0,0001$) car la probabilité de détecter de petites différences sur un grand échantillon est grande. Ces résultats n'ont probablement pas d'interprétation clinique à ce niveau (toutes pathologies confondues) car toutes les hospitalisations sont incluses sans distinction. Cependant, cela confirme le principe selon lequel si un échantillon est tiré de façon aléatoire et inclut un certain nombre d'observations, alors l'estimation de distribution de n'importe quelle variable est extrapolable à la population de référence.

Table 18 – Paramètres de distribution du montant total des remboursements (euros) dans l'échantillon EPS-TCT et dans la population TCT de référence

Année	Données	Moyenne (erreur standard)	Médiane [IC 95%]	Q1-Q3	Min-max	Valeur p*
2008	EPS-TCT	2 660,80 (13,46)	1 338,7 [1 325,3;1 352,1]	645,1-2 872,3	42-159 061	
	TCT	2 744,05 (2,63)	1 360,1	650,4-2 932,5	0-785 323	
	Différence	**83,25**	**21,3**			**p<0,0001**
2009	EPS-TCT	2 874,21 (15,59)	1 352,0 [1 338,5;1 365,6]	662,9-3 032,5	4-202 304	
	TCT	2 977,52 (3,04)	1 366,5	668,4-3 083,0	0-490 069	
	Différence	**103,31**	**14,5**			**p<0,0001**

**Kolmogorov Smirnov one-sample test.*

Figure 18 – Distribution empirique du montant total des remboursements : juxtaposition des courbes de l'échantillon EPS-TCT (poids TCT) et de la population TCT de référence – 2009

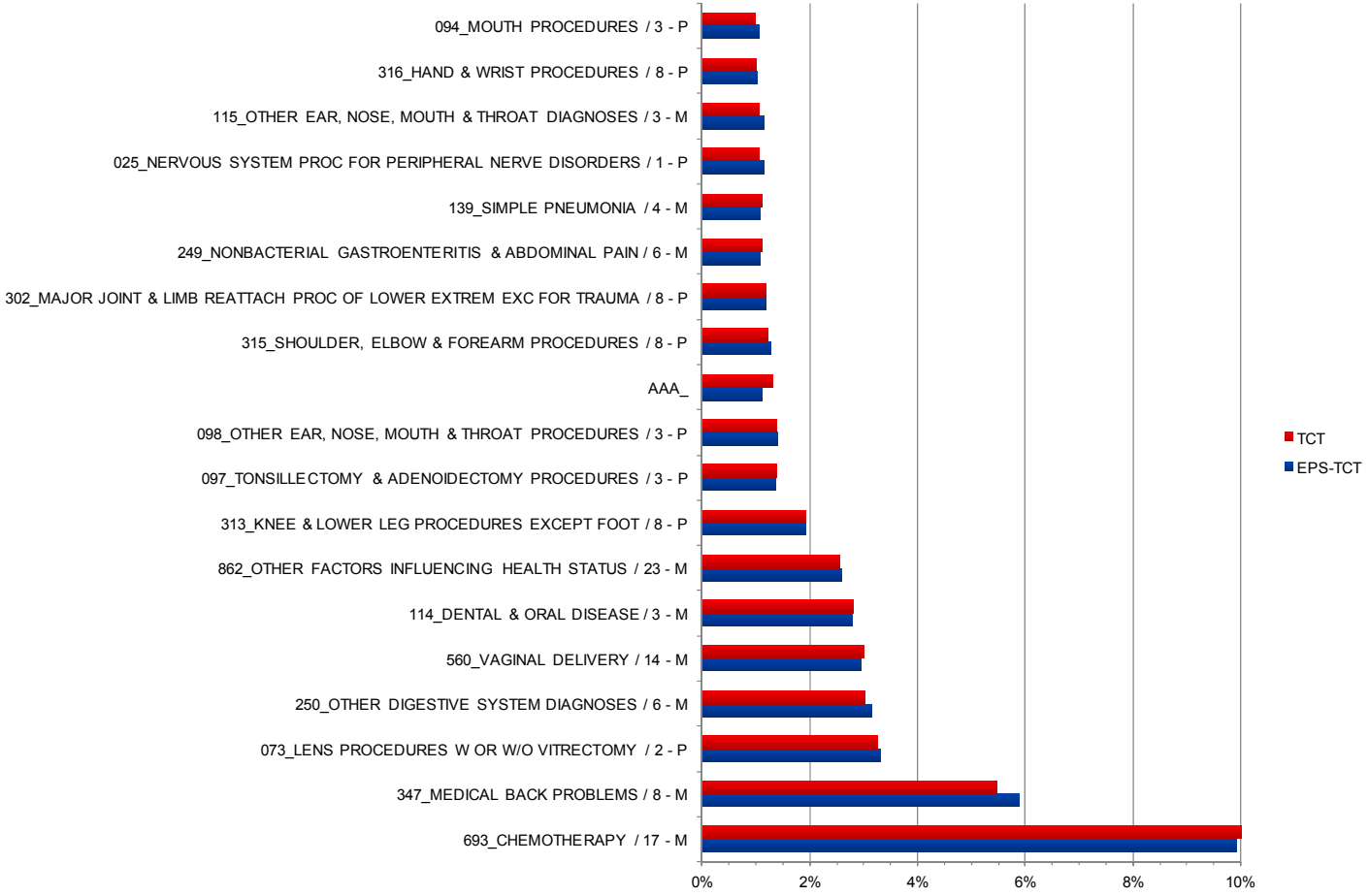


Répartition par APR-DRG et par diagnostic principal

La Figure 19 présente la répartition des APR-DRG dans l'échantillon (EPS-TCT) et dans les données nationales (TCT). Pour le graphe, nous nous sommes limités a posteriori aux APR-DRG comprenant plus de 1% de tous les séjours.

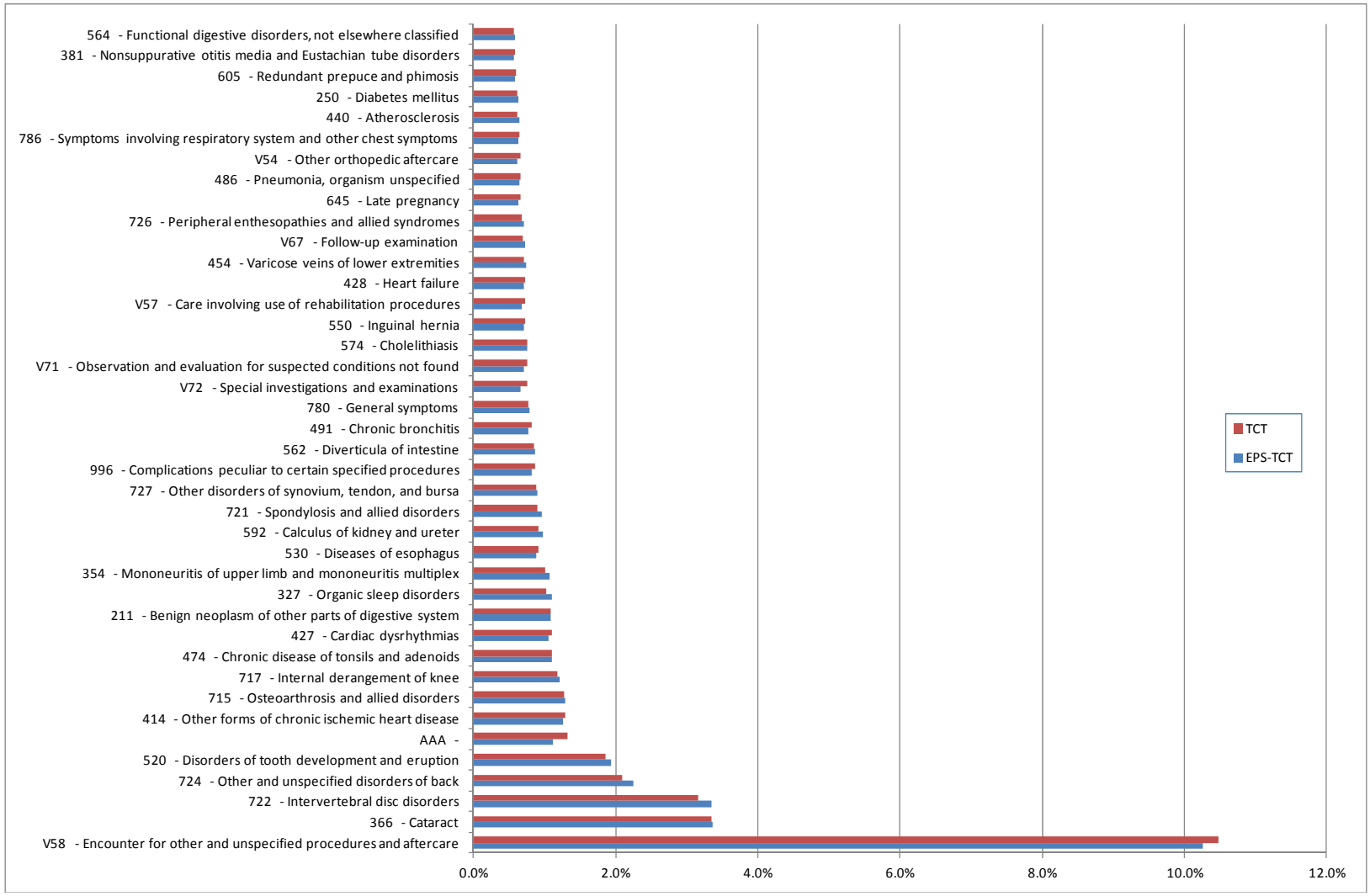
La Figure 20 présente la répartition des diagnostics principaux dans l'échantillon (EPS-TCT) et dans les données nationales (TCT). Pour le graphe, nous nous sommes limités a posteriori aux diagnostics comprenant plus de 0,5% de tous les séjours.

Figure 19 – Répartition des APR-DRG (APR-DRG représentant plus de 0,5% des séjours) dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - 2009



**La catégorie AAA_ reprend les séjours entièrement psychiatriques dans les hôpitaux généraux (groupement particulier à la Belgique).*

Figure 20 – Répartition des diagnostics principaux (diagnostics représentant plus de 0,5% des séjours) dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - 2009



**La catégorie AAA_ reprend les séjours entièrement psychiatriques dans les hôpitaux généraux (groupement particulier à la Belgique).*

Pour chaque variable, la table récapitulative (Table 19) reprend les écarts maximaux observés, parmi toutes les catégories de la variable, entre l'estimation de la répartition dans l'échantillon et la répartition dans les données nationales. La valeur de p (testant si les distributions sont semblables) est également présentée. D'autre part, les différences relatives maximales sont présentées pour pouvoir évaluer l'erreur faite en estimant le nombre de séjours à partir des données de l'échantillon.

On observe que pour les APR-DRG, une surestimation maximale de 300% est détectée (pour 2008 dans l'APR-DRG 613 – Nouveau-nés, 1500-1999 gr. avec infections congénitales ou périnatales et pour 2009 dans l'APR-DRG 660 – Agranulocytose et autre neutropénie). Pour l'APR-DRG 613, ceci s'explique par le fait que dans la population, le nombre de séjours est relativement petit et que par hasard, une proportion non négligeable a été échantillonnée (échantillon de 6 séjours versus 67 dans le national). Le même phénomène est observé encore plus nettement pour les diagnostics principaux. Cela peut se produire pour toute variable dont une des catégories a un faible nombre de séjours au niveau national. En revanche, en ce qui concerne l'APR-DRG 660, il s'agit d'une modification dans l'algorithme d'attribution des APR-DRG. Jusqu'en 2008 plusieurs diagnostics principaux pouvaient mener à cet APR-DRG. En 2009, seule la neutropénie congénitale y mène (ICD-9-CM 288.01), les autres diagnostics menant alors à un autre APR-DRG. Le nombre de séjours TCT de l'APR-DRG 660 est donc tombé de 2763 à 5 au niveau national et de 102 à 1 dans l'échantillon EPS-TCT. Ceci démontre la nécessité impérative de bien connaitre les données avant toute interprétation de résultats et de tenir compte des modifications potentielles de la classification clinique lors d'études longitudinales.

Table 19 – Table récapitulative des tests de représentativité par variable – au niveau global

		Echantillon (poids AIM)			Echantillon (poids TCT)		
Année	Variable (nombre de catégories)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel max (nombre séjours)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel. max. avec la référence (nombre séjours)
2008	Sexe (2)	0,2%	p=0,25	-5,4%	0,02%	p=0,91	-1,2%
	Catégorie d'âge (21)	0,4%	p<0,001	-17%	0,04%	p=1,00	-1,6%
	Statut social (2)	0,15%	p=0,56	-5,9%	0,05%	p=0,84	-1,3%
	Type d'hospitalisation (2)	0,79%	p<0,001	-7,1%	0,74%	p<0,001	-2,6%
	Région (patient) (3)	0,30%	p=0,04	-8,3%	0,32%	p=0,04	-3,9%
	Province (patient) (11)	0,52%	p<0,001	-10,6%	0,52%	p<0,001	-6,2%
	Arrondissement (patient) (43)	0,53%	p<0,001	-24,4%	0,53%	p<0,001	-20,7%
	Région (hôpital) (3)	0,28%	p=0,17	-7,1%	0,29%	p=0,17	-2,5%
	Province (hôpital) (11)	0,65%	p<0,001	-9,9%	0,64%	p<0,001	-5,7%
	Arrondissement (hôpital) (39)	0,64%	p<0,001	-23,3%	0,64%	p<0,001	-20,7%
	Diagnostic Principal (935)	0,26%	**	**+799,1%**	0,26%	**	**+778,8%**

Année	Variable (nombre de catégories)	Echantillon (poids AIM)			Echantillon (poids TCT)		
		Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel max (nombre séjours)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel. max. avec la référence (nombre séjours)
	APR-DRG (360)	0,26%	**	**+301,2%**	0,24%	**	**+295,6%**
	Montant total des remboursements*	€15*	p<0,001		€21*	p<0,001	
2009	Sexe (2)	0,3%	p=0,07	-6,9%	0,01%	p=0,94	-1,1%
	Catégorie d'âge (21)	0,26%	p<0,001	-17%	0,04%	p=1,00	-1,5%
	Statut social (2)	0,10%	p=0,38	-6,4%	0,01%	p=0,50	-1,1%
	Type d'hospitalisation (2)	0,68%	p<0,001	7,5%	0,35%	p<0,001	2,2%
	Région (patient) (3)	0,18%	p=0,51	-6,7%	0,19%	p=0,47	-1,3%
	Province (patient) (11)	0,51%	p<0,001	-10,0%	0,52%	p<0,001	-5,1%
	Arrondissement (patient) (43)	0,72%	p<0,001	-31,1%	0,72%	p<0,001	-27,4%
	Région (hôpital) (3)	0,18%	p=0,49	-6,6%	0,21%	p=0,43	-1,4%
	Province (hôpital) (11)	0,42%	p<0,001	-12,3%	0,44%	p<0,001	-7,3%
	Arrondissement (hôpital) (38)	0,64%	p<0,001	-23,6%	0,64%	p<0,001	-19,3%
	Diagnostic Principal (955)	0,21%	**	**+1235%**	0,25%	**	**+1151%**
	APR-DRG (360)	0,69%	**	**+301,4%**	0,37%	**	**+312,9%**
	Montant total des remboursements	€18*	p<0,001		€14*	p<0,001	

** Différence entre les montants médians.*
*** Le nombre de séjours est nul dans l'EPS-TCT et non nul dans TCT pour une des classes de la variable. La méthode (SURVEYFREQ) ne contenant pas de facteur de correction pour les cellules vides, le Chi-square de Rao-Scott n'a pas pu être estimé.*

En conclusion :

En considérant tous les séjours hospitaliers, tous APR-DRG confondus, nous avons montré que l'on pouvait utiliser les données de l'échantillon EPS-TCT couplé pour décrire globalement les données d'hospitalisation (TCT). Cela rejoint la théorie de l'échantillonnage selon laquelle, tout échantillon tiré d'une population de façon aléatoire est représentatif de la population (selon certaines conditions sur le nombre d'unités échantillonnées).

Dans notre cas, le nombre de séjours (tous APR-DRG confondus) est assez large pour soutenir ces conclusions.

5.2.2 Résultats au niveau APR-DRG

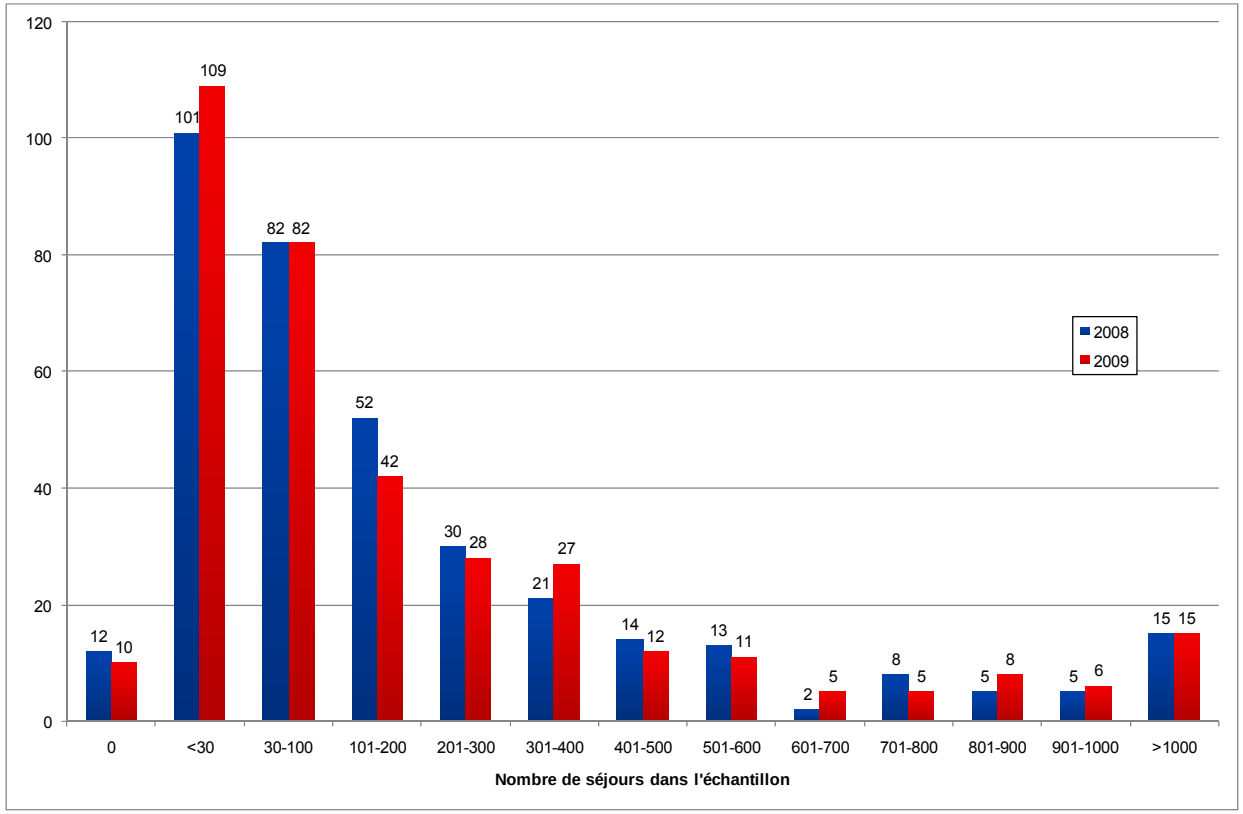
Analyses sur les séjours par APR-DRG

Dans le cadre d'études, nous sommes souvent amenés à étudier non pas l'entièreté des données hospitalières mais uniquement celles reliées à un certain APR-DRG (ou groupe d'APR-DRG). Etant donné que l'échantillon n'est pas un échantillon stratifié selon ce facteur, il est très probable d'observer des disparités entre l'échantillon et les données nationales (au niveau APR-DRG) et surtout pour les APR-DRG avec un petit nombre de séjours. Pour décrire ces résultats, nous avons repris certaines valeurs clefs pour présenter les différences observées entre les estimations et les valeurs issues des données nationales :

- Différences relatives par APR-DRG : elles représentent les écarts possibles entre les fréquences pondérées et les fréquences réelles issues des données nationales (TCT)
- Écart maximal entre les taux de répartition estimés (EPS-TCT) et les taux de répartition observés (TCT) par variable. Cet écart correspond à la différence entre la répartition estimée (EPS-TCT) et celle observée (TCT) dans la catégorie où cette différence est la plus grande (par exemple la catégorie des 50-54 ans).

La Figure 21 présente le nombre d'APR-DRG par nombre de séjours dans l'échantillon. Nous observons que pour la plupart des APR-DRG, le nombre de séjours tiré de la population TCT est de moins de 100 séjours. Etant donné que la précision des estimations est liée de façon directe à la taille de l'échantillon, nous pouvons déjà conclure que les analyses spécifiques pour les APR-DRG incluant un petit nombre de séjours seront moins fiables que ceux incluant un grand nombre de séjours (l'incertitude sera plus grande).

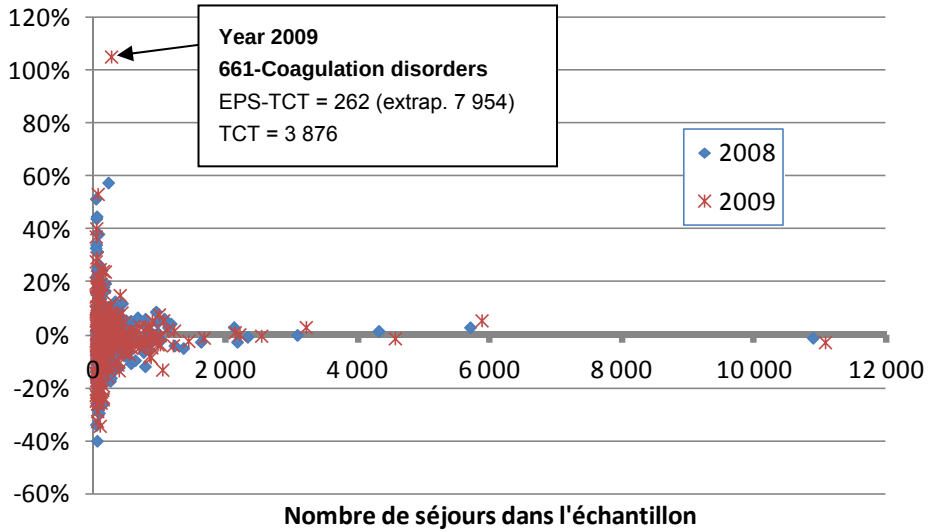
Figure 21 – Nombre d'APR-DRG en fonction du nombre de séjours dans l'échantillon EPS-TCT

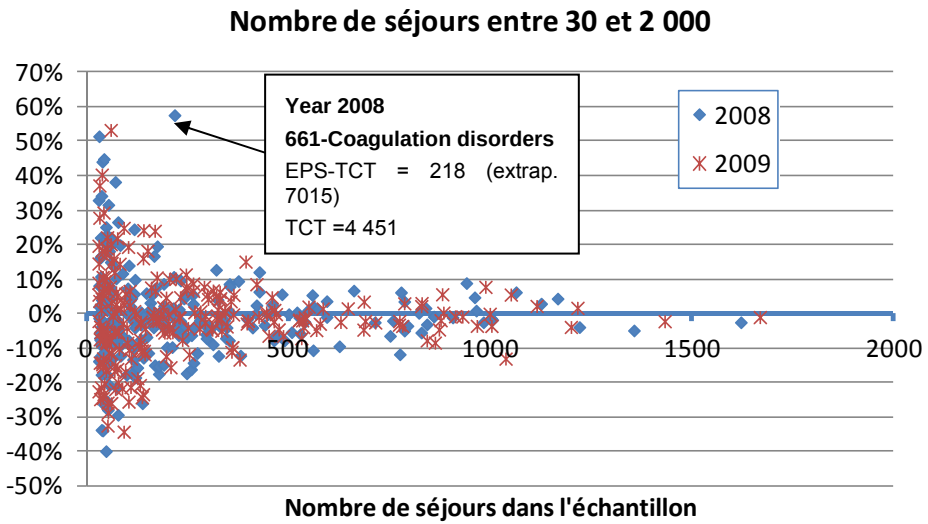


Différences relatives par APR-DRG

La Figure 22 de gauche représente les différences relatives par APR-DRG. Celle de droite, zoom de celle de gauche, focalise sur les APR-DRG dont le nombre de séjours dans l'échantillon est compris entre 30 et 2 000. Comme attendu, plus la taille de l'échantillon (ou de l'APR-DRG) est petite, plus cette différence est grande et variable. Nous avons sélectionné les APR-DRG dont le nombre de séjours dans l'échantillon est d'au moins 30 séjours.

Figure 22 – Différence relative entre les répartitions par APR-DRG dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence

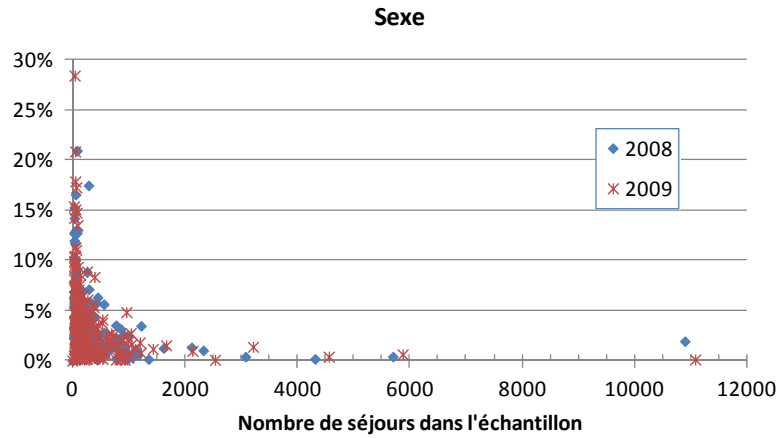


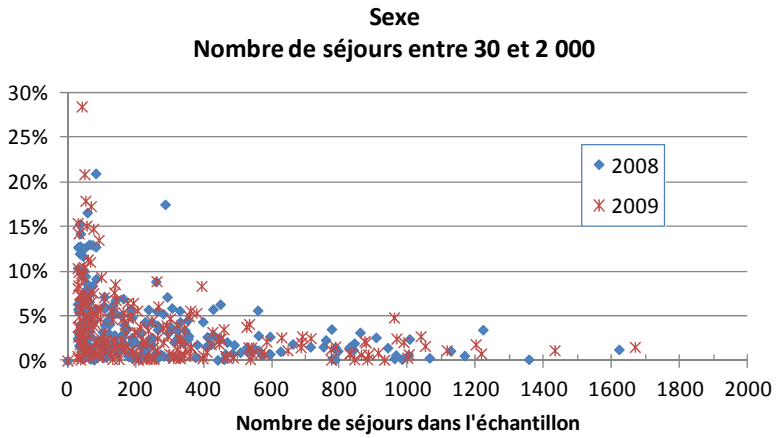


Répartition par sexe (poids TCT)

Nous observons (Figure 23) que sur les 220 APR-DRG ayant un nombre de séjours supérieur à 30 et non spécifique à un sexe (comme par exemple : accouchements ou procédures sur le pénis), il y a 166 APR-DRG (75%) dont les différences entre répartition sont inférieures à 5%. Les différences répartitions estimées (EPS-TCT) et observées (TCT) peuvent s'expliquer par le faible nombre de séjours repris dans l'échantillon (par exemple : *229 - autres procédures sur le système digestif* : 35 séjours (51.2%) de femmes et 33 (48.8%) d'hommes pour l'échantillon versus 1 323 (59%) de femmes et 918 (41%) d'hommes dans les données nationales) mais pas uniquement (par exemple : *141 - asthme et bronchiolites* : 91 séjours (35%) de femmes et 171 (65%) d'hommes pour l'échantillon versus 1 753 (45%) de femmes et 2 123 (55%) d'hommes pour les données nationales (TCT)). Il est donc recommandé d'investiguer plus spécifiquement les résultats de l'APR-DRG avant de travailler sur l'échantillon pour en déduire des résultats globaux.

Figure 23 – Différence maximale entre les répartitions par sexe dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence

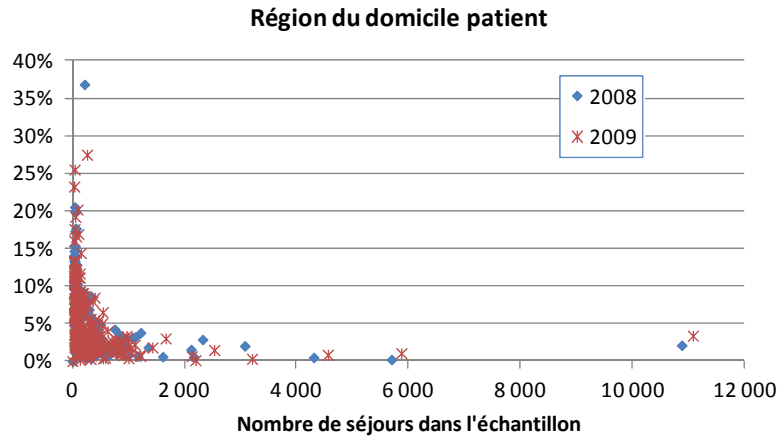




Autres facteurs

Les résultats observés sont les mêmes pour les autres facteurs (voir figures ci-dessous) que pour la répartition par sexe.

Figure 24 – Différence maximale entre les répartitions par région et par province du domicile du patient dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence



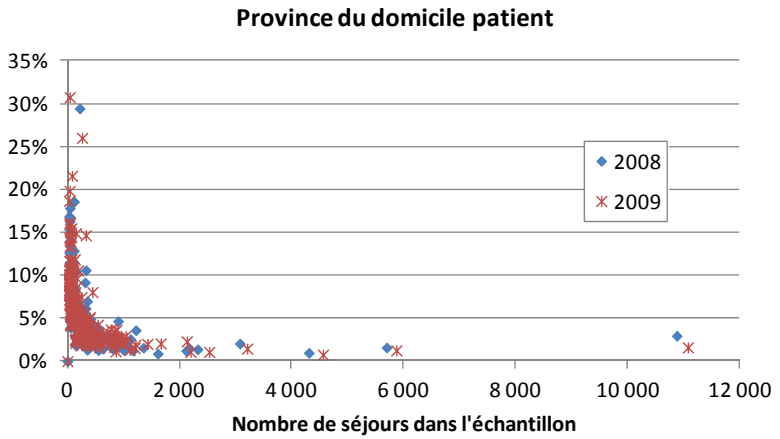
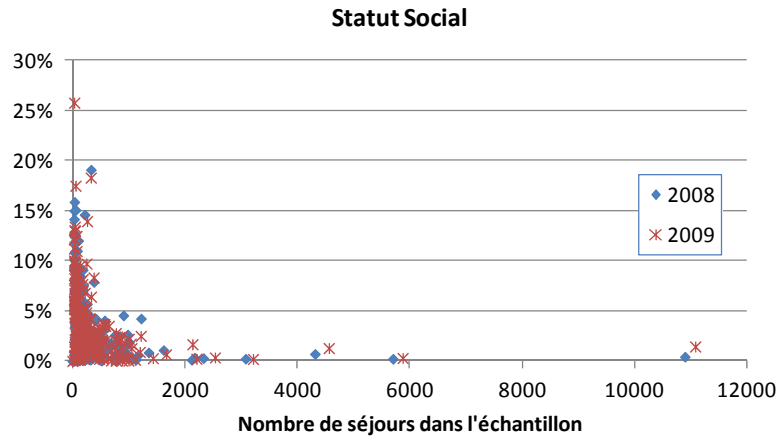
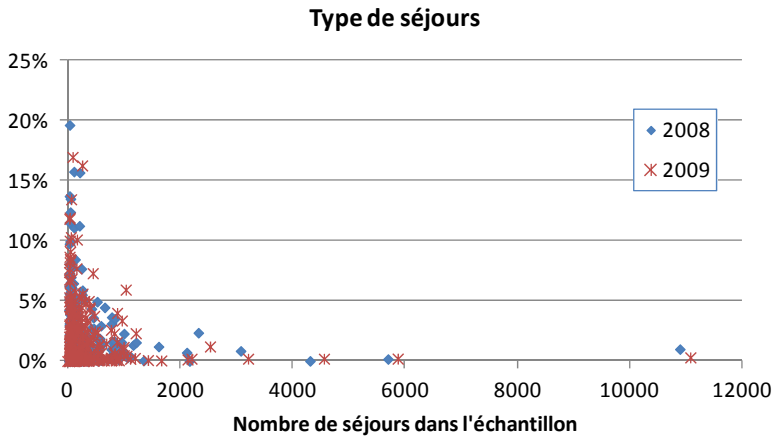


Figure 25 – Différence maximale entre les répartitions par statut social et par type d'hospitalisation dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence

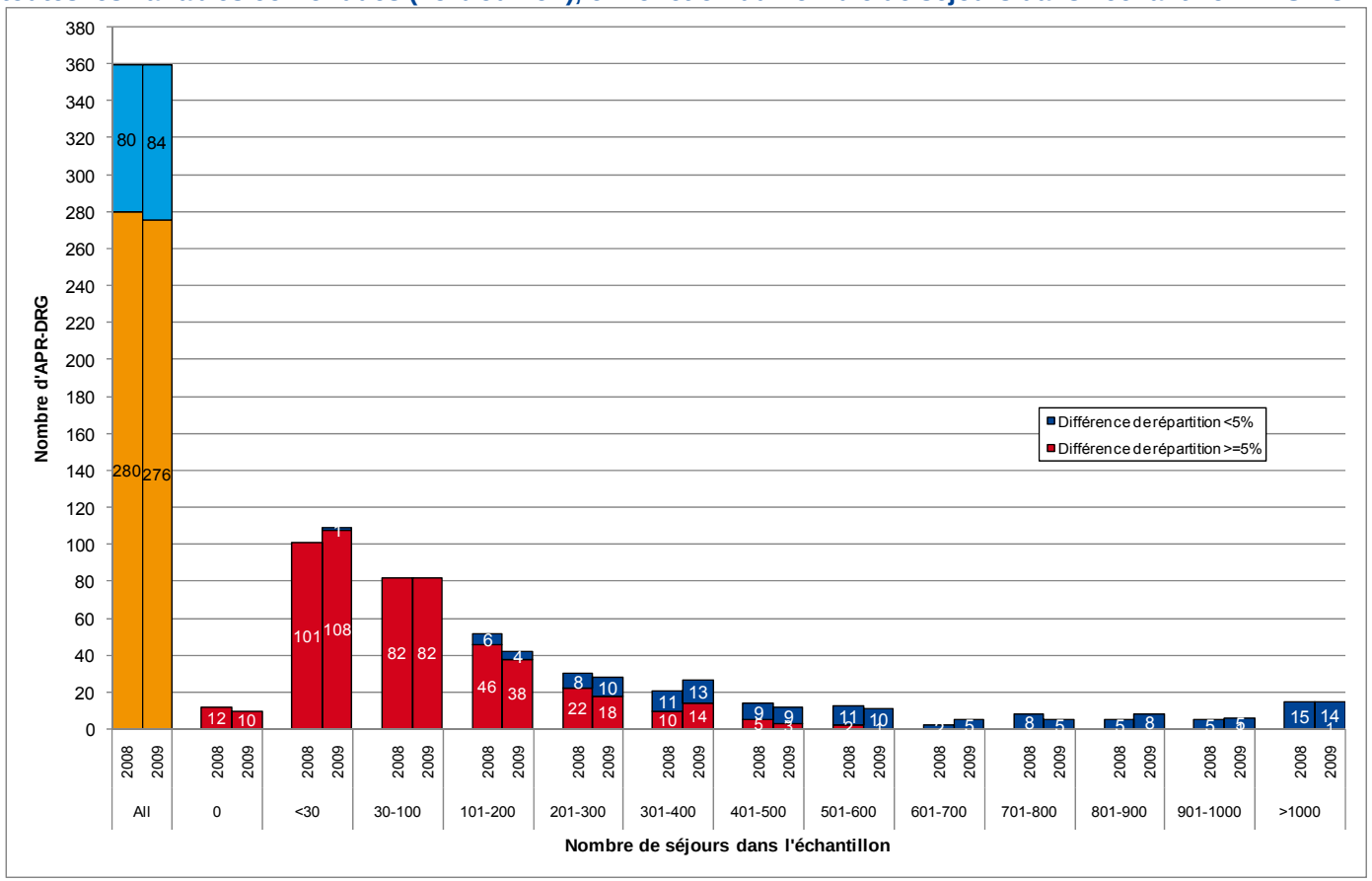




En résumé, nous avons pour 2008 et 2009, respectivement, 80 et 84 APR-DRG dont la différence maximale entre les répartitions estimées dans l'échantillon EPS-TCT et celles observées dans les données nationales (TCT) pour l'ensemble des variables catégorielles n'excède pas 5% (marge d'erreur fixée arbitrairement). Ces APR-DRG totalisent néanmoins 76% des séjours.

La Figure 26 présente le nombre d'APR-DRG ayant une différence en termes de répartition pour toutes les variables en dessous et au-dessus de 5% en fonction du nombre de séjours hospitaliers repris dans l'EPS-TCT.

Figure 26 – Nombre d'APR-DRG classés selon la différence maximale entre les répartitions (EPS-TCT versus population TCT de référence) pour toutes les variables confondues (<5% ou non), en fonction du nombre de séjours dans l'échantillon EPS-TCT



Pour les analyses par APR-DRG et niveau de sévérité , il y a environ :

- 3% des APR-DRG pour lesquels aucun séjour n'a été sélectionné dans l'échantillon (3,3% (12/360) en 2008 et 2,8% (10/360) en 2009)
- 32% des APR-DRG pour lesquels un des niveaux de sévérité n'est pas représenté dans l'EPS-TCT alors qu'il est présent dans les données nationales (33,3% (120/360) en 2008 et 31,7% (114/360) en 2009.

Conclusion pour les analyses par APR-DRG

Comme la méthode d'échantillonnage ne prend pas en compte le facteur APR-DRG, il n'y a pas représentativité systématique et il faut donc analyser les résultats au cas par cas. Quelque soit l'analyse, les trois principes suivants restent valables :

1. Taille de l'échantillon : plus l'échantillon est petit, plus le risque de non-représentativité est grand.
2. Nombre de niveaux de la variable d'intérêt : plus le nombre de niveaux est grand, plus les fréquences sont faibles et moins la représentativité est démontrable.
3. Résultats pour tous les facteurs : la représentativité doit être observée sur l'ensemble des facteurs utiles à l'analyse, car elle pourrait être atteinte fortuitement pour un des facteurs.

Comme dans tout analyse d'échantillon, les analyses de sous-groupes ne peuvent être effectuées que sous certaines conditions et leurs résultats peuvent être dus au hasard. Il était prévu d'analyser les résultats de représentativité pour chaque APR-DRG et niveau de sévérité. Cependant, les résultats pour chaque variable au niveau APR-DRG et niveau de sévérité ne peuvent être synthétisés mais doivent être analysés individuellement par APR-DRG.

5.2.3 Exemple d'analyses par APR-DRG

Sur avis du médecin expert de l'équipe de recherche, les deux 2 APR-DRG comprenaient un premier APR-DRG médical à priori homogène d'un point de vue clinique et un second APR-DRG, chirurgical et plutôt hétérogène[s] :

1. Les accouchements par voie vaginales (APR-DRG « 560 »)
2. Les interventions des membres inférieurs et genoux, excepté le pied (APR-DRG « 313 »).

5.2.3.1 APR-DRG 560 Accouchement par voie vaginale.

En 2008, il y a 92 137 séjours dans les données TCT nationales reliés à cet APR-DRG dont 2 171 repris dans l'échantillon EPS-TCT. Pour 2009, nous avons 94 182 séjours au niveau national dont 2 207 dans l'échantillon.

Lorsqu'on étudie la répartition des séjours pour accouchement par voie vaginale par catégorie d'âge (Table 20), les estimations sur base de l'échantillon sont très proches des résultats sur la population de référence TCT (erreur de maximum 1% par rapport à la référence). Les tests comparant les deux distributions de l'âge est non-significatif (p>0.5). L'échantillon et les données nationales ne diffèrent donc pas du point de vue de la répartition des séjours par âge. Utiliser les fréquences pondérées (Table 21) à la place des fréquences observables dans les données TCT amènerait à sous-estimer ou surestimer le nombre de séjours au sein de chaque catégorie d'âge, la surévaluation pouvant aller jusqu'à 700% par rapport au national (catégorie des 50-54 ans). En effet, il n'y a eu que 5 accouchements par voie vaginale en 2008 pour les 50-54 ans dont 1 est repris dans l'échantillon. En prenant une pondération de plus ou moins 40, cela nous mène à une estimation de 40 accouchements (au lieu de 5).

En ce qui concerne le niveau de sévérité (Table 22 et Table 23), les résultats sont assez satisfaisants du point de vue de la représentativité des données de l'échantillon.

[s] Cet APR-DRG a d'ailleurs été divisé en sous-groupes dans le système des montants de référence.[38]

Il est important de considérer tous les résultats globalisés (Table 26) afin de pouvoir décider si, en toute connaissance de cause, on veut utiliser les données de l'échantillon pour cet APR-DRG.

Répartition des séjours par âge

Table 20 – Répartition des séjours par catégorie dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux d'âge - APR-DRG 560 Accouchement par voie vaginale

		Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
Année	Catégorie d'âge (ans)	% séjours	% séjours [IC 95%]	Différence avec la Référence*	% séjours [IC 95%]	Différence avec la Référence*
2008	10-14	<0,1%				
	15-19	2,0%	1,8% [1,3%; 2,4%]	-0,18%	1,9% [1,3%; 2,4%]	-0,15%
	20-24	14,0%	13,4% [11,9%;14,8%]	-0,63%	13,7% [12,2%;15,1%]	-0,33%
	25-29	35,3%	35,7% [33,7%;37,8%]	+0,45%	34,9% [32,9%;36,9%]	-0,39%
	30-34	32,4%	32,6% [30,7%;34,6%]	+0,22%	32,8% [30,8%;34,8%]	+0,39%
	35-39	13,6%	14,0% [12,6%;15,5%]	+0,47%	14,5% [13,0%;16,0%]	+0,88%
	40-44	2,6%	2,2% [1,6%; 2,8%]	-0,39%	2,1% [1,5%; 2,7%]	-0,47%
	45-49	0,1%	0,1% [<0,1%; 0,3%]	+0,03%	0,1% [<0,1%; 0,3%]	+0,02%
	50-54	<0,1%	<0,1% [<0,1%; 0,1%]	+0,04%	<0,1% [<0,1%; 0,1%]	+0,04%
		100%	100%	p=0,65	100%	p=0,53
2009	10-14	<0,1%	.	.	.	.,
	15-19	2,0%	2,3% [1,6%; 2,9%]	+0,31%	2,3% [1,7%; 3,0%]	+0,37%
	20-24	14,0%	14,5% [13,0%;15,9%]	+0,46%	15,1% [13,6%;16,6%]	+1,11%
	25-29	35,1%	34,2% [32,2%;36,2%]	-0,90%	34,7% [32,7%;36,7%]	-0,42%
	30-34	32,7%	33,1% [31,1%;35,0%]	+0,42%	32,2% [30,3%;34,2%]	-0,44%
	35-39	13,5%	12,9% [11,5%;14,3%]	-0,57%	12,8% [11,4%;14,2%]	-0,63%
	40-44	2,7%	2,9% [2,2%; 3,6%]	+0,24%	2,6% [2,0%; 3,3%]	-0,02%
	45-49	0,1%	0,2% [<0,1%; 0,4%]	+0,05%	0,2% [<0,1%; 0,3%]	+0,04%
	50-54	<0,1%				
	Total	100%	100%	p=0,78	100%	p=0,61

Table 21 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par catégorie d'âge - APR-DRG 560 Accouchement par voie vaginale

Année	Catégorie d'âge (ans)	Référence (= TCT) Nb. Séjours TCT	Echantillon Nb. Séjours EPS-TCT	Echantillon (Poids AIM) Nb. Séjours extrapolés + Diff. Relative* (%)	Echantillon (Poids TCT) Nb. Séjours extrapolés + Diff. Relative* (%)
2008	10-14	7			
	15-19	1 851	39	1 564 (-15,5%)	1 672 (-9,7%)
	20-24	12 901	291	11 460 (-11,2%)	12 278 (-4,8%)
	25-29	32 513	773	30 627 (-5,8%)	31 349 (-3,6%)
	30-34	29 874	715	27 976 (-6,4%)	29 477 (-1,3%)
	35-39	12 507	302	12 031 (-3,8%)	12 983 (+3,8%)
	40-44	2 377	47	1 877 (-21,1%)	1 897 (-20,2%)
	45-49	102	3	119 (+17,1%)	122 (+19,2%)
	50-54	5	1	40 (+697,3%)	42 (+736,9%)
	Total	92 137	2 171	85 693 (-7,0%)	89 820 (-2,5%)
2009	10-14	5	.	.	.
	15-19	1 850	50	1 982 (+7,1%)	2 206 (+19,3%)
	20-24	13 195	318	12 631 (-4,3%)	14 285 (+8,3%)
	25-29	33 044	755	29 836 (-9,7%)	32 756 (-0,9%)
	30-34	30 769	735	28 875 (-6,2%)	30 454 (-1,0%)
	35-39	12 687	281	11 255 (-11,3%)	12 135 (-4,4%)
	40-44	2 504	64	2 533 (+1,2%)	2 490 (-0,6%)
	45-49	127	4	160 (+25,7%)	160 (+26,3%)
	50-54	1			
	Total	94 182	2 207	87 272 (-7,3%)	94 487 (+0,3%)

Répartition par niveau de sévérité

Table 22 – Répartition des séjours par niveau de sévérité dans l'échantillon EPS-TCT et dans la population TCT de référence et différence entre les deux - APR-DRG 560 Accouchement par voie vaginale

		Référence (= TCT)	Echantillon (Poids AIM)		Echantillon (Poids TCT)	
Année	Niveau de sévérité	Répartition	Répartition [IC 95%]	Différence avec la Référence*	Répartition [IC 95%]	Différence avec la Référence*
2008	Mineur (SOI 1)	52,7%	53,9% [51,8%;56,0%]	+1,14%	53,9% [51,8%;56,0%]	+1,13%
	Modéré (SOI 2)	40,0%	38,8% [36,8%;40,9%]	-1,15%	38,8% [36,8%;40,9%]	-1,16%
	Majeur (SOI 3)	7,2%	7,2% [6,1%; 8,3%]	+0,03%	7,2% [6,2%; 8,3%]	+0,04%
	Extrême (SOI 4)	0,1%	<0,1% [<0,1%; 0,1%]	-0,02%	<0,1% [<0,1%; 0,1%]	-0,02%
	Total	100%	100%	p=0,73	100%	p=0,73
2009	Mineur (SOI 1)	51,2%	51,7% [49,6%;53,8%]	+0,56%	51,8% [49,7%;53,9%]	+0,64%
	Modéré (SOI 2)	41,5%	41,1% [39,1%;43,2%]	-0,39%	41,0% [39,0%;43,1%]	-0,46%
	Majeur (SOI 3)	7,3%	7,1% [6,0%; 8,2%]	-0,16%	7,1% [6,0%; 8,2%]	-0,17%
	Extrême (SOI 4)	0,1%	<0,1% [<0,1%; 0,1%]	-0,01%	<0,1% [<0,1%; 0,1%]	-0,01%
	Total	100%	100%	P=0,93	100%	p=0,94

**La valeur de p compare la distribution des séjours dans les catégories entre l'échantillon et la population de référence (sur base du Rao-Scott Chi square).*

Table 23 – Nombre de séjours (population TCT de référence, échantillon EPS-TCT, extrapolés) et différences relatives par niveau de sévérité - APR-DRG 560 Accouchement par voie vaginale

		Référence (= TCT)	Echantillon	Echantillon (Poids AIM)	Echantillon (Poids TCT)
Année	Niveau de sévérité	Nb. Séjours TCT	Nb. Séjours EPS-TCT	Nb. Séjours extrapolés + Diff. Relative* (%)	Nb. Séjours extrapolés + Diff. Relative* (%)
2008	Mineur (SOI 1)	48 602	1170	46 177 (-5,0%)	48 396 (-0,4%)
	Modéré (SOI 2)	36 840	843	33 277 (-9,7%)	34 877 (-5,3%)
	Majeur (SOI 3)	6 636	157	6 199 (-6,6%)	6 507 (-1,9%)
	Extrême (SOI 4)	59	1	40 (-32,3%)	40 (-31,6%)
	Total	92 137	2 171	85 694 (-7,0%)	89 820 (-2,5%)
2009	Mineur (SOI 1)	48192	1142	45 143 (-6,3%)	48 955 (+1,6%)
	Modéré (SOI 2)	39091	907	35 880 (-8,2%)	38 780 (-0,8%)
	Majeur (SOI 3)	6850	157	6 210 (-9,4%)	6 708 (-2,1%)
	Extrême (SOI 4)	49	1	40 (-19,4%)	43 (-11,5%)
	Total	94 182	2 207	87 272 (-7,3%)	94 487 (+0,3%)

**Différence relative par rapport aux données de référence.*

Montant total des remboursements (en Euros)

La Table 24 montre les différents paramètres de distribution du montant total des remboursements et les différences avec les données nationales. Si le test de Kolmogorov est significatif ($p<0,0001$), le test de la médiane est quant à lui non-significatif ($p=0,39$). La Table 25 et la Figure 27 montrent les résultats par niveau de sévérité. Il s'avère que les différences entre les résultats obtenus avec les données EPS-TCT et TCT sont petites pour tous les niveaux de sévérité sauf pour le niveau « extrême ». Ceci s'explique par le fait, que pour ce niveau, il n'y a qu'un seul séjour sélectionné dans l'échantillon sur les 59 ou 49 séjours que compte la population TCT (respectivement pour l'année 2008 et 2009).

Table 24 – Paramètres de distribution du montant total des remboursements (euros) dans l'échantillon EPS-TCT et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale

Année	Données	Moyenne (erreur standard)	Médiane [IC 95%]	Q1-Q3	Min-max	Valeur p*
2008	EPS-TCT	2 782,17 (19,622)	2 650,7 [2 624,6;2 676,7]	2 343,2-3 043,6	222-17 572	
	TCT	2 780,65 (3,671)	2 653,9	2 336,4-3 033,7	222-74 998	
	Différence	1,53	-3,3			p<0,0001
2009	EPS-TCT	2 962,94 (25,977)	2 816,9 [2 786,6;2 847,2]	2 477,0-3 205,8	700-35 051	
	TCT	2 944,31 (3,693)	2 812,7	2 478,0-3 205,7	239-82 649	
	Différence	18,63	4,2			p<0,0001

**Kolmogorov Smirnov one-sample test.*

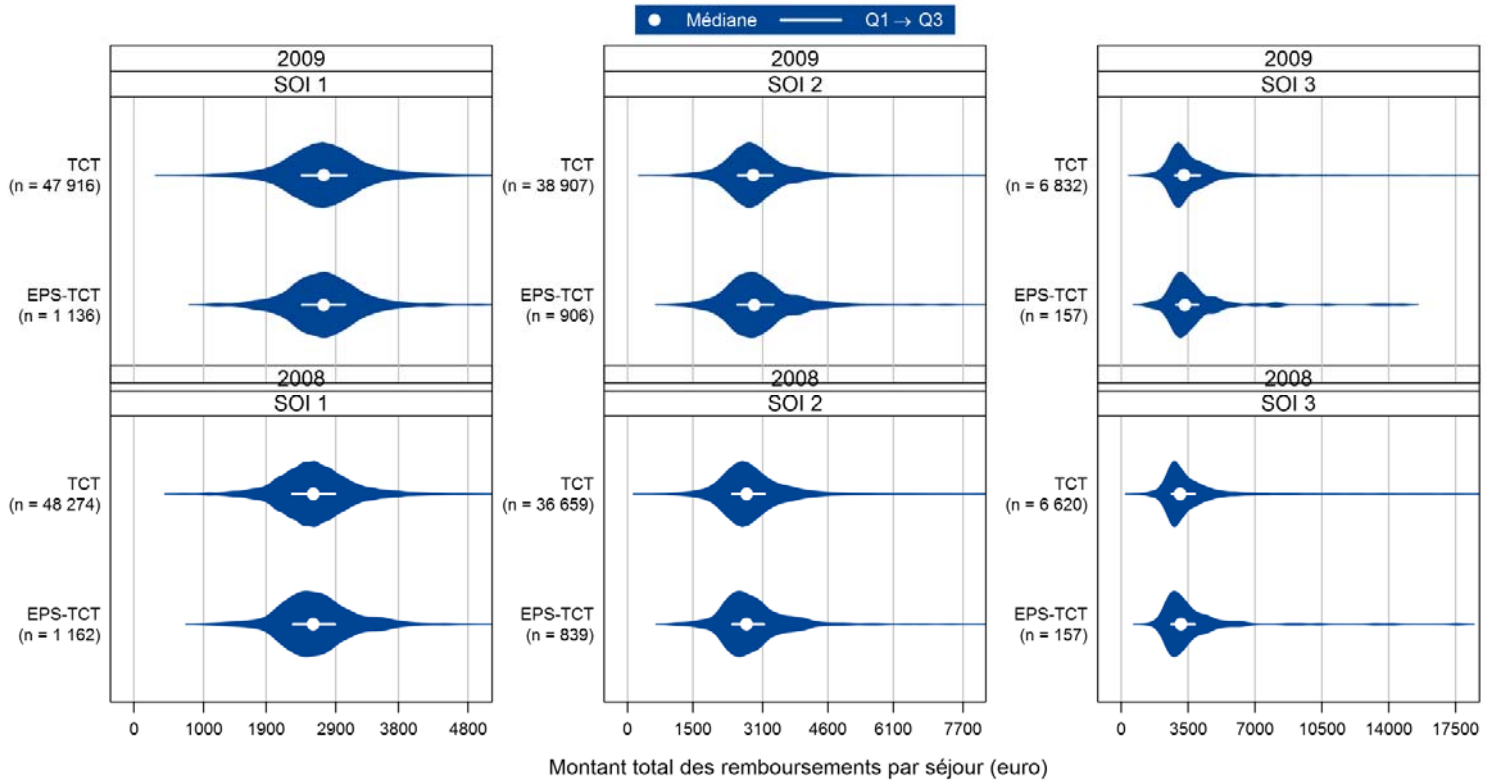
Table 25 – Paramètres de distribution du montant total des remboursements (euros) par niveau de sévérité dans l'échantillon EPS-TCT et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale

Année	Niveau de Sévérité	Données	Moyenne (erreur standard)	Médiane [IC 95%]	Q1-Q3	Min-max	Valeur p*
2008	Mineur (SOI 1)	EPS-TCT (n=1170)	2 613,27 (17,00)	2 573,5 [2 534,4;2 612,7]	2 280,7-2 881,9	609-8 529	
		TCT (n=48 602)	2 597,71 (2,73)	2572,9	2 268,7-2 884,6	222-20 242	
		Différence	15,56	0,6			p<0,001
	Modéré (SOI 2)	EPS-TCT (n=843)	2 853,04 (28,91)	2 725,8 [2 670,9;2 780,8]	2 396,7-3 130,7	222-12 623	
		TCT (n=36 840)	2 864,22 (5,80)	2731,2	2 395,6-3 147,6	222-74 998	
		Différence	-11,18	-5,3			p<0,001
	Majeur (SOI 3)	EPS-TCT (n=157)	3 662,11 (166,02)	3 128,1 [2 944,7;3 311,6]	2 629,6-3 855,3	1547-17 572	
		TCT (n=6 636)	3 626,75 (30,81)	3078,9	2 649,3-3 851,6	637-46 067	
		Différence	35,36	49,2			p<0,001
	Extrême (SOI 4)	EPS-TCT (n=1)	2 199,80	2 199,8	2 199,8-2 199,8	2 200-2 200	

Année	Niveau de Sévérité	Données	Moyenne (erreur standard)	Médiane [IC 95%]	Q1-Q3	Min-max	Valeur p*
		TCT (n=59)	6 129,15 (762,92)	4 848,1	3 620,4-6 474,2	2 200-44 313	
		Différence	-3 929,35	-2 648,3			p<0,001
2009	Mineur (SOI 1)	EPS-TCT (n=1142)	2 743,31 (17,93)	2 722,6 [2 697,9;2 747,3]	2 416,3-3 030,0	700-8 562	
		TCT (n=48 602)	2 751,82 (3,25)	2 723,4	2 406,0-3 051,8	239-72 419	
		Différence	-8,51	-0,8			p<0,001
	Modéré (SOI 2)	EPS-TCT (n=907)	3 068,80 (46,92)	2 905,9 [2 858,7;2 953,1]	2 522,0-3 343,1	800-35 051	
		TCT (n=39 091)	3 025,07 (5,49)	2 879,9	2 530,0-3 320,6	244-42 482	
		Différence	43,73	26,0			p<0,001
	Majeur (SOI 3)	EPS-TCT (n=157)	3 846,42 (149,40)	3 328,4 [3 157,3;3 499,5]	2 877,5-4 029,7	1467-14 700	
		TCT (n=6 850)	3 803,58 (29,59)	3 275,2	2 816,5-4 119,6	731-82 649	
		Différence	42,83	53,2			p<0,001
	Extrême (SOI 4)	EPS-TCT (n=1)	19 566,00	19 566,3	19 566,3-19 566,3	19 566-19 566	
		TCT (n=49)	7 713,75 (678,22)	5 683,1	4 442,9-9 315,9	1 609-26 661	
		Différence	1 1852,25	13 883,1			p<0,001

**Kolmogorov Smirnov one-sample test.*

Figure 27 – Distribution du montant total de remboursement par niveau de sévérité (hospitalisations classiques) dans l'échantillon EPS-TCT (poids TCT) et dans la population de référence TCT - APR-DRG 560 Accouchement par voie vaginale



Répartition par diagnostic principal

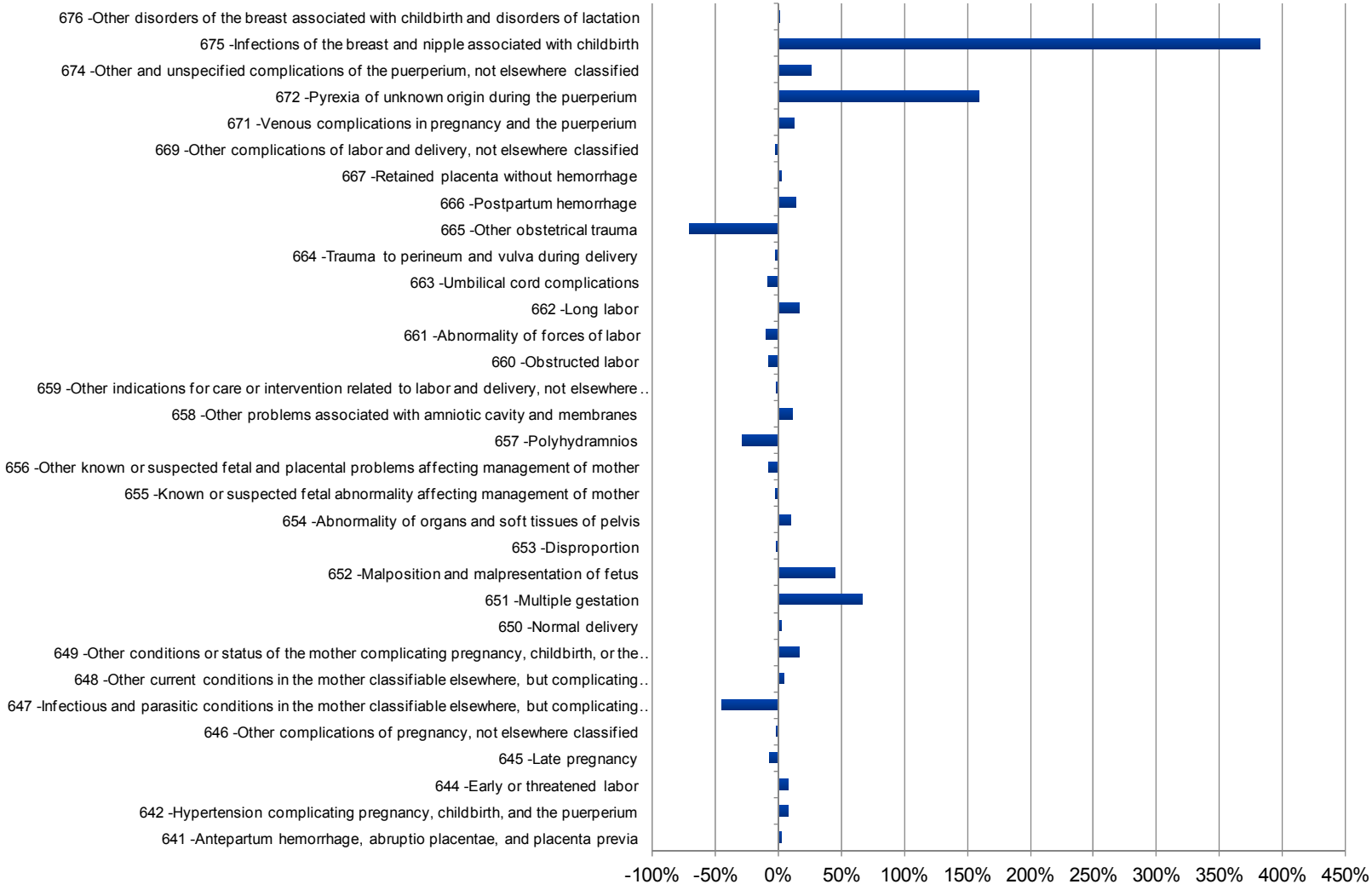
La Figure 28 nous montre que la répartition par diagnostics principaux dans l'EPS-TCT est semblable à celle dans la population TCT. Les différences relatives sont quant à elles très variables (Figure 29). Par exemple, on observe une différence relative de 382% pour le diagnostic principal « 675 -Infections of the breast and nipple associated with childbirth » car nous avons 9 séjours dans les données TCT dont 1 séjour a été sélectionné dans l'échantillon.

Figure 28 – Répartition des diagnostics principaux dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale - 2009



XXX – Dummy diagnosis.

Figure 29 – Différence relative entre les répartitions par diagnostic principal dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 560 Accouchement par voie vaginale - 2009



Les diagnostics principaux présents dans les données TCT et dont aucun séjour n'a été sélectionné dans l'échantillon ne sont plus repris dans ce graphique (640, 643, 668, 670, 673, 678 et XXX).

Table 26 – Tableau récapitulatif des tests de représentativité par variable - APR-DRG 560 Accouchement par voie vaginale

		Echantillon (poids AIM)			Echantillon (poids TCT)		
Année	Variable (nombre de catégories)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel max (nombre séjours)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel. max. avec la référence (nombre séjours)
2008	Catégorie d'âge (9)	0,63%	0,65	-21,1%	0,88%	0,53	-20,2%
	Statut social (2)	0,39%	0,54	-10,9%	0,34%	0,60	-6,1%
	Type d'hospitalisation (2)	0,01%	0,93	-9,4%	0,0%	0,98	-3,2%
	Région (patient) (3)	0,59%	0,69	-8,6%	0,58%	0,68	-4,2%
	Province (patient) (11)	1,52%	0,49	-22%	1,51%	0,48	-18,2%
	Arrondissement (patient) (43)	1,07%	0,62	-43,8%	1,06%	0,61	-50,1%
	Région (hôpital) (3)	0,34%	0,91	-7,6%	0,42%	0,88	-3,3%
	Province (hôpital) (11)	1,56%	0,17	-22,4%	1,56%	0,17	-18,7%
	Arrondissement (hôpital) (38)	1,13%	0,63	-52,9%	1,13%	0,61	-59,9%
	Niveau de sévérité (4)	1,15%	0,73	-32,3%	1,16%	0,73	-31,6%
	Diagnostic Principal (37)	1,33%	**	+117%	1,34%	**	+129%
	Montant total des remboursements	-3,1€*	<0,001		-3,3€*	<0,001	

Année	Variable (nombre de catégories)	Echantillon (poids AIM)			Echantillon (poids TCT)		
		Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel max (nombre séjours)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel. max. avec la référence (nombre séjours)
2009	Catégorie d'âge (9)	0,90%	0,78	+25,7%	1,11%	0,61	+26,3%
	Statut social (2)	0,13%	0,85	-7,5%	0,23%	0,73	+2,5%
	Type d'hospitalisation (2)	0,19%	0,21	-42,2%	0,20%	0,20	-38,1%
	Région (patient) (3)	0,14%	0,98	-7,8%	0,13%	0,99	+1,2%
	Province (patient) (11)	0,80%	0,58	-29,1%	0,76%	0,57	-22,9%
	Arrondissement (patient) (43)	1,29%	0,10	-83,7%	1,28	0,09	-82,1%
	Région (hôpital) (3)	0,23%	0,97	-8,0%	0,15%	0,99	-3,3%
	Province (hôpital) (11)	1,23%	0,62	-26,7%	1,24%	0,63	-20,2%
	Arrondissement (hôpital) (38)	1,27%	0,05	-69,9%	1,27%	0,04	-72,7%
	Niveau de sévérité (4)	0,56%	0,96	-19,4%	0,64%	0,94	-11,5%
	Diagnostic Principal (39)	1,46%	**	+339%	1,47%	**	+382%
	Montant total des remboursements	4,1€*	<0,001		4,2€*	<0,001	

** Différence entre les montants médians. ** Le nombre de séjours est nul dans l'EPS-TCT et non nul dans TCT pour une des classes de la variable. La méthode (SURVEYFREQ) ne contenant pas de facteur de correction pour les cellules vides, le Chi-square de Rao-Scott n'a pas pu être estimé.*

Conclusion pour l'APR-DRG 560 Accouchement par voie vaginale

Pour les accouchements par voie vaginale, les données de l'échantillon EPS-TCT dévient très peu des données nationales (TCT) en ce qui concerne les répartitions des variables étudiées. Cependant, il faut être vigilant quant à l'utilisation des fréquences pondérées pouvant sur- ou sous-estimer le nombre de séjours réels dans certaines catégories.

5.2.3.2 APR-DRG 313 Interventions des membres inférieurs et genoux, excepté le pied.

Le même exercice a été effectué pour l'APR-DRG 313 (Interventions des membres inférieurs et genoux, excepté le pied) et les différents résultats sont présentés en détail en annexe 5.1.2. La Table 27 est la table récapitulative. Le nombre de séjours avec une intervention des membres inférieurs et genoux est de 1621 et 1668 respectivement pour l'année 2008 et 2009.

Les Figure 30 et Figure 31 représentent la répartition des diagnostics et les différences relatives.

Les Figure 32 et Figure 33 représentent la distribution des montants totaux de remboursements par niveau de sévérité pour respectivement les hospitalisations classiques et les hospitalisations de jour.

Figure 30 – Répartition des diagnostics principaux dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 313 (Interventions des membres inférieurs et genoux, excepté le pied) - 2009

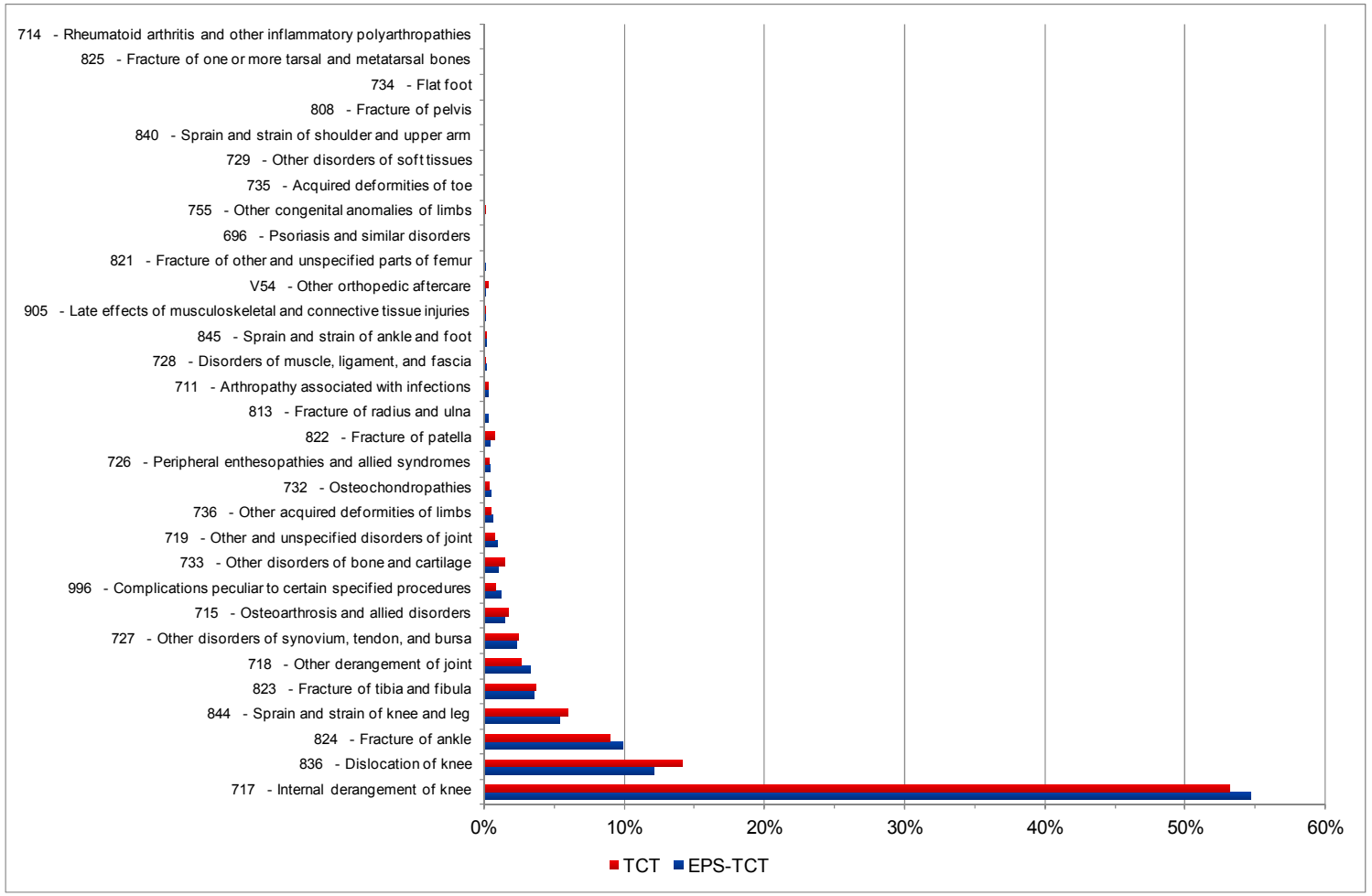


Figure 31 – Différence relative entre les répartitions par diagnostic principal dans l'échantillon EPS-TCT (poids TCT) et dans la population TCT de référence - APR-DRG 313 (Interventions des membres inférieurs et genoux, excepté le pied) - 2009

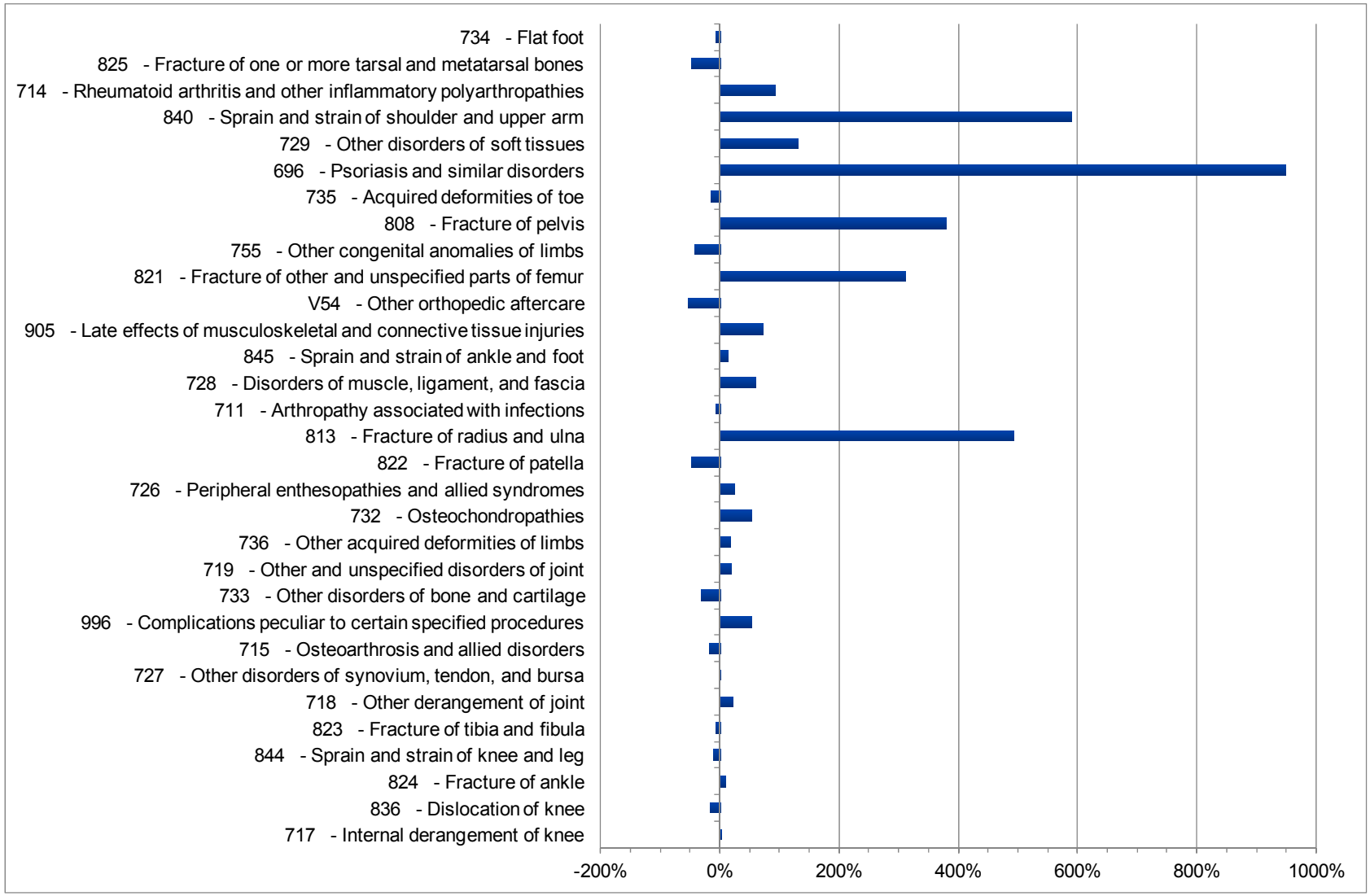
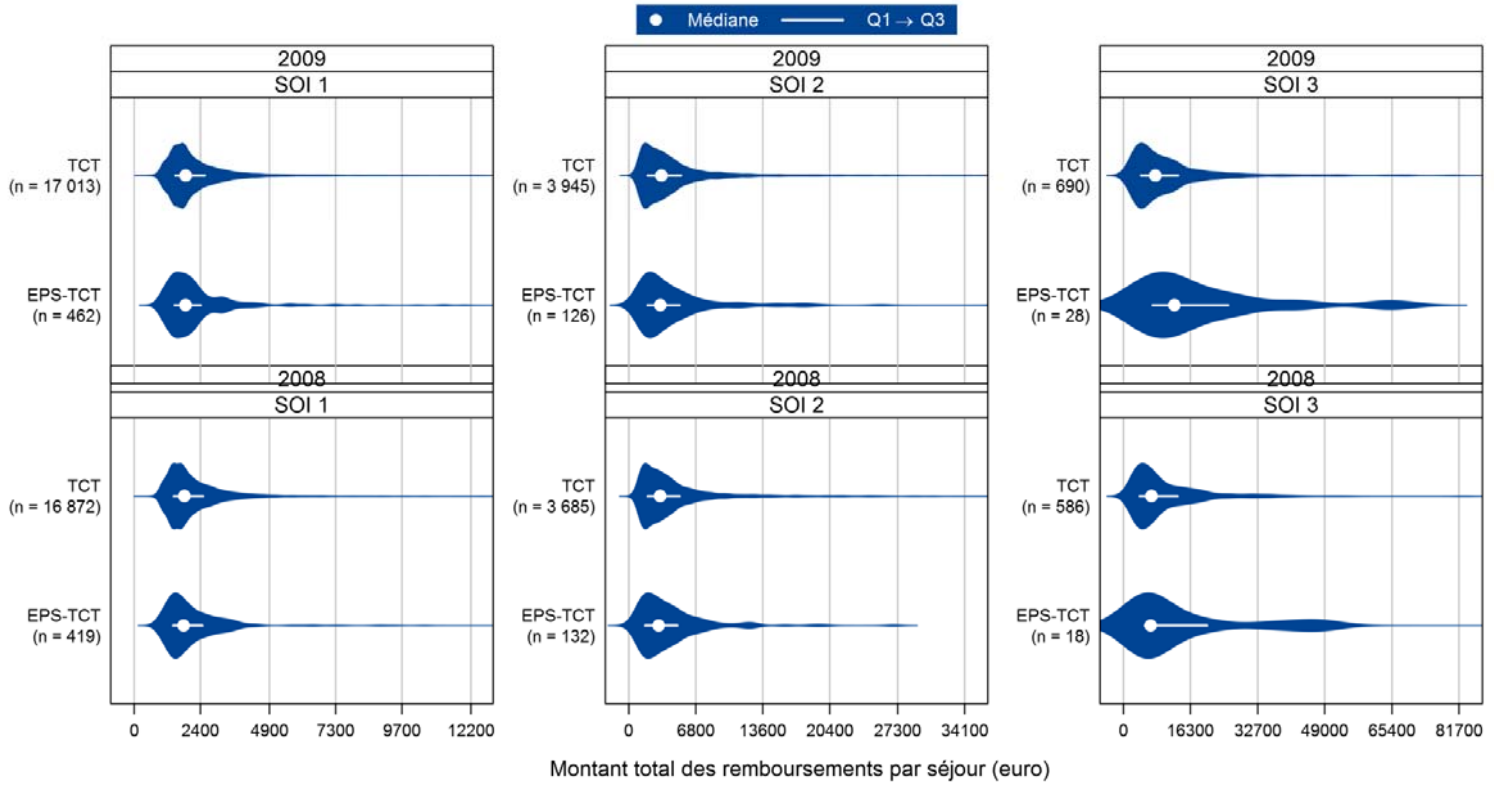
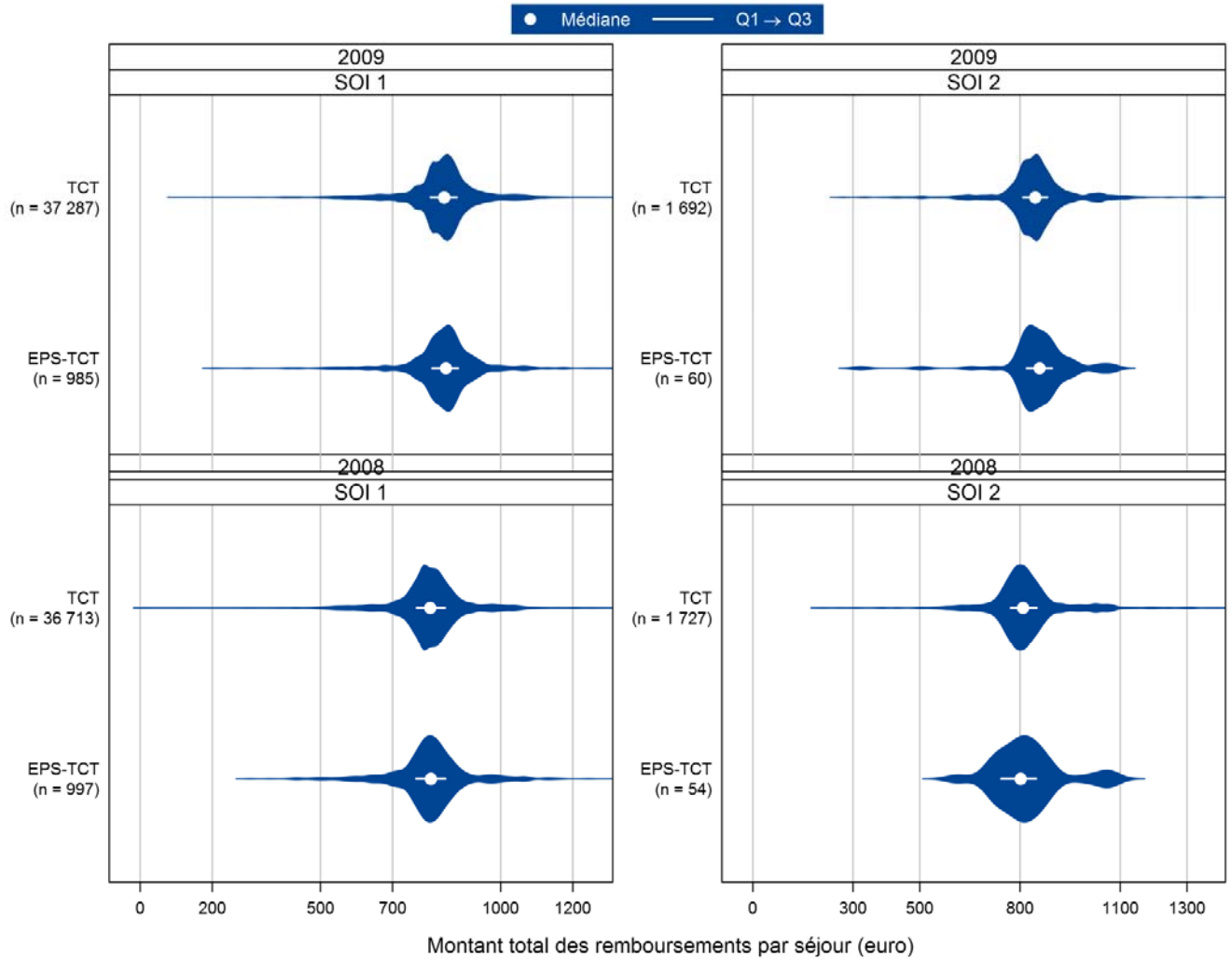


Figure 32 – Distribution du montant total de remboursement par niveau de sévérité dans l'échantillon EPS-TCT (poids TCT) et dans la population de référence TCT – Hospitalisations Classiques - APR-DRG 313 Interventions des membres inférieurs et genoux, excepté le pied



L'estimation par noyau de la densité de probabilité est une technique de lissage impliquant que la courbe déborde des limites des données de base.

Figure 33 – Distribution du montant total de remboursement par niveau de sévérité dans l'échantillon EPS-TCT (poids TCT) et dans la population de référence TCT – Hospitalisations de jour - APR-DRG 313 Interventions des membres inférieurs et genoux, excepté le pied



L'estimation par noyau de la densité de probabilité est une technique de lissage impliquant que la courbe déborde des limites des données de base.

Table 27 – Tableau récapitulatif des tests de représentativité par variable – APR-DRG 313 Interventions des membres inférieurs et genoux, excepté le pied

		Echantillon (poids AIM)			Echantillon (poids TCT)		
Année	Variable (nombre de catégories)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel max (nombre séjours)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel. max. avec la référence (nombre séjours)
2008	Catégorie d'âge (21)	1,27%	0,54	-28,8%	1,78%	0,46	+34,1%
	Sexe (2)	0,75%	0,56	-9,2%	1,2%	0,33	-5,4%
	Statut social (2)	1,20%	0,13	-8,8%	1,10%	0,17	+6,4%
	Type d'hospitalisation (2)	1,28%	0,30	-10,8%	1,21%	0,33	-5,8%
	Région (patient) (3)	0,60%	0,59	-18,1%	0,60%	0,58	-13,7%
	Province (patient) (11)	0,91%	0,90	-15,4%	0,88%	0,92	-10,0%
	Arrondissement (patient) (43)	1,10%	0,79	-65,8%	1,00%	0,81	-68,3%
	Région (hôpital) (3)	0,41%	0,91	-9,3%	0,36%	0,92	-4,2%
	Province (hôpital) (11)	0,98%	0,83	-27,7%	1,00%	0,81	-24,1%
	Arrondissement (hôpital) (38)	0,93%	0,42	-83,9%	0,91%	0,45	-82,5%
	Niveau de sévérité (4)	0,90%	0,22	-72,9%	0,85%	0,23	-72,7%
	Diagnostic Principal(69)	0,94%	**	+1 227%	0,66%	**	+1 345%
	Montant total des remboursements	-€2,4*	<0,001		-€2,3*	<0,001	

Année	Variable (nombre de catégories)	Echantillon (poids AIM)			Echantillon (poids TCT)		
		Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel max (nombre séjours)	Diff max. entre l'estimation et la référence (répartition)	p-value	Diff. Rel. max. avec la référence (nombre séjours)
2009	Catégorie d'âge (21)	1,42%	0,11	-30,5%	1,91%	0,05	+30,8%
	Sexe (2)	1,91%	0,13	-9,5%	1,5%	0,23	-3,6%
	Statut social (2)	0,72%	0,37	-7,1%	0,67%	0,40	+4,5%
	Type d'hospitalisation (2)	0,06%	0,95	-6,5%	0,05%	0,98	-1,1%
	Région (patient) (3)	2,44%	0,02	-15,4%	3,03%	0,02	+12,9%
	Province (patient) (11)	2,04%	0,01	-34,7%	2,06%	0,02	-30,2%
	Arrondissement (patient) (43)	1,19%	0,06	-129,9%	1,18%	0,06	-143,1%
	Région (hôpital) (3)	2,71%	0,05	-15,2%	2,72%	0,05	+14,1%
	Province (hôpital) (11)	2,20%	0,06	-27,3%	2,23%	0,06	-23,3%
	Arrondissement (hôpital) (38)	1,30%	0,01	-152,1%	1,31%	0,01	-167,3%
	Niveau de sévérité (4)	0,59%	0,50	+65,6%	0,55%	0,51	+79,5%
	Diagnostic Principal (71)	1,97%	**	+903%	2,01%	**	+948%
	Montant total des remboursements	€6,1*	<0,001		6,1€*		<0,001

** différence entre les montants médians.*
*** Le nombre de séjours est nul dans l'EPS-TCT et non nul dans TCT pour une des classes de la variable. La méthode (SURVEYFREQ) ne contenant pas de facteur de correction pour les cellules vides, le Chi-square de Rao-Scott n'a pas pu être estimé.*

Conclusion

Pour les interventions des membres inférieurs, les données de l'échantillon EPS-TCT dévient très peu des données nationales (TCT) en ce qui concerne les répartitions des variables étudiées sauf pour les variables géographiques pour lesquelles les différences peuvent aller jusqu'au environ de 3%. Si on considère que cette différence est trop importante pour conduire les analyses, il ne faudra pas prendre les variables géographiques en compte dans d'éventuels modèles statistiques.

5.3 Conclusion générale sur les tests de représentativité

Le tirage aléatoire stratifié par âge et sexe des bénéficiaires de l'EPS combiné au nombre suffisamment large de séjours TCT couplés à cet échantillon assure la représentativité de l'EPS-TCT. Cette représentativité permet d'extrapoler les résultats de l'échantillon (EPS-TCT) au niveau national (TCT) avec une marge d'erreur relativement faible.

Nous avons observé que sur l'ensemble des données, les différences entre, d'une part, les répartitions des variables étudiées (sexe, âge, statut social, type d'hospitalisation…) obtenues grâce aux méthodes d'extrapolation de l'EPS-TCT et, d'autre part, les répartitions de ces variables dans les données nationales (TCT), sont relativement faibles (<1%) même si statistiquement elles sont significatives au seuil de 5%. Ceci s'explique par le fait que les données EPS-TCT représentent un nombre important de séjours hospitaliers (~100 000 par année) et que la moindre différence est détectée du point de vue statistique. Nous pouvons donc être confiants dans l'utilisation des données EPS-TCT pour en déduire des conclusions nationales (en appliquant les méthodes d'analyses adéquates).

L'étude de 2 APR-DRG plus spécifiques (~2 000 et ~1 700 séjours dans l'EPS-TCT) permet de montrer que pour certaines variables les différences entre répartitions (EPS-TCT versus TCT) sont plus importantes : elles peuvent atteindre 3% dans le cas de ces 2 APR-DRG. Ces différences pourraient même être d'autant plus importantes que le nombre de séjours dans l'EPS-TCT pour l'APR-DRG d'intérêt est petit. Cela montre bien l'importance d'évaluer les marges d'erreur possibles avant de pouvoir utiliser les données. Il est nécessaire aussi de toujours mentionner les résultats avec leur intervalle de confiance afin de permettre une meilleure évaluation des variations possibles.

Il est aussi important de considérer les limitations suivantes selon l'utilisation des données de cet échantillon EPS-TCT :

1. Analyses de sous-groupe : toute analyse n'impliquant qu'une partie des données de l'échantillon ne peut être considérée comme valide que si la représentativité des données de ce sous-groupe a été testée (exemple : analyse d'un APR-DRG particulier).
2. Analyses d'autres variables : l'analyse d'autres variables pourrait être biaisée du fait, par exemple, que les effectifs par catégorie sont trop petits.
3. Analyses des fréquences pondérées : il est tentant d'utiliser directement les fréquences pondérées pour décrire l'épidémiologie d'une pathologie. Cependant, les différences relatives pouvant être assez larges, il est essentiel de ne pas les combiner aveuglément à d'autres résultats issus d'autres bases. Dans le but d'obtenir par exemple le taux d'hospitalisation pour une pathologie X par 1 000 habitants, la combinaison de données de l'EPS-TCT et de données démographiques nationales peut mener à un taux sous ou surestimé dont la marge d'erreur est importante. Il est préférable, dans ce cas, de combiner les données EPS-TCT avec les données démographiques de l'IMA tout en tenant en compte du taux de non couplage.
4. Méthodes d'analyse : les coefficients de pondération et les méthodes d'inférences sont importants pour obtenir les résultats les plus fiables possibles. Nous avons observé de grandes différences dans les fréquences pondérées en utilisant les poids déduits des données IMA d'une part et ceux déduits des données TCT d'autre part. Il est à noter que les poids à attribuer post-stratification (poids TCT) permettent d'ajuster les résultats compte tenu de la distribution des facteurs de stratification sexe et âge dans l'EPS-TCT d'une part et dans les données nationales (TCT) d'autre part. Cependant, pour des analyses spécifiques (à un niveau APR-DRG ou APR-DRG/SOI, ou pour un groupe de codes ICD-9-CM), il est recommandé de considérer des poids post-stratification standardisés [39, 40] permettant d'ajuster les résultats afin qu'ils soient les plus représentatifs possibles. Les méthodes d'analyses propres aux sondages (Survey) utilisent les facteurs de stratifications et les pondérations dans leurs calculs de variances et permettent d'obtenir des résultats plus robustes.

6 DISCUSSION ET CONCLUSIONS

Pertinence et mise en œuvre du projet

Le couplage des données du RHM (ou RCM dans le passé) avec les données nationales de remboursements hospitaliers des organismes assureurs, rassemblées par l'Agence InterMutualiste, s'est révélée indispensable à de nombreuses études : « Prévention médicamenteuse des fractures ostéoporotiques » (rapport KCE 159, 2011),[41] « Indicateurs de qualité en oncologie pour les cancers du sein et du testicule »(rapports KCE 149 et 150, 2011)[42, 43] ou encore « La stimulation cardiaque chez les patients bradycardes en Belgique » (rapport KCE 137, 2010),[44] « Élargissement du système de montants de référence pour des interventions standards avec une période de carence et l'intégration des séjours en hospitalisation de jour » (rapport AIM, 2010)[45] ou encore « Analyse de la politique d'admission-réadmission hospitalière, relative aux séjours cardiologiques, orthopédiques et psychiatriques » (rapport AIM, 2007).[46]

D'un côté le RHM permet d'identifier les pathologies dont souffrent les patients hospitalisés, à travers les diagnostics principaux ou secondaires et le regroupement par APR-DRG.[t] De l'autre, les données de population de l'AIM contiennent des données essentielles telles que le mois et l'année d'un éventuel décès, qu'il soit advenu en milieu hospitalier ou non, alors que le RHM ne contient de renseignements que sur les décès à l'hôpital. D'autres informations issues des deux sources se complètent également : l'enregistrement des procédures sous la classification ICD-9-CM permet une autre approche de l'information médicale que celle apportée par la nomenclature INAMI.

L'intérêt de disposer d'un échantillon EPS-TCT est de pouvoir accéder plus rapidement au couplage de données de pathologies avec des données de soins prodigués en milieu hospitalier, sans devoir automatiquement requérir un couplage des bases de données RHM ou TCT et AIM (soins de santé et population). L'autorisation systématique du Comité sectoriel de la Sécurité Sociale et de la Santé ainsi que le transfert de données extraites des bases AIM et TCT ne seraient donc plus nécessaires à chaque étude.

Un schéma de flux de données entre les acteurs a été mis au point et décrit dans la demande d'autorisation du Comité sectoriel de la Sécurité Sociale et de la Santé. Les différents acteurs comprenaient les organismes assureurs, le Collège Intermutualiste National (CIN), la plateforme eHealth, la Banque-Carrefour de la sécurité sociale, le SPF Santé publique, la Cellule Technique et finalement l'AIM et le KCE. Le schéma détaillait les échanges des tables de correspondance, comprenant plusieurs recodages des identifiants des patients ou de ceux des séjours, ainsi que les échanges des données en tant que telles. Notons que les identifiants des personnes physiques prestataires de soins ont également été anonymisés.

L'échange de tables de correspondance reprenant une clé d'identification unique du patient garantissant l'anonymat des données individuelles existait déjà depuis des années entre les acteurs (SPF-SP-TCT, INAMI-TCT, OA-AIM) et le recours à l'expertise de la TCT en matière de couplage des données RHM en provenance du SPF Santé publique et des données SHA/HJA en provenance de l'INAMI représente un atout majeur du schéma d'échange de données qui a été suivi pour ce projet. La plateforme eHealth, quant à elle, a joué son rôle de garant de la protection de la sécurité des données en servant d'intermédiaire recodant les clés d'identification.

La Cellule Technique a analysé le risque d'identification de personne au sein de « small cells » (courrier en annexe 3). Le risque a été éliminé par le nombre limité de variables demandées et par le recodage de certaines variables (p.ex. groupement des âges de plus de cent ans). Le nombre de séjours concernés atteignait moins de 0.01%.

[t] Une donnée précieuse, qu'est la présence ou non des diagnostics avant l'admission, est venue en outre compléter les données RHM dès 2008.

Résultats de la reconstitution des séjours hospitaliers dans l'EPS

L'algorithme de reconstitution des séjours hospitaliers dans l'EPS a été construit sur les précédents algorithmes utilisés par l'AIM au cours de ses différentes études. Le procédé a donc bénéficié d'une expérience de plusieurs années à l'AIM. Deux possibilités existaient pour déterminer la date de fin d'un séjour et le nombre de jours (à partir des dates de début de prestation ou à partir de la date de fin de prestation). L'étude a démontré l'équivalence de ses deux méthodes, qui sont identiques dans 99.9% des cas (compte tenu d'une différence acceptable de 1 jour). En outre, le couplage ultérieur avec les séjours dans les données TCT a permis de confirmer la validité de l'algorithme choisi.

Certains choix méthodologiques ont dû être faits, comme par exemple l'élimination des mini-forfaits qui n'ont pas été considérés comme des séjours hospitaliers à part entière. D'autres choix pourraient être faits pour des études spécifiques, requérant une adaptation de l'algorithme décrit dans le présent projet.

Résultats du couplage proprement dit

En fin de compte, 93,8% des séjours EPS ont pu être couplés (compte tenu d'une différence acceptable de 1 jour dans la date de début de séjour), ce qui est comparable au taux de couplage des études précédentes réalisées sur des extractions de l'EPS et que nous considérons comme un succès pour une première sur l'entièreté de l'EPS. Le couplage est donc possible, même si l'analyse des rejets (séjours non couplés) peut toujours à l'avenir permettre de le perfectionner.

A un 'méso'-niveau, la représentativité de chaque institution ne relevait pas de la délimitation du champ du projet et n'a donc pas été étudiée. Il n'empêche que les résultats du couplage EPS-TCT ont montré que le taux de couplage des séjours au niveau de chaque hôpital joue à cet égard un rôle très important. Ce taux varie en effet fortement selon l'institution et influence le taux de couplage au niveau de l'échantillon en fonction de sa sur- ou sous-représentativité dans cet échantillon.

Résultats des tests de représentativité

En considérant tous les séjours hospitaliers, tous APR-DRG confondus, nous avons montré que l'on pouvait utiliser les données de l'échantillon EPS-TCT couplé pour décrire les données d'hospitalisation (TCT) ; le tirage aléatoire combiné au nombre suffisamment large de séjours tirés assurant la représentativité de l'échantillon par rapport à certaines variables d'intérêt général (sexe, âge, région, statut social, type d'hospitalisation...). Cela rejoint la théorie de l'échantillonnage selon laquelle tout échantillon tiré d'une population de façon aléatoire est représentatif de la population en question.

L'échantillonnage n'utilisant pas la variable APR-DRG comme facteur de stratification, il est nécessaire, préalablement à toute analyse, d'effectuer les tests de représentativité sur l'APR-DRG. Ceci afin de connaître les limites des données de l'échantillon et d'être plus confiants quant à leur interprétation élargie à la population hospitalière toute entière.

Par exemple, la représentativité ayant pu être démontrée pour les différentes variables dans le cas de l'APR-DRG 560 « Accouchement par voie vaginale » (tous niveaux de sévérité confondus), les données EPS-TCT 2008-2009 peuvent être utilisées dans le cadre d'études traitant des accouchements médicaux par voie vaginale en Belgique.

En revanche, la représentativité au niveau géographique n'est pas aussi satisfaisante en 2009 pour l'APR-DRG 313 « Interventions des membres inférieurs et genoux, excepté le pied ». Aucun résultat obtenu par région ni/ou province ne peut être généralisé au niveau belge. En outre, l'utilisation de ces facteurs dans des modèles statistiques pourraient biaiser les résultats.

Quant au niveau de granularité inférieur ou de sous-groupes, il convient d'étudier la représentativité de l'échantillon en fonction des besoins spécifiques à chaque étude.

Limites

Les facteurs testés dans le présent projet constituent les variables principalement utilisées lorsque l'on étudie les données hospitalières d'un point de vue statistique, financier ou épidémiologique. Il existe d'autres facteurs sur lesquels une représentativité n'a pas été testée. La méthodologie utilisée peut du reste tout à fait être appliquée à d'autres variables, en fonction de questions de recherche plus spécifiques. Il est même indispensable de tester cette représentativité dans les trois cas suivants :

1. utilisation d'autres variables (comme par exemple le passage dans les services d'urgence, ou le recours à certains services),
2. analyse à des niveaux supplémentaires de détail des variables déjà étudiées (sous-groupes, croisement de variables). Les sous-chapitres de diagnostics (principaux) non étudiés, forment par exemple, des groupes potentiels cliniquement pertinents,
3. utilisation d'autres années que 2008-2009.

Il faut également rester vigilant lors de toute étude en connaissant et gardant à l'esprit les spécificités de chaque base de données. Par exemple, l'échantillon EPS-TCT ne comprend pas les hospitalisations de personnes non affiliées à un organisme assureur puisqu'elles ne font pas partie de la population de l'EPS ; les séjours ayant débuté avant l'année de couplage (en l'occurrence 2008), sont dans les données TCT mais n'ont pas pu être couplées avec les informations de l'EPS, ce qui peut s'avérer important pour des analyses saisonnières (début d'année). Cette dernière limitation n'existerait plus dans le cas d'un échantillon couplé pérenne. En outre, les finalités des bases de données sont différentes : le RHM a une finalité clinique ainsi que de financement hospitalier alors que les données de l'AIM suivent une logique de facturation (ce qui explique les différences d'enregistrements concernant les nouveau-nés). Enfin, les modifications légales (ex. intégration des petits risques des travailleurs indépendants pour l'EPS) ou en termes de classification clinique (ex. changement de version ICD-9-CM ou APR-DRG pour le RHM) constituent des pièges dangereux pour les études longitudinales.

La validité des données ne relevant pas du champ de l'étude, les valeurs aberrantes n'ont pas fait l'objet de contrôles, que ce soit au niveau de l'échantillon EPS-TCT ou au niveau des données TCT nationales.

Nous avons déjà évoqué l'influence du taux de couplage variant d'une institution à l'autre sur le couplage EPS-TCT. La manière de coder les procédures et les diagnostiques peut également varier d'un hôpital à l'autre, ce qui pourrait introduire un biais si un hôpital est surreprésenté au niveau de l'échantillon, d'autant plus s'il s'agit d'un hôpital de grande taille.

En ce qui concerne les « coûts », seuls les remboursements par les organismes assureurs aux hôpitaux ont été pris en compte. Les frais supportés par le patient qui sont disponibles dans l'EPS sont donc exclus de l'étude. En outre, aucune tentative d'approcher les coûts réels n'a été entreprise. L'objectif était de vérifier qu'aucun biais n'existait au niveau du montant total remboursé entre échantillon et données TCT nationales. Le calcul de ce montant a été effectué par la Cellule Technique avant livraison des données TCT et comprenait un recalcul du prix de journée à 100%.

La représentativité du montant total des remboursements a été démontrée au niveau global ainsi que pour les deux APR-DRG sélectionnés (elle ne l'est pas au niveau de chaque APR-DRG, et a fortiori non plus à tous les niveaux APR-DRG/niveau de sévérité). Cependant il faut tenir compte de la particularité suivante du système belge de couverture des soins de santé. Le montant total inclut, d'une part, les montants remboursés par acte ou par forfait (prestations dont la biologie clinique et l'imagerie médicale, médicaments, implants...) et, d'autre part, le prix de journée extrapolé à 100%. Les premiers montants varient selon le statut préférentiel du patient. De même, la proportion du prix de journée extrapolé à 100% couverte par le ticket modérateur varie en fonction du statut préférentiel. Mais le montant total de ce prix de journée extrapolé est indépendante du statut préférentiel. Dans le cas de certaines pathologies pour lesquelles le prix de journée constitue une grosse partie des remboursements, l'influence du statut préférentiel sera donc moindre que dans le cas contraire. Il est donc essentiel de connaître l'influence du statut préférentiel sur les éléments composant la facture avant d'entamer toute étude sur ces montants.

Pour les études portant sur le financement, la méthodologie devra tenir compte de toute la complexité du système de financement hospitalier belge. Même si l'extrapolation du prix de journée à 100% vise à tenir compte de la partie du financement couverte par le versement en douzièmes du Budget des Moyens Financiers, les prestations peuvent être financées à l'acte et/ou forfaitairement. Pour ces études, il n'est donc pas possible de faire l'économie d'une méthodologie claire et argumentée sur le calcul des coûts.

Futur

Un échantillon EPS-TCT pérenne, issu d'un couplage répété chaque année, permettrait une meilleure gestion des ressources et de la faisabilité de certains projets de recherche. Une extraction permet en effet de tester la faisabilité d'une étude avant de devoir entamer la procédure nécessaire lorsque les données RHM et AIM doivent être couplées au niveau national. En outre, cette pérennité permettrait de mener des études longitudinales, en suivant le patient à travers le temps grâce à la conservation d'une même clé d'identification du patient d'une année à l'autre.

S'il faut une autorisation de communication de données au Comité Sectoriel de la Sécurité Sociale et de la Santé pour chaque couplage 'ad-hoc', en revanche une base légale est nécessaire pour permettre un couplage pérenne.

Dans toute nouvelle initiative c'est le premier pas qui coûte. Le processus a été complexe à mettre en place, mais peut désormais fonctionner. A cet égard, une systématisation de l'échange de tables de correspondance entre les différents acteurs dans le cadre d'un échantillon EPS-TCT pérenne ne peut que perfectionner la qualité du couplage, la qualité des tables de correspondance s'améliorant d'année en année.

En cas de couplage des années antérieures (2002-2007), il faudra tenir compte des changements introduits par l'intégration des petits risques du régime des travailleurs indépendants dans l'assurance obligatoire dès 2008. Ainsi, les hospitalisations de jour pour les travailleurs indépendants (considérés comme petits risques jusqu'alors) sont absentes des données EPS avant le 1er janvier 2008. En outre, la répartition géographique hétérogène de ces travailleurs indépendants pourrait avoir un impact sur la représentativité de l'EPS-TCT.

Conclusions

D'une part l'algorithme de reconstitution des séjours hospitaliers utilisé dans l'étude a pu être validé. D'autre part, la méthodologie de couplage a été mise au point et les variables nécessaires identifiées. Le couplage s'est révélé satisfaisant. Quant à la méthodologie élaborée pour tester la représentativité, elle a permis de montrer que les ordres de grandeur étaient globalement respectés. Elle a en outre décrit des zones moins bien représentées dans les données (par APR-DRG et APR-DRG/niveau de sévérité). Ainsi, les études nécessitant l'analyse des données EPS-TCT 2008-2009 peuvent être exécutées en toute connaissance de cause.

Bien entendu, la représentativité décrite pour 2008-2009 peut ne pas être similaire pour les autres années. Mais lorsque les utilisateurs des données EPS-TCT couplées doivent connaître la représentativité de cet échantillon et ses limites, la méthodologie statistique développée sur les données 2008-2009 peut être reproduite sur les années ultérieures (ou antérieures).

RÉFÉRENCES

1. Januel JM, Luthi JC, Quan H, Borst F, Taffe P, Ghali WA, et al. Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. BMC Health Serv Res. 2011;11 :194.
2. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. Med Care. 2005;43(5) :480-5.
3. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived rom ICD-9-CCM administrative data. Med Care. 2002;40(8) :675-85.
4. Vlayen J, Leys M, Ramaekers D, Camberlin C, Paulus D, Van De Water G, et al. Indicateurs de qualité cliniques. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2006. (D/2006/10.273/44) Available from : https://kce.fgov.be/sites/default/files/page_documents/d20061027344.pdf
5. Van De Sande S, De Wachter D, Swartenbroekx N, Peers J, Debruyne H, Moldenaers I, et al. Inventaire des bases de données de soins de santé. Objective Elements - Communication (OEC). Bruxelles : Centre Fédéral d'Expertise des Soins de Santé (KCE); 2006 19/05/2006. KCE Reports 30B (D2006/10.273/15) Available from : https://kce.fgov.be/sites/default/files/page_documents/d20061027315.pdf
6. Loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel M. B. 18 mars 1993. Available from : http://www.ejustice.just.fgov.be/loi/loi.htm
7. Loi du 15 janvier 1990 relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale, M. B. 22 février 1990. Available from : http ://www.ejustice.just.fgov.be/loi/loi.htm
8. Loi du 27 décembre 2005 portant des dispositions diverses, M. B. 30 décembre 2005.

9. Van De Sande S, De Wachter D. Technische nota bij Hoofdstuk 3 Datavoorbereiding & Berekeningsmodule met Alternatieven. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2005. Available from : https://kce.fgov.be/sites/default/files/page_documents/d20051027336(3).pdf
10. De Roquefeuil L, Studer A, Neumann A, Y M. L'échantillon généraliste de bénéficiaires : représentativité, portée et limites. Pratiques et Organisation des Soins. 2009;40(3) :11.
11. Haut Conseil de la Santé publique. Pour une meilleure utilisation des bases de données nationales pour la santé publique et la recherche. 2012, Mars. Collection Documents Available from : http://www.hcsp.fr/explore.cgi/hcspr20120309_bddadmination.pdf
12. Loi-programme (I) du 24 décembre 2002, M. B. 31 décembre 2002. Available from : http ://www.ejustice.just.fgov.be/loi/loi.htm
13. Arrêté royal du 29 septembre 1998 portant exécution de l'article 206, § 2, de la loi relative à l'assurance obligatoire soins de santé et indemnités, coordonnée le 14 juillet 1994, M. B. 27 octobre 1998. Available from : http ://www.ejustice.just.fgov.be/doc/rech_f.htm
14. Arrêté royal du 16 octobre 1998 portant exécution de l'article 156, alinéa 3, de la loi du 29 avril 1996 portant des dispositions sociales en ce qui concerne le délai dans lequel et la manière dont le Ministère des Affaires sociales, de la Santé publique et de l'Environnement transmet le résumé clinique minimum concernant les années 1995, 1996 et 1997 à la cellule technique M. B. 1 décembre 1998. Available from : http://www.ejustice.just.fgov.be/doc/rech_f.htm
15. Arrêté royal du 22 mars 1999 portant exécution de l'article 156, alinéa 5, de la loi du 29 avril 1996 portant des dispositions sociales, en ce qui concerne les modalités selon lesquelles les organismes assureurs sont tenus de transmettre à la cellule technique les informations nécessaires à la fusion des données cliniques minimum et financières anonymes des années 1995 et 1997, M. B. 7 avril 1999. Available from : http://www.ejustice.just.fgov.be/doc/rech_f.htm
16. Arrêté royal du 22 mars 1999 portant exécution de l'article 156, alinéa 5, de la loi du 29 avril 1996 portant des dispositions sociales, en ce qui concerne les modalités selon lesquelles les hôpitaux généraux non psychiatriques sont tenus de transmettre à la cellule technique les informations nécessaires à la fusion des données cliniques minimum et financières anonymes des années 1995 et 1997 M. B. 7 avril 1999. Available from : http://www.ejustice.just.fgov.be/doc/rech_f.htm
17. Arrêté royal du 24 septembre 2001 modifiant l'arrêté royal du 22 mars 1999 portant exécution de l'article 156, alinéa 5, de la loi du 29 avril 1996 portant des dispositions sociales, en ce qui concerne les modalités selon lesquelles les organismes assureurs sont tenus de transmettre à la cellule technique les informations nécessaires à la fusion des données cliniques minimum et financières anonymes des années 1995 et 1997 M. B. 9 octobre 2001. Available from : http://www.ejustice.just.fgov.be/doc/rech_f.htm
18. Arrêté royal du 24 septembre 2001 modifiant l'arrêté royal du 22 mars 1999 portant exécution de l'article 156, alinéa 5, de la loi du 29 avril 1996 portant des dispositions sociales, en ce qui concerne les modalités selon lesquelles les hôpitaux généraux non psychiatriques sont tenus de transmettre à la cellule technique les informations nécessaires à la fusion des données cliniques minimum et financières anonymes des années 1995 et 1997 M. B. 9 octobre 2001. Available from : http://www.ejustice.just.fgov.be/doc/rech_f.htm
19. Loi du 29 avril 1996 portant des dispositions sociales M. B. 30 avril 1996. Available from : http ://www.ejustice.just.fgov.be/loi/loi.htm
20. Anderson R. Security Engineering : A Guide to Building Dependable Distributed Systems, 2nd Edition. In : Wiley, editor.; 2008.
21. Hash function [Web page]. 2013 [cited 29 July 29 2013]. Available from : http ://en.wikipedia.org/wiki/Hash_function
22. Pommerening K, Reng M. Secondary Use of the EHR via Pseudonymisation [Web page]. Telematikplattform für die medizinischen Forschungsnetze [cited 2013, February 11]. Available from : http://www.staff.uni-mainz.de/pommeren/Vortraege/PseudoSlides.pdf

23. Lo Iacono L. Multi-centric universal pseudonymisation for secondary use of the EHR. Stud Health Technol Inform. 2007;126 :239-47.
24. Arrêté royal du 27 avril 2007 déterminant les règles suivant lesquelles certaines données hospitalières doivent être communiquées au Ministre qui a la Santé publique dans ses attributions, M. B. 10 juillet 2007. Available from : http://www.ejustice.just.fgov.be/doc/rech_f.htm
25. INAMI. Convention nationale entre les établissements hospitaliers et les organismes assureurs du 24 juin 2011. In : Service des soins de santé, editor. Bruxelles; 2011. p. 34.
26. Arrêté royal du 25 avril 2002 relatif à la fixation et la liquidation du budget des moyens financiers des hôpitaux, M. B. 30 mai 2002. Available from : http ://www.ejustice.just.fgov.be/doc/rech_f.htm
27. INAMI. Instructions aux établissements hospitaliers, aux laboratoires de biologie clinique agréés, aux praticiens de l'art infirmier et à tous les autres établissements et dispensateurs qui recourent au système de deliverance des fichiers de facturation et aux organismes assureurs - Edition 2006. In. Bruxelles; 2006. p. 524.
28. Van De Sande S, De Ryck D, De Gauquier K, Hilderson R, Neyt M, Peeters G, et al. Étude de faisabilité de l'introduction en Belgique d'un système de financement hospitalier «all-in» par pathologie. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2010. (D/2010/10.273/02) Available from : https://kce.fgov.be/sites/default/files/page_documents/d20101027302.pdf
29. INAMI. Document SHA - Edition 16. Instructions aux organismes assureurs relatives à la transmission des tableaux statistiques par séjour hospitalier anonyme. Mise à jour d'application pour les données comptabilisées du 01/07/2006 au 30/06/2008. Bruxelles; 2001.
30. Schumacher S. Probabilistic Versus Deterministic Data Matching : Making an Accurate Decision [Web page]. 2007 [cited 2013, February 11]. Available from : http ://www.information-management.com/specialreports/20070118/1071712-1.html
31. SAS Institute Inc. SAS 9.3 BASE & STAT/GRAPH. Cary, NC : SAS Institute Inc.; 2012. Available from : http ://www.sas.com
32. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Statistical Science. 2001;16(2) :18.
33. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Health Care. 2005;14(5) :347-51.
34. Vaillant J. Initiation à la théorie de l'échantillonnage [Web page]. 2005. Available from : http://monnano.weebly.com/uploads/1/6/6/3/1663287/theoechtage.pdf
35. Shqiponje T, Hutton D, Davies P, Adnett N, Coe R. Competition Within Schools : Representativeness of Yellis Sample Schools in a Study of Subject Enrollment of 14-16 Year Olds. In : Institute for Education Policy Research; 2004. p. 17.
36. Rao JNK, Scott AJ. On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. The Annals of Statistics. 1984;12(1) :15.
37. R Core Team. R 3.0.1 : A Language and Environment for Statistical Computing. Vienna, Austria : R Foundation for Statistical Computing; 2013. Available from : http ://www.R-project.org/
38. Loi du 22 août 2002 portant des mesures en matière de soins de santé, M. B. 22 septembre 2002. Available from : http://www.ejustice.just.fgov.be/loi/loi.htm
39. Little RJA. Inference with survey weights. Journal of Official Statistics. 1991;7(4) :405-24.
40. Little RJA. Post Stratification : a Modeler's Perspective. Journal of the American Statistical Association. 1993;88 :1001-12.
41. Roberfroid D, Camberlin C, Dubois C. Prévention médicamenteuse des fractures ostéoporotiques. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2011. (D/2011/10.273/33) Available from: https://kce.fgov.be/sites/default/files/page_documents/kce_159b_fractures_ostéoporotiques.pdf

42. Vlayen J, Vrijens F, Beirens K, Stordeur S, Devriese S, Van Eycken E. Indicateurs de qualité en oncologie : cancer du testicule. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2010. (D/2010/10.273/97) Available from : https://kce.fgov.be/sites/default/files/page_documents/kce_149b_cancer_du_testicule.pdf
43. Stordeur S, Vrijens F, Vlayen J, Beirens K, Devriese S, Van Eycken E. Indicateurs de qualité en oncologie : Cancer du sein. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2010. (D/2010/10.273/100) Available from : https://kce.fgov.be/sites/default/files/page_documents/kce_150b_cancer_du_sein.pdf
44. Van Brabandt H, Neyt M, Stroobandt S, Van De Sande S, Schwierz C. La stimulation cardiaque chez les patients bradycardes en Belgique. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE); 2010. (D/2010/10.273/57) Available from : https://kce.fgov.be/sites/default/files/page_documents/kce_137b_la_stimulation_cardiaque.pdf
45. AIM. Élargissement du système de montants de référence pour des interventions standards avec une période de carence et l'intégration des séjours en hospitalisation de jour. 2010.
46. AIM. Analyse de la politique d'admission-réadmission hospitalière, relative aux séjours cardiologiques, orthopédiques et psychiatriques. 2007.