

# Validité des données hospitalières pour leur exploitation en épidémiologie

---

## **PROJET AGORA N° SSTC/AG/10/031**

M. Gilbert, D. Thimus, F.H. Roger-France  
Ecole de Santé Publique, Health Services Research, Unité d'Informatique Médicale  
Université Catholique de Louvain

C. Camberlain, I. Mertens  
Ministère de la Santé Publique

A. Figeys, D. Moens  
Institut National d'Assurance contre la Maladie et l'Invalidité

### ***Collaborateurs principaux:***

Dr F. Capet, Mme G. Ducoffre et Dr G. Hanquet  
Institut de santé Publique, section d'épidémiologie

Dr Ch. Daumerie  
Cliniques Universitaires St-Luc, Université Catholique de Louvain

Mme D. Deliège, Dr M. Jeanjean  
Ecole de Santé Publique, Université Catholique de Louvain

Dr Y. Gillerot  
Institut de génétique de Lovreval, Eurocat pour les provinces de Namur et Hainaut oriental

Dr M. Haelterman, Dr L. Van Eycken  
Registre National du Cancer et Registre flamand du cancer

### ***Remerciements***

Nous remercions les Services Fédéraux des Affaires Scientifiques, Techniques et Culturelles (SSTC) pour la subvention de ce projet.

## Sommaire

|  |     |
|--|-----|
| 1. <a href="#">Introduction– Objectif général de l'étude</a> ..... | P3  |
| 2. <a href="#">Méthologie d'exploitation des données</a> .....     | P4  |
| 3. <a href="#">L'infarctus du myocarde</a> .....                   | P9  |
| 4. <a href="#">Le cancer du sein chez la femme</a> .....           | P13 |
| 5. <a href="#">Le cancer de la thyroïde</a> .....                  | P21 |
| 6. <a href="#">Les infections à méningocoques</a> .....            | P31 |
| 7. <a href="#">L'hypospadias</a> .....                             | P37 |
| 8. <a href="#">Le diabète sucré</a> .....                          | P42 |
| 9. <a href="#">Conclusions et perspectives</a> .....               | P53 |
| 10. <a href="#">Annexe : programme SAS</a> .....                   | P54 |

Le présent rapport se situe dans un centre d'intérêt pour de nombreuses disciplines travaillant dans le domaine de la santé publique : personnel administratif, gestionnaires, économistes, médecins... Afin de favoriser la lisibilité pour chacun, ses différents chapitres sont concis et leur lecture peut être indépendante de l'ensemble du rapport.

L'analyse de chaque maladie traçante est présentée dans les paragraphes suivants :

- Objectif particulier de l'étude
- Matériel et méthode
- Résultats
- Discussion
- Conclusion
- Références bibliographiques
- Collaborateurs
- Annexe : figures illustrant les résultats

Pour chaque maladie, un schéma résumant la faisabilité et les limites des études épidémiologiques sur les données hospitalières est inclus dans la discussion des résultats.

L'étude est centrée principalement sur les données du Résumé Clinique Minimum (RCM) du Ministère de la Santé publique, qui enregistrent notamment le diagnostic des maladies et les prestations médicales s'y référant. Les données des prestations de l'INAMI ont aussi été exploitées, soit dans des fichiers de données agrégées par code de cette nomenclature, soit dans les fichiers contenant le Résumé Financier Minimum (RFM) alors liées au RCM par le numéro de séjour du patient. Les fichiers RFM n'ont été disponibles que pour la fin de notre analyse, lors de l'étude de la pathologie diabétique.

## I. Introduction - Objectif général de l'étude

En Belgique, les dépenses pour les prestations hospitalières sont remboursées par l'Institut National d'Assurance contre la Maladie et l'Invalidité (I.N.A.M.I.) Depuis 1990, pour tenter de contrôler ces dépenses, le Ministère de la Santé Publique demande aux hôpitaux dispensant des soins aigus, de lui envoyer les données administratives et médicales de tous les séjours, sous forme d'un Résumé Clinique Minimum (RCM).

Sur le RCM, figurent des données administratives concernant l'hôpital et le service, l'âge et le sexe du patient, l'arrondissement de son domicile, ainsi que des données médicales sur le séjour du patient, principalement les diagnostics et certaines prestations. Les RCM sont envoyées au Ministère, sous un format anonymisé. Ils sont couplés aux données du Résumé Financier Minimum (RFM) contenant l'information sur le remboursement des prestations, médicaments et autres fournitures. Tous les séjours sont classifiés dans les groupes appelés Diagnosis Related Groups (DRG) L'hôpital peut être pénalisé en fonction de la durée du séjour ou du nombre de certaines prestations, comparés aux moyennes nationales pour le même DRG dans la même catégorie d'hôpitaux.

L'étude présente a été demandée conjointement par le Ministère de la santé Publique et l'INAMI. Son objectif est de valoriser ces données pour leur utilisation en épidémiologie.

En 2000 et 2002, le Ministère de la Santé Publique a déjà publié un atlas montrant la répartition géographique de certaines maladies dans les hôpitaux belges en 1996 et en 1998 [1,2]. Cette analyse épidémiologique globale des RCM était faite par séjour hospitalier, sur base du diagnostic principal enregistré et du domicile du patient.

Certaines méthodes pour étudier la validité des données hospitalières ont déjà fait l'objet d'autres études européennes. Ainsi, une étude finlandaise a choisi la validation des données hospitalières en les comparant avec celles de l'étude MONICA [Réf. 1 Infarctus]. Une étude française a récemment publié les résultats d'une analyse de données hospitalières pour l'incidence du cancer du sein en les comparant à ceux d'un registre régional du cancer [Réf. 3 Cancer du sein]. Nous n'avons pas examiné systématiquement la correspondance entre le contenu du RCM et celui du dossier médical mais cet examen a déjà été effectué précédemment en Belgique et dans d'autres pays européens avec des résultats variables selon les maladies étudiées [Réf. 4 Cancer du sein]

Outre les études très classiques d'incidence, les données hospitalières ont parfois servi des études cliniques par exemple de la prévalence hospitalière d'une maladie ou de ces complications [Réf. 4 Diabète]

Beaucoup de recherches ont donc déjà testé la validité des données hospitalières pour leur usage en épidémiologie. Après avoir établi de manière théorique les critères de faisabilité, notre étude essaye de répertorier systématiquement les différents biais pour différentes analyses et détermine quelques procédures pour les surmonter.

- 
1. Bogaert J, Pincé H. RCM 1996 en images. Variation géographique de la pathologie dans les *hôpitaux belges*. Ministère des Affaires Sociales, de la Santé Publique et de l'Environnement. Administration des soins de santé. Direction de la politique des soins de santé,(Service Traitement de l'information médicale), 2000.
  2. RCM 1998. Variation géographique de la pathologie dans les hôpitaux belges. Ministère de la Santé Publique. Direction de la politique des soins de santé. Service traitement de l'information médicale. <http://www.health.fgov.be/vesalius/startf.htm>

## II. La méthodologie

Le choix des maladies étudiées a été fait selon trois critères principaux:

- l'intérêt de leur étude pour la santé publique,
- la faisabilité de l'étude établie théoriquement a priori
- et surtout la disponibilité d'autres sources de données ou de résultats comparables.

Les maladies choisies, le type d'étude effectué, la méthode de validation et les autres sources disponibles pour étudier ces maladies sont décrites dans le tableau 1.

| Maladies traçantes        | Type d'étude / année(s)                              | Méthode                   | Bases de référence  |
|---------------------------|--|---------------------------|---|
| Infarctus du myocarde     | Incidence* / 1994-1996                               | Comparaison des données   | Registre MONICA   |
| Cancer du sein            | Incidence / 1996-1997                                | Comparaison des résultats | Registre National du Cancer                               |
| Cancer de la thyroïde     | Incidence / 1993-1998                                |                           | Références internationales<br>Registre flamand            |
| Infection à méningocoques | Incidence / 1993-1998                                |                           | Laboratoires vigies - (ISP)                               |
| Hypospadias               | Prévalence* / 1993-1998                              |                           | Registre régional.  |
| Diabète                   | Prévalence hospitalière<br>Analyses cliniques / 1997 |                           | Médecins vigies - (ISP)<br>Enquête de santé par interview |

Tableau 1 - Maladies traçantes et méthodes de validation des études sur les données RCM+RFM

- La base de données de travail est construite :
  - A partir des données médicales, afin de sélectionner les séjours pour la maladie choisie: diagnostics principaux et secondaires, procédures médicales et éventuellement médicaments.
  - A partir des données administratives, afin de définir les patients atteints de cette maladie: numéro anonyme de l'hôpital, numéro anonyme du patient, âge, sexe et code de l'arrondissement du domicile du patient (Cfr. Programme SAS en annexe de ce rapport)
- Pour établir a priori les critères de faisabilité des études et localiser les limites théoriques de leur validité, nous présentons dans la fig. 1, les facteurs qui peuvent expliquer la différence entre la morbidité de la maladie dans la population et celle que nous pouvons mesurer à l'aide des données hospitalières.

La morbidité apparente d'une maladie sur les données hospitalières dépend des facteurs suivants:

- la morbidité réelle au niveau régional,
- le système de référence existant pour les admissions à l'hôpital ou les transferts éventuels,
- les démarches diagnostiques et thérapeutiques en faveur du patient
- et la qualité de l'enregistrement des données.

\* Incidence = nombre de nouveaux cas dans une période de temps déterminée (souvent , incidence annuelle)

\* Prévalence = proportion des patients atteints à un moment précis dans le temps ou une période de vie déterminée

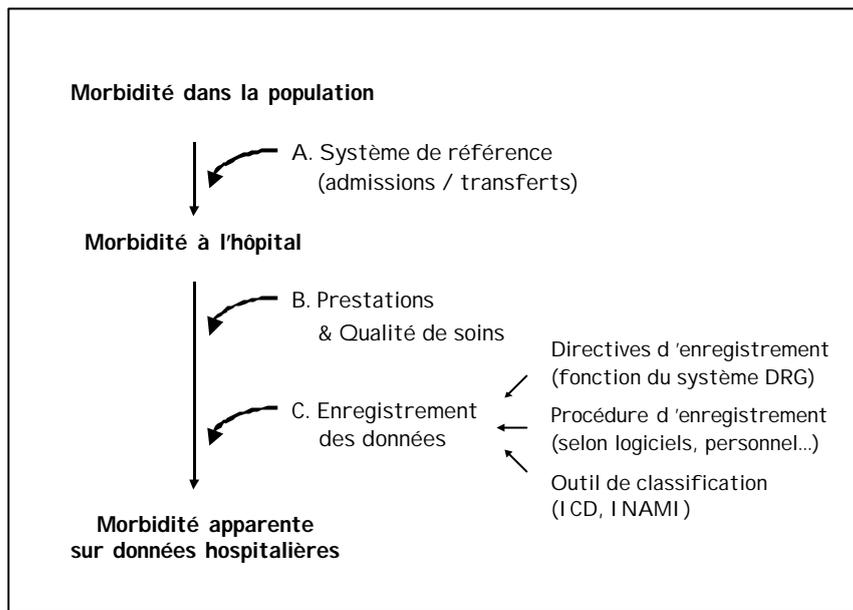


Fig. 1. Facteurs influençant la morbidité apparente d'une maladie sur les données hospitalières

#### A. Le système de référence de l'hôpital

La condition sine qua non pour utiliser les données hospitalières en épidémiologie est évidemment l'hospitalisation fréquente, en "séjour aigu", des patients atteints de la maladie que l'on veut étudier.

*Il est par exemple illusoire d'étudier la prévalence de la dépression chronique sur RCM. En effet, les patients qui en sont atteints ne séjournent pas souvent à l'hôpital ou si c'est le cas, une partie d'entre eux fréquente les hôpitaux psychiatriques, institutions dispensant des soins chroniques et n'enregistrent pas de données RCM.*

Pour les études de prévalence hospitalière, cette seule condition est suffisante. La prévalence hospitalière n'est pas souvent représentative de la prévalence dans la population générale. Cependant, son étude est intéressante pour certaines maladies chroniques particulièrement quand elles sont associées à des maladies plus aiguës. Dans ce cas, la prévalence hospitalière attendue est plus importante que la prévalence dans la population générale.

*La prévalence hospitalière du diabète a été associée dans notre étude à l'incidence de l'infarctus du myocarde*

Pour les études d'incidence sur la population générale, il est nécessaire que les patients atteints soient presque systématiquement hospitalisés lors de la découverte de leur maladie.

*Les études d'incidence seront faites plus facilement sur les maladies aiguës, comme l'infarctus du myocarde ou la méningite*

Le calcul de l'incidence peut être biaisé quand les patients poursuivent parfois leur traitement dans plusieurs hôpitaux différents (Cfr. Paragraphe "Prestations" ci-dessous)

*Nous verrons les cas des patientes atteintes d'un cancer du sein qui peuvent être prises en charge par un système de soins assez complexe.*

En résumé, les études d'incidence sont donc plus aisées pour les maladies aiguës, nécessitant un traitement hospitalier unique.

*Les maladies choisies pour nos études d'incidence se rapprochent de ce cas de figure.*

## B. Les prestations

Dans notre analyse, nous avons utilisé les données des prestations médicales et des prescriptions médicamenteuses,

- soit pour détecter la présence des maladies dont l'enregistrement du diagnostic faisait défaut  
*Dans notre étude de la prévalence hospitalière du diabète, nous avons détecté par ce moyen, 24% des patients diabétiques pour lesquels aucun diagnostic de diabète n'avait été enregistré dans l'année. De même, chez les patientes opérées d'un cancer du sein, le diagnostic de la maladie n'était parfois pas noté au moment de l'intervention mais seulement lors d'un séjour antérieur.*
- soit pour marquer l'incidence d'une maladie si celle-ci nécessite un traitement chronique.  
*Le diagnostic de cancer du sein ou de la thyroïde, corroboré par une intervention chirurgicale curative, marque le plus souvent la découverte de la maladie.*

En effet, sous le format actuel des RCM, il n'est pas possible de distinguer une maladie nouvellement diagnostiquée à l'hôpital d'un diagnostic établi à domicile ou durant une hospitalisation antérieure.

Si nous considérons la présence des prestations médicales ou du traitement médicamenteux comme un test de la présence de la maladie, une prestation est définie comme tout à fait sensible si elle est toujours utilisée pour traiter la maladie étudiée, une prestation est considérée comme tout à fait spécifique si elle n'est utilisée que pour traiter cette maladie et aucune autre.

La prestation marquant l'incidence de la maladie doit donc être spécifique. Si cette prestation est à la fois spécifique et sensible, son seul enregistrement peut suffire à marquer l'incidence, même si le diagnostic n'est pas enregistré.

*Un exemple est la cure radicale d'hypospadias.*

Nos études d'incidence adoptant cette méthodologie sont cependant influencées par la qualité des soins:

- Il peut exister une sous-estimation quand le malade ne bénéficie pas du traitement classique
- Il peut exister une surestimation quand le malade est traité dans des hôpitaux différents ou dans des périodes annuelles différentes

Pour évaluer l'importance de celle-ci, il est intéressant de calculer la proportion d'interventions multiples, sur les différents séjours des mêmes patients dans le même hôpital. Si cette proportion est importante, la possibilité de doublons d'une année à l'autre ou d'un hôpital à l'autre peut être aussi conséquente.

*Pour le cancer du sein par exemple, les « faux négatifs » représentent certaines femmes plus âgées ne pouvant être opérées et les "faux positifs" sont les femmes qui subissent deux interventions, par exemple une tumorectomie suivie d'une mastectomie dans deux hôpitaux différents ou durant une période à cheval sur deux années.*

- Selon les directives et procédures d'enregistrement

L'objectif financier du traitement des données RCM conditionne les directives pour leur enregistrement:

Le diagnostic principal doit être en relation avec le motif d'admission et l'éventuelle procédure chirurgicale. Ces directives d'enregistrement orientent souvent le choix du diagnostic principal vers les maladies les plus aiguës. Dès lors, un diagnostic important est parfois refoulé au rang des diagnostics secondaires en faveur d'un autre diagnostic plus représentatif des dépenses économiques ou de la durée du séjour, quelques logiciels d'encodage permettant cette analyse. Certaines maladies, principalement les maladies chroniques, apparaissent plus souvent en diagnostics secondaires.

Nos études montrent que tout diagnostic (principal et secondaire) doit être inclus dans l'analyse épidémiologique pour que celle-ci soit valable.

Pour les diagnostics secondaires, si l'enregistrement d'une maladie n'influence pas le classement du séjour en DRG, le risque qu'elle ne soit pas enregistrée peut être conséquent.

*Il paraît par exemple hasardeux de vouloir étudier la prévalence hospitalière de l'hypertension artérielle ou de l'obésité comme nous l'avons testé chez les patients diabétiques.*

De plus, la classification AP-DRG adoptée jusque récemment en Belgique, ne considérait qu'un seul diagnostic secondaire dans l'algorithme. Elle ne favorisait donc pas l'encodage de tous les diagnostics secondaires. De ce fait aussi, l'enregistrement d'un diagnostic peut faire défaut. C'est le cas spécialement pour les maladies chroniques.

Cependant, l'analyse de tous les séjours regroupés par patient permet de récupérer certains diagnostics qui ne seraient pas détectés sur une analyse séparée de ces séjours.

*Dans l'étude de prévalence du diabète, cette procédure nous a permis de détecter un nombre important de complications chroniques qui n'étaient pas enregistrées sur chaque séjour.*

- Selon le système de codification

La classification ICD-9-CM ne permet pas toujours d'identifier précisément toute maladie, surtout s'il s'agit de maladies rares. La recherche d'une prestation spécifique est parfois plus précise dans la codification propre à l'INAMI

*Pour l'étude de prévalence de l'hypospadias, La sélection des séjours était basée sur la codification de l'INAMI, l'ICD-9-CM ne distinguant pas l'hypospadias de l'épispadias, ni par le diagnostic, ni par la procédure chirurgicale.*

L'ICD-9-CM distingue rarement le grade ou la sévérité d'une maladie. Les codes morphologiques (codes M) sont répertoriés mais ne sont pas utilisés systématiquement.

*Il existe un code ICD-9-CM pour distinguer le cancer in situ dans le sein mais pas spécifiquement dans la thyroïde. De plus, d'après notre analyse, la classification morphologique n'est enregistrée que dans 34% des cas pour ce cancer. De ce fait, dans notre étude d'incidence, la constatation du nombre croissant d'interventions chirurgicales pour traiter le cancer thyroïdien ne permet pas de conclure à une augmentation de l'incidence de la maladie.*

L'examen détaillé de la classification ICD-9-CM avant la sélection du fichier permet de sélectionner non seulement les séjours mentionnant le code de la maladie mais aussi tout séjour incluant un code d'un diagnostic ou d'un traitement spécifiquement associé.

*Le diagnostic de diabète a pu par exemple être retenu grâce à l'enregistrement du seul code de la rétinopathie diabétique.*

- En résumé, voici les questions indispensables auxquelles il faut répondre avant d'envisager l'étude d'une maladie sur les données hospitalières :

En rapport avec le système de référence de l'hôpital :

- Le malade atteint est-il parfois hospitalisé ?
- Est-il systématiquement hospitalisé lors de la découverte de sa maladie?
- Est-il susceptible d'être transféré dans un autre hôpital pour des soins ultérieurs ?

En rapport avec les prestations médicales :

- Existe-t-il des prestations sensibles\* et spécifiques\*, pouvant marquer l'incidence de la maladie?
- La maladie est-elle aiguë ? Nécessite-t-elle des prestations médicales uniques ou multiples ?

En rapport avec l'enregistrement des données :

- L'enregistrement de la maladie et/ou des prestations ont-ils une influence sur le classement du séjour en DRG?
- Existe-t-il des codes (ICD-9-CM et INAMI) sensibles\* et spécifiques\* ?

- Une réponse négative à la première question ci-dessus rend évidemment l'étude impossible, une réponse négative à une des questions suivantes peut limiter l'étude. Rappelons ici les points essentiels de la méthodologie qui inclut certaines procédures pouvant surmonter partiellement ces limites:

1. En préalable à la sélection du fichier, l'examen détaillé de la classification ICD-9-CM permet de sélectionner non seulement les séjours mentionnant le code de la maladie mais aussi tout séjour incluant le code d'une maladie ou d'un traitement spécifique qui lui est associé.
2. Tout diagnostic (principal ou secondaire) doit être inclus dans l'analyse.
3. L'analyse de tous les séjours regroupés par patient au sein d'un hôpital permet de récupérer certains diagnostics qui ne seraient pas détectés sur une analyse séparée de ces séjours.
4. Un médicament spécifique\* ou une prestation médicale spécifique\* permet parfois de détecter un diagnostic dont l'enregistrement fait défaut sur RCM.
5. Une prestation médicale spécifique\* peut marquer l'incidence d'une maladie, particulièrement dans le cas de certaines maladies chroniques.

---

\* Une prestation médicale est considérée comme tout à fait spécifique par rapport à la maladie étudiée si elle n'est utilisée que pour traiter cette maladie et aucune autre,

\* Une prestation est considérée comme tout à fait sensible si elle est toujours utilisée pour traiter la maladie étudiée.

### III. L'infarctus du myocarde

L'infarctus du myocarde est la seule maladie que nous avons choisie pour valider les données du Résumé Clinique Minimum en les comparant aux données d'une base de référence, en l'occurrence la base MONICA. Pour les autres maladies, nous aurons un objectif de recherche épidémiologique particulier qui sera présenté en début de chapitre.

#### Matériel et méthode

La méthodologie de cette étude se base sur celle utilisée dans une étude finlandaise de validité des données hospitalières sur les maladies coronariennes {1}

La validité des données épidémiologiques du RCM est analysée en prenant comme référence les données cardio-vasculaires du Registre Monica luxembourgeois. L'analyse statistique est réalisée à l'aide des indicateurs de sensibilité, de spécificité et des valeurs prédictives positive et négative.

Le Registre Monica récolte les données relatives à tout accident coronarien survenu dans la province du Luxembourg en Belgique. Les informations proviennent des hôpitaux et des médecins généralistes. Le Registre Monica récolte les données de diagnostic, de symptômes, de tests effectués (ECG, dosage des enzymes sériques, autopsie), de la présence d'antécédent d'infarctus chez le patient, etc.

La méthode consiste à comparer les données des patients « au cas par cas ». Pour chaque patient, la présence ou non d'un diagnostic d'infarctus aigu du myocarde est observée dans les deux bases de données. L'analyse statistique utilisée est celle de la mesure de la sensibilité et de la spécificité du RCM.

Dans le RCM, les codes de diagnostics sont issus de la classification internationale des maladies et des opérations 9<sup>ème</sup> révision (ICD-9-CM). Les codes spécifiant l'infarctus aigu du myocarde sont les codes 410.0 à 410.9.

Dans le Registre Monica, le patient est considéré comme ayant un infarctus aigu du myocarde certain dans plusieurs cas :

- L'ECG confirme le diagnostic du médecin
- L'ECG rend probable le diagnostic et les enzymes sériques sont anormales
- Les enzymes sériques sont anormales et les symptômes sont typiques
- L'autopsie est positive

La sélection des patients dans les deux bases de données est effectuée sur les critères suivants :

- Années 1994 à 1996
- Domicile du patient situé en province du Luxembourg, Belgique
- Patients âgés de 25 à 74 ans

De plus, les patients de la base de donnée RCM sont sélectionnés suivant la présence des codes ICD-9-CM de diagnostic 410\* à 414\*. Ces codes peuvent apparaître en diagnostic principal du séjour ou en diagnostic secondaire. Ces codes représentent les pathologies suivantes :

- 410\* : infarctus aigu du myocarde
- 411\* : cardiopathies ischémiques aiguës et subaiguës, autre
- 412\* : infarctus anciens du myocarde
- 413\* : angines de poitrine, angors
- 414\* : cardiopathies ischémiques chroniques, autres

Dans le Registre Monica, les patients sélectionnés sont ceux ayant eu un événement coronarien et ayant été hospitalisés.

La méthodologie consiste ensuite à fusionner les deux bases de données et à retrouver dans le Registre Monica les patients correspondant aux patients du RCM.

Les variables utilisées pour la fusion des deux bases de données sont les suivantes:

- L'hôpital de traitement
- L'année et le mois de l'incident coronarien ou de l'admission à l'hôpital
- L'âge
- Le sexe
- Le code d'arrondissement du domicile (code de l'Institut National de Statistique)

L'analyse statistique de la validité épidémiologique du RCM est réalisée à l'aide des indicateurs de sensibilité, de spécificité et de valeur prédictive positive ou négative. Pour chaque patient, la présence ou non d'un diagnostic d'infarctus aigu du myocarde est observée dans les deux bases de données. Le Registre Monica est considéré comme la référence.

La validité intrinsèque du RCM est sa capacité à détecter un infarctus aigu du myocarde lorsque ce patient présente effectivement cette pathologie suivant le Registre Monica et à ne pas détecter d'infarctus lorsque le patient n'en a pas. Les indicateurs qui mesurent la validité intrinsèque d'un test sont la sensibilité et la spécificité<sup>1</sup>.

La valeur prédictive du RCM est la probabilité que le patient ait selon le Registre Monica un infarctus aigu du myocarde lorsque ce diagnostic est noté dans le RCM ou la probabilité que le patient n'ait pas d'infarctus aigu du myocarde si le diagnostic n'est pas codé dans le RCM. Les indicateurs qui mesurent la valeur prédictive d'un test sont la valeur prédictive positive et la valeur prédictive négative<sup>2</sup>.

## Résultats

La base de données RCM sélectionnée comprend 6794 séjours hospitaliers tandis que la partie du Registre Monica sélectionnée correspond à 830 événements coronariens. Après fusion entre les deux bases de données, 561 patients ont été sélectionnés.

Le taux de fusion est peu élevé. 68% des incidents coronariens du Registre Monica ont retrouvé leur séjour hospitalier correspondant dans le RCM. Ce faible pourcentage s'explique pour plusieurs raisons.

A. Les variables de fusion ne sont pas suffisamment discriminantes.

Dans la base de données RCM, plusieurs séjours peuvent avoir la même combinaison de valeurs pour les variables de fusion. Plusieurs patients peuvent être hospitalisés le même mois de la même année, dans le même hôpital pour des pathologies appartenant aux diagnostics 410\* à 414\* et qui, de plus, ont le même sexe, le même âge et sont domiciliés dans le même arrondissement INS. 13 % des séjours ont donc été retirés des 6794 séjours du RCM car ils formaient des doublons. Pour pallier ce problème, il aurait été intéressant de disposer de variables de fusion plus précises comme le jour d'hospitalisation et le code postal du domicile du patient. Ces données ne sont pas fournies pour des raisons de confidentialité.

---

<sup>1</sup> Sensibilité : capacité d'un test de donner un résultat positif quand la maladie est présente

Spécificité : capacité d'un test de donner un résultat négatif quand la maladie est absente

<sup>2</sup> Valeur prédictive positive : la probabilité que la maladie soit présente lorsque le test est positif,

Valeur prédictive négative : la probabilité que la maladie soit absente lorsque le test est négatif.

B. L'âge du patient n'est pas calculé de manière identique dans les deux bases de données

L'âge des patients du Registre Monica est calculé en fonction de la date de naissance du patient et de la date de l'incident coronarien. Par contre, le RCM ne dispose pas, pour des raisons de confidentialité, de la date de naissance des patients. L'âge est calculé en considérant que tous les patients sont nés un 1 juillet. L'âge du RCM ne correspond donc pas toujours à celui du Registre Monica.

Le tableau suivant reprend la distribution des patients suivant la présence ou non du diagnostic 410\* dans le RCM et la présence ou non d'un infarctus certain dans le Registre Monica.

|     |                  | Registre Monica |                 |       |
|-----|------------------|-----------------|-----------------|-------|
|     |                  | Infarctus       | Pas d'infarctus | Total |
| RCM | Code 410*        | 349             | 72              | 421   |
|     | Pas de code 410* | 52              | 88              | 140   |
|     | Total            | 401             | 160             | 561   |

Tableau II: *Distribution des patients suivant la présence ou l'absence du diagnostic 410\* dans les deux bases de données*

En prenant le Registre Monica comme référence, nous obtenons pour le RCM :

- Sensibilité = 87 % (349/401)
- Spécificité = 55 % (88/160)
- VPP = 83 % (349/421)
- VPN = 63 % ((88/140)

Le pourcentage de faux positifs, c'est-à-dire la proportion de patients non malades parmi les patients présentant le code de diagnostic dans leur RCM, est de 17 %.

Lorsque nous analysons les 72 patients faux positifs, nous remarquons que le Registre Monica indique pour 55 d'entre eux que le diagnostic d'infarctus n'est pas certain mais est possible. Ces patients présentent par exemple des symptômes typiques d'un infarctus mais l'électrocardiogramme et les enzymes sériques ne confirment pas le diagnostic.

Le pourcentage de faux négatifs, c'est-à-dire la proportion de patients malades parmi les patients ne présentant pas le code de diagnostic dans leur RCM, est de 37 %.

Lorsque nous analysons les 52 patients faux négatifs, nous remarquons que la moitié de ces patients ont été hospitalisés dans le même hôpital et qu'ils présentent dans deux tiers des cas le code 414\* des cardiopathies ischémiques chroniques. Ces faux négatifs proviennent peut-être en partie d'un problème de codification dans cet hôpital.

## Discussion et conclusion

En ce qui concerne la validité intrinsèque du RCM, une sensibilité de 87 % est un bon résultat. Chez 87 % des patients ayant un infarctus, le RCM détecte cette pathologie.

La spécificité du RCM est quant à elle de 55 %. Pour 55 % des non malades, le RCM ne code pas de diagnostic d'infarctus et donc pour 45 % des non malades le RCM code le diagnostic d'infarctus (faux positif). Comme vu précédemment, cette forte proportion de faux positifs est peut-être due à un diagnostic trop rapide de l'infarctus aigu du myocarde se basant sur les symptômes typiques de l'infarctus mais sans confirmation par un ECG ou l'analyse des enzymes sériques.

En ce qui concerne la validité prédictive du RCM dans le diagnostic de l'infarctus aigu du myocarde, la valeur prédictive positive est de 83 %, ce qui est un bon résultat. Dans 83 % des cas où le diagnostic 410\* est présent dans le RCM, le patient a un infarctus selon la référence Monica. Cependant, cette forte valeur prédictive positive peut s'expliquer par la forte prévalence de la pathologie dans la population étudiée<sup>3</sup>. Dans la population du Registre Monica, la prévalence de l'infarctus est de 71 %.

La valeur prédictive négative est plus faible, pour 63 % des patients ne présentant pas de code de diagnostic dans le RCM la maladie est bien absente suivant le Registre Monica. La proportion de faux négatifs est donc de 37 %. Comme vu précédemment dans les résultats, cette forte proportion pourrait être due à une erreur de codification.

En conclusion, en prenant comme référence le Registre Monica, l'enregistrement RCM est sensible à 87 % pour l'analyse épidémiologique de l'infarctus aigu du myocarde mais est seulement spécifique à 55 %.

## Références bibliographiques

Mähönen M, Salomaa V, Brommels M et al. The validity of hospital discharge register data on coronary heart disease in Finland. *European Journal of Epidemiology* 1997, 13: 403-415.

## Collaborateurs

M. Jeanjean, Association inter universitaire pour la prévention des maladies cardio-vasculaires, Ecole de santé Publique, Université Catholique de Louvain

---

<sup>3</sup> A sensibilité et spécificité égales, lorsque la fréquence de la maladie dans la population étudiée augmente, la valeur prédictive positive augmente également.

## IV. Le cancer du sein chez la femme

### Objectif particulier de l'étude

Le cancer du sein est la première cause de mortalité par cancer chez la femme. Depuis quatre décennies, les statistiques révèlent un nombre croissant de femmes décédant de ce cancer. Actuellement, 3% de femmes meurent de cette maladie avant l'âge de 75 ans {1}. Heureusement, depuis les années 80, la mortalité standardisée par âge sur la population européenne, a baissé en Belgique comme dans plusieurs pays européens.

Depuis une quinzaine d'années, le Registre National du Cancer n'a pas montré d'augmentation d'incidence de la maladie parallèle à l'accroissement du nombre de décès. Néanmoins le nombre de prestations chirurgicales spécifiques du cancer du sein, enregistré par l'INAMI, laisse supposer une augmentation régulière du nombre de patients opérés {2}, comme dans la plupart des pays occidentaux.

Ces constatations discordantes nous ont amené à tester l'enregistrement des données hospitalières pour étudier l'incidence de cette maladie. En 1996, le Registre National débutait de nouvelles procédures d'enregistrement pour la région flamande. Nous avons saisi cette opportunité pour comparer les résultats de nos analyses respectives pour les années 1996 et 1997.

D'autres études récentes dans nos pays voisins ont eu la même initiative à un niveau plus local avec des résultats assez probants {3, 4}

### Matériel et méthodes

Le concept de diagnostic, enregistré sur les données hospitalières, ne nous permet pas de distinguer un nouveau diagnostic d'un diagnostic posé antérieurement. Notre étude d'incidence a dès lors été réduite à celle des patientes opérées.

Dans une première approche de l'analyse, nous avons d'abord confronté les données des interventions (INAMI ou sur RCM) avec les diagnostics (sur RCM), pour essayer de calculer, une valeur prédictive positive (VPP\*) pour chaque intervention et de savoir dans quelle mesure les variations du nombre de prestations pouvaient être utilisées comme témoin des variations de l'incidence de la maladie.

Nous avons sélectionné, parmi les données des RCM des hôpitaux accueillant les patients en séjours aigus en 1996 et en 1997, tous les séjours pour lesquels avaient été enregistrés au moins un des diagnostics ou une des prestations médicales correspondant aux codes suivants :

- pour les diagnostics, en ICD-9-CM: 174\* Tumeur maligne du sein chez la femme, 2330 Carcinome du sein in situ, 2383 Tumeur du sein de comportement incertain ou autre
- pour les prestations médicales,
  - en ICD9-CM : 85.2\* Excision ou destruction de tissu du sein, 85.4\* Mastectomie, 40.23 Excision de ganglion axillaire, 40.51 Excision radicale de ganglion axillaire
  - en nomenclature INAMI : 227032-227043 Exérèse d'1 tumeur ou d'un kyste de la glande mammaire, 227010-227021 Exérèse d'1 tumeur située au-dessus du fascia dans les parties molles mais avec résection totale de l'organe dans lequel se situe la tumeur, 226936-226940 Evidemment ganglionnaire de l'aisselle, 227054-227065 Mammectomie partielle ou

---

\* « VPP » pour une prestation chirurgicale = nombre de séjours mentionnant à la fois l'intervention et le diagnostic de cancer divisé par le nombre total des séjours mentionnant l'intervention.

tumorectomie associée à un curage ganglionnaire axillaire, 226951-226962 Intervention selon Urban, 226973-226984 Intervention selon Halsted ou Pattey avec examen anatomo-pathologique extemporané, 226995-227006 Intervention selon Halsted ou Pattey.

Sur RCM, le calcul d'incidence concerne les patients pour lesquels ont été enregistrés au moins une fois dans l'année un des codes 174\* ou 2330 pour le diagnostic ainsi qu'un des codes opératoires ci-dessus. Les patients sont définis dans chaque hôpital par le numéro anonyme, l'âge et le sexe et l'arrondissement du domicile.

La sélection inclut tous les diagnostics et tous les types de séjours, quelle que soit leur durée.

Pour mettre en évidence les variations géographiques de l'incidence, nous avons calculé, pour chaque arrondissement, un rapport d'incidence standardisée (incidence observée / incidence standardisée) suivant la méthode de standardisation indirecte et nous avons appliqué la correction de Bonferroni pour interpréter la signification des résultats (correction pour tests multiples).

## Résultats

Le nombre de patientes pour lesquelles on relève sur RCM, un diagnostic de cancer du sein et une intervention sur le sein dans la même année est impressionnant: 7712 patientes en 1996 et 7805 patientes en 1997. Ces chiffres correspondent à un taux brut d'incidence annuelle de 149 pour 100 000 femmes en 1996 et de 150 en 1997.

La répartition géographique des cas par arrondissements est très hétéroclite mais varie peu de 1996 à 1997. La région la plus touchée pour les deux années est le centre du pays: les arrondissements de Bruxelles, Halle, Nivelles (Fig. S1)

L'incidence de la maladie augmente régulièrement à partir de l'âge de 30 ans jusqu'à l'âge de la ménopause (Fig. S3). Après 50 ans, l'incidence reste élevée et augmente encore après l'âge de 75 ans.

Les patientes étaient opérées le plus souvent par mastectomie. En 1997, 53% des patientes subissaient cette intervention. Parmi celles-ci, 13% ont été opérée deux fois dont 9% d'une tumorectomie suivie d'une mastectomie dans le même hôpital et durant la même année. Le taux de mastectomie était un peu plus important en Wallonie (56%) qu'en Flandre (51%) ou dans la région bruxelloise (54%). Ce taux était aussi nettement plus élevé chez la femme âgée (Fig. S4).

Les proportions des opérations pour cancers in situ sont assez basses (3% en 1996, 4% en 1997). Elles semblent un peu plus élevées pour certaines régions de Flandres, comme le Limbourg ou la province d'Anvers, où l'incidence globale des cancers opérés est par contre plus basse pour les deux années étudiées (Fig. S5).

La VPP\* des mastectomies, calculée en 1997 est de 95% ; pour les tumorectomies, elle est de 45% (96% pour les tumorectomies avec curage axillaire). 65% des prestations chirurgicales enregistrées comme tumorectomies seraient donc effectuées chez des patientes qui ne sont pas atteintes de cancer.

Si on tient compte de ces résultats, il est possible d'évaluer grossièrement l'incidence à partir du seul nombre d'interventions répertorié dans les bases de données INAMI :

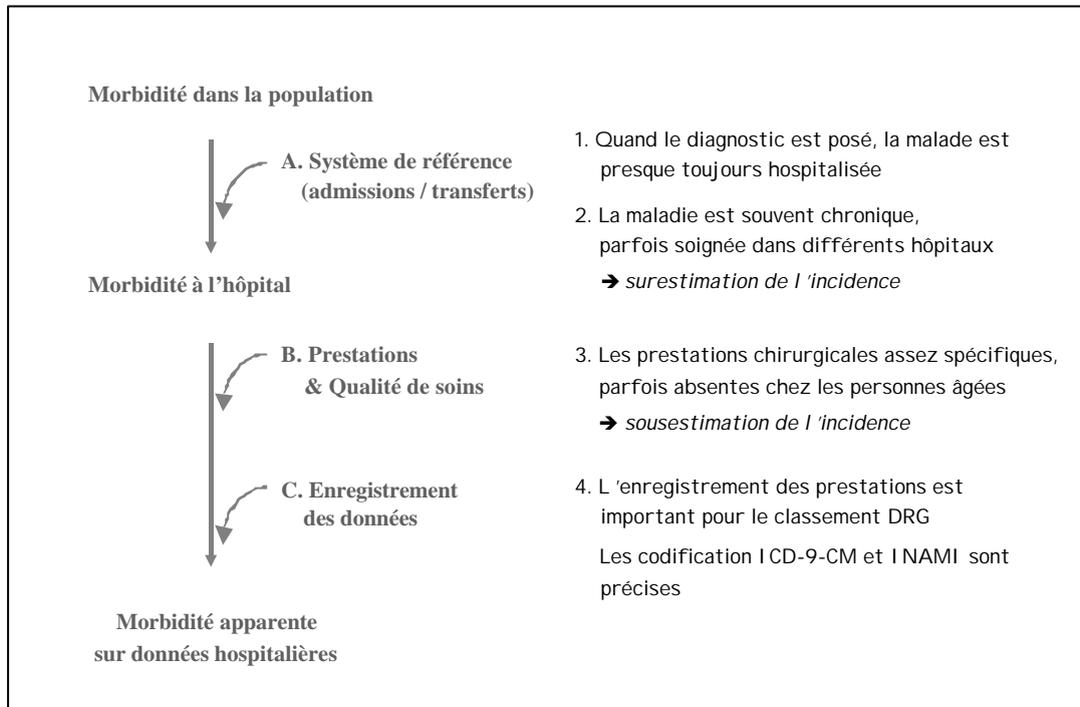
- N mastectomie(s) associé au diagnostic de cancer du sein: 4416 actes x 95% = 4195
  - N tumorectomie(s) associé au diagnostic de cancer du sein : 9767 actes x 45% = 4395
  - N patientes opérées pour cancer du sein : (4195+4395) – 13% d'interventions multiples = 7474
- Le taux d'incidence calculé (144 /100000 femmes) est assez proche de celui calculé sur RCM.

---

\* « VPP » pour une prestation chirurgicale = nombre de séjours mentionnant à la fois l'intervention et le diagnostic de cancer divisé par le nombre total des séjours mentionnant l'intervention.

## Discussion

Les limites théoriques de validité des données hospitalières pour l'étude de l'incidence du cancer du sein sont résumées dans la figure S6 dans le texte ci-dessous.



**Fig. S6** *Limites théoriques de validité des données hospitalières pour l'étude de l'incidence du cancer du sein*

Le cancer du sein, comme le cancer de la thyroïde ou de la prostate, est une maladie insidieuse mais lorsque le diagnostic est posé, le patiente est presque toujours hospitalisé pour subir un traitement chirurgical. L'incidence de la maladie opérée à l'hôpital représente donc bien l'incidence de la maladie diagnostiquée dans la population générale.

La maladie peut être longtemps latente. Le délai entre l'apparition du cancer et son diagnostic, conditionnant fortement le pronostic vital de la patiente, peut varier en fonction de l'accès aux procédures diagnostiques, principalement à la mammographie. Cet état de fait explique la répartition géographique très hétéroclite que nous présente la carte d'incidence annuelle sur la Belgique (Fig. S1).

L'incidence annuelle dans une région peut en effet changer brusquement d'une année à l'autre lors du démarrage d'une campagne de dépistage.

Notre analyse des données RCM révèle une incidence nettement supérieure à celle enregistrée par le Registre National. La figure S2 compare par arrondissement, les enregistrements respectifs pour l'année 1997. Elle met en évidence une différence nette entre le Nord et le Sud du pays. Pour la région wallonne, le nombre de cancers opérés, enregistrés sur le RCM est beaucoup plus important; pour la région flamande, le Registre National révèle un nombre de patientes atteints pratiquement identique à celui des patientes opérées enregistrées à l'hôpital (4282 sur le registre versus 4283 sur les RCM) mais il existe des discordances parfois nettes pour chaque arrondissement. Une explication au moins partielle de ces différences peut être trouvée dans la figure S3.

Dans ce graphique présentant les distributions d'incidence par âge, les différences entre notre analyse et celle du registre flamand semblent minimales. Si on prend le registre flamand comme référence, on peut considérer comme :

- "faux négatifs", les cas enregistrés par le registre et non sur les données RCM. Ce sont les femmes plus âgées qui ne bénéficient pas du traitement chirurgical mais sont traitées plus souvent par un traitement médicamenteux chronique.
- "faux positifs", les cas enregistrés sur les RCM et non par le registre. Nous avons vu (Fig. S4) que ces doubles interventions (tumorectomie suivie de mammectomie) étaient déjà plus fréquentes chez les plus jeunes patientes dans un même hôpital. Il est fort probable que les patientes les plus jeunes sont aussi plus souvent opérées dans des hôpitaux différents. L'anonymisation des données avant l'envoi au ministère empêche le suivi de ces patientes d'un hôpital à l'autre, elles seront donc comptabilisées deux fois dans l'analyse.

Le cancer in situ est un cancer diagnostiqué à un stade précoce. Sa proportion sur l'ensemble des cancers peut donc être considérée comme un indicateur de qualité du dépistage de la maladie. D'après notre analyse sur RCM, ce taux serait de 3,1 % en 1996 à 4,0% en 1997. Les taux révélés par le Registre National pour la Flandre et le profil de ces taux par âge semblent peu différents (3,4% en 1996 et 4,7% en 1997) (Fig. 6).

Certains registres en France affichent un taux de cancers in situ supérieur à 10% et aux Etats-Unis, la proportion se situerait entre 10 et 25%. La campagne de dépistage élèvera probablement notre taux de cancers in situ au rang de certains pays voisins. Quoique légèrement sous-enregistré sur les données hospitalières, l'analyse régulière de ce taux sur RCM permettrait ce suivi, principalement pour les populations entre 50 et 69 ans, concernées par le nouveau programme de dépistage par mammographie.

## Conclusion

L'incidence du cancer du sein en Belgique a peut être atteint un record international. Ce taux estimé à partir des données hospitalières est significativement plus élevé que celui relevé par le Registre National pour ces années 1996-1997 mais il est cependant identique à celui trouvé plus récemment par le nouveau registre flamand {5}.

Nous avons montré aussi la possibilité d'utiliser les données du RCM pour analyser la qualité des soins:

- pour déterminer le type d'interventions subies par les patientes ainsi que le nombre des interventions multiples souvent inutiles
- et surtout pour suivre les taux annuels de cancers in situ. En effet ce taux peut être considéré comme un marqueur de la qualité du dépistage par mammographie, par exemple pour les femmes âgées de 50 à 69 ans, population ciblée par le programme national en cours depuis deux ans.

Notre analyse sur RCM est simple et ne montre pas de problème important de validité. La difficulté la plus importante dans l'estimation de l'incidence semble liée aux interventions multiples dans des hôpitaux différents. Ce problème sera partiellement surmonté quand les données RCM et les données RFM seront liées par un numéro unique pour le patient. Ce traitement permettra alors de suivre le patient d'un hôpital à l'autre mais pas d'une année à l'autre.

Si on ne dispose que des données INAMI, l'utilisation de ces seules données pour suivre l'évolution de l'incidence pourrait être intéressante si l'analyse de la VPP\* et celle des taux d'interventions multiples est répétée régulièrement sur un échantillon représentatif de données RCM car la VPP peut varier en fonction de l'évolution des pratiques médicales.

En conclusion, il semble intéressant d'exploiter les données RCM en routine pour contribuer d'une part, à la surveillance épidémiologique de ce problème prioritaire en santé publique et d'autre part, au suivi diagnostique et thérapeutique qui permettront de contrôler la mortalité spécifique de cette maladie.

## Références bibliographiques

1. Trend of mortality breast cancer in Belgium  
<http://www.iph.fgov.be/epidemiologie/epien/cervixen/breasttrend.pdf>
2. Gilbert M., Delège D., Surveillance de la morbidité par cancer du sein en Belgique : Evolution de l'utilisation des prestations techniques spécifiques de 1988 à 1997. Rapport SESA non publié, Ecole de Santé Publique, UCL
3. Ganry O, Kadi D, Taleb A, Ledoux B, Raverdy N, Dubreuil A –Accuracy of DRGs data for estimation of breast cancer incidence in France. *Annuel EUPHA meeting 2001 – Abstract book.*
4. Schiffano P, Scranci M, Papini P, Agabiti N, Borgia P, Perrucci C A.- Accuracy of hospital discharge data on breast cancer surgery in the Lazio region, Italy. *Annuel EUPHA meeting 2001 – Abstract book.*
5. Gilbert M, Thimus D, Roger-France F, Haelterman M, Camberlain C, Mertens I, Capet F. Contribution of hospital discharge data to epidemiological surveillance – example of breast cancer incidence. *Annuel EUPHA meeting 2001 – Abstract book.*

## Collaborateurs

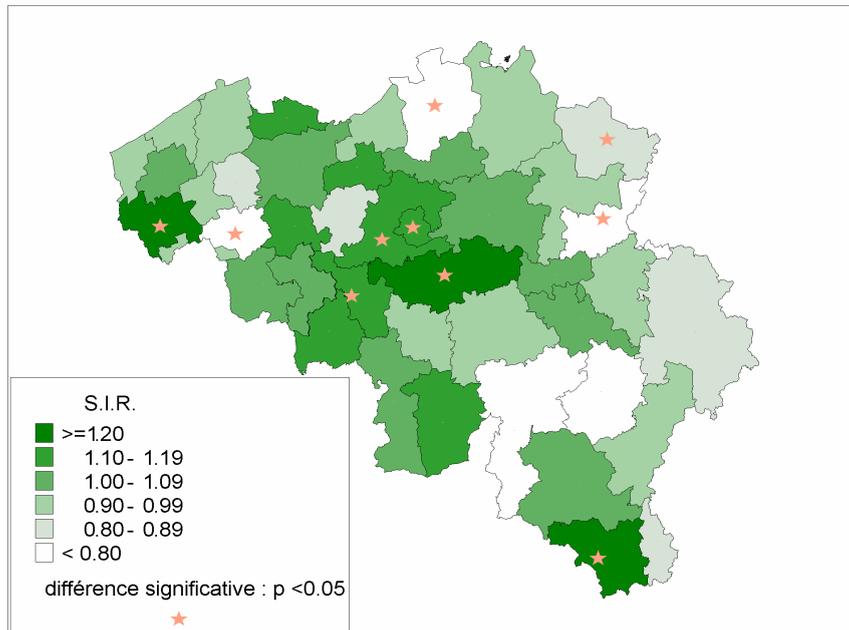
D. Delière, Ecole de Santé Publique, Université Catholique de Louvain - pour l'étude des données de l'INAMI

M. Haelterman, Œuvre Belge du Cancer - pour l'étude des données du Registre National du Cancer.

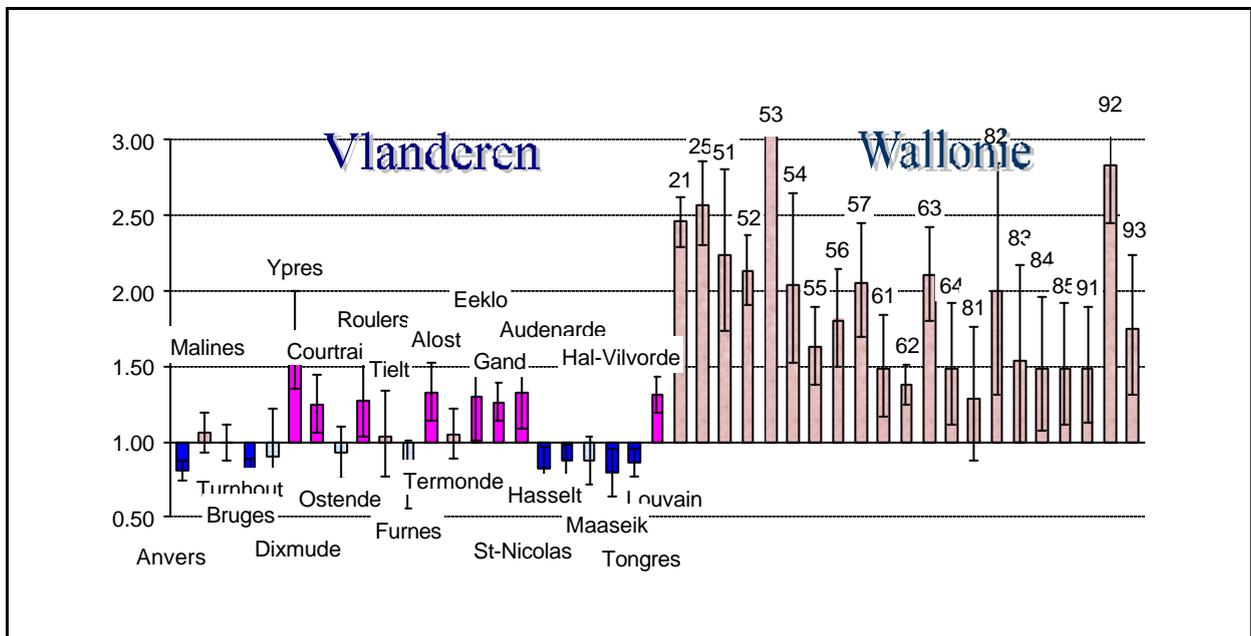
---

\* « VPP » pour une prestation chirurgicale = nombre de séjours mentionnant à la fois l'intervention et le diagnostic de cancer divisé par le nombre total des séjours mentionnant l'intervention.

**Figures illustrant les résultats**



**Fig. S1 Cancer du sein opéré chez la femme - Rapport d'incidence standardisée en 1997 -- Etude nationale sur RCM - Méthode de standardisation indirecte, avec correction de Bonferoni\***



**Fig. S2 Cancer du sein chez la femme en 1997, par arrondissement - Rapport de l'incidence annuelle des cancers opérés relevés sur les RCM, sur l'incidence annuelle des cancers relevés par le Registre National du Cancer (Intervalle de confiance pour un seuil de signification de 95%)**

\* La différence est significative si l'intervalle de confiance du SIR (pour un seuil de signification de 95%) ne contient pas la valeur 1 La correction de Bonferoni est appliquée pour réduire l'effet du hasard dans la signification lors de tests multiples

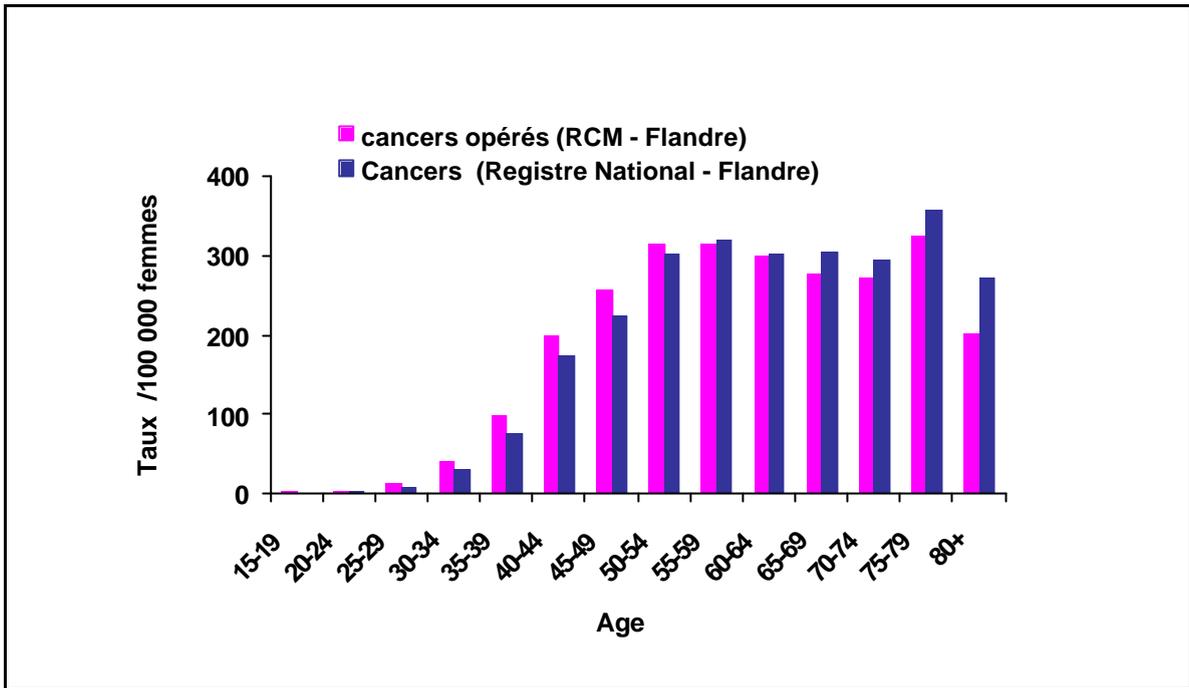


Fig. S3 Incidence annuelle du cancer du sein chez la femme en région flamande 1997. Comparaison entre l'incidence des cancers opérés (sur données hospitalières RCM) et l'incidence des cancers (enregistrés par le Registre National pour la région flamande).

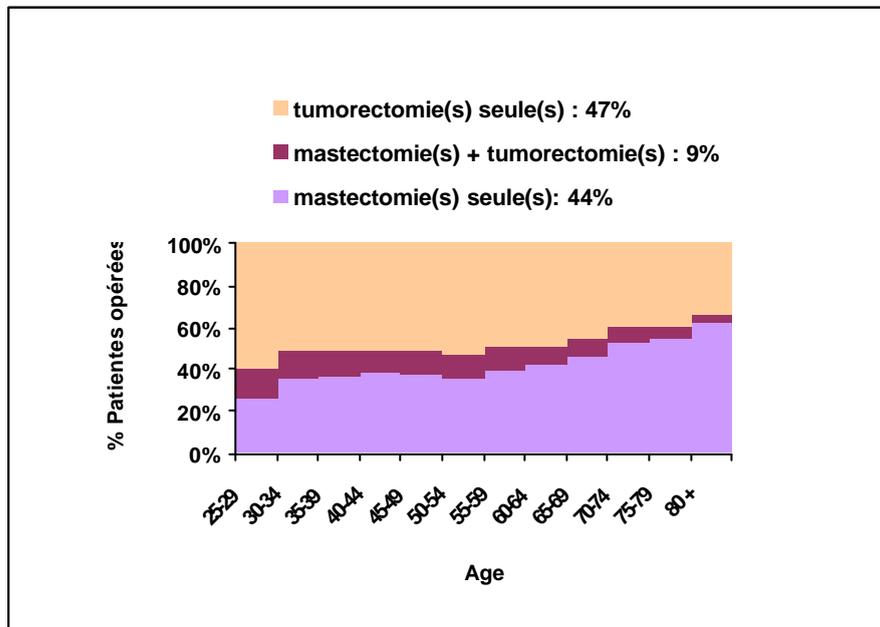


Fig. S4 Traitement chirurgical du cancer du sein chez la femme en 1997 - Types de procédures chirurgicales Etude nationale sur les données hospitalières RCM.

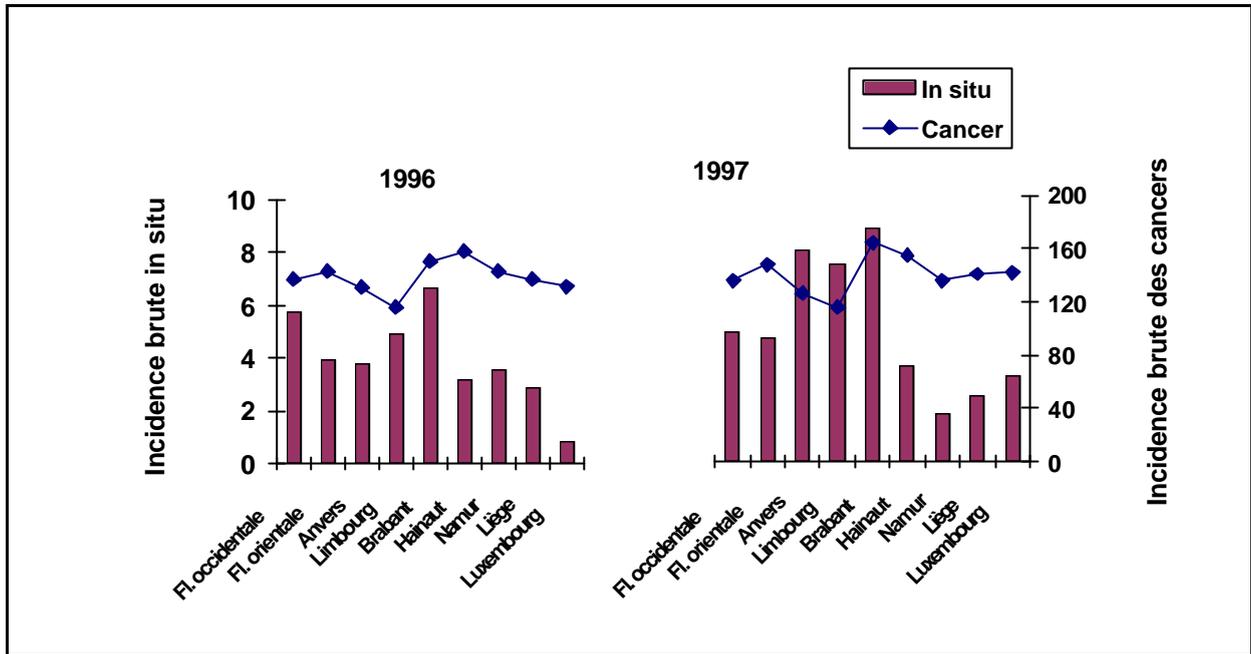


Fig. S5 Cancer du sein opéré chez la femme en 1996 et en 1997 - Comparaison des incidences annuelles brutes des cancers du sein (non spécifié in situ) et des cancers du sein in situ - Etude nationale sur RCM.

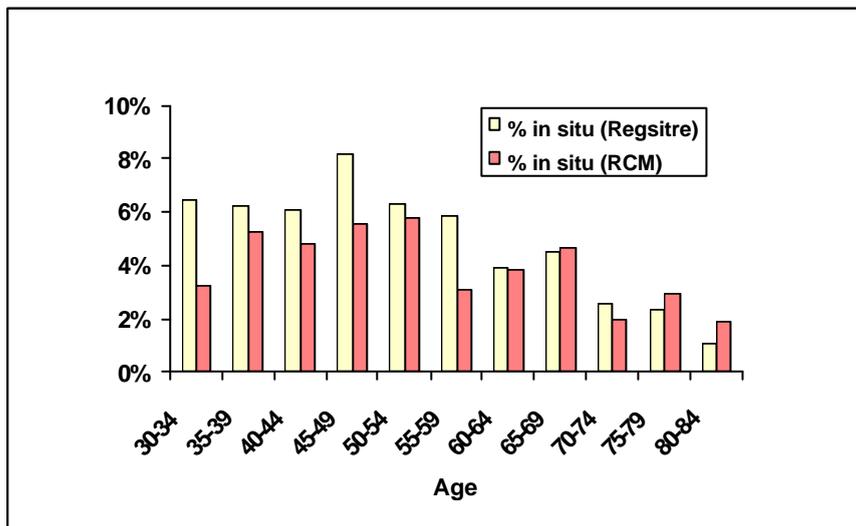


Fig. S6 Proportion des cancers chez la femme déclarés "in situ" en 1997 - Comparaison entre l'enregistrement des cancers du sein opérés, sur les données RCM et l'enregistrement sur des cancers par le Registre National du Cancer pour la Flandre.

## V. Le cancer de la thyroïde

### Objectif particulier de l'étude

L'accident de Tchernobyl survenu le 26 avril 1986 fut responsable d'une épidémie de cancers thyroïdiens en Ukraine et dans les pays avoisinants. Selon le rapport du Belgium Royal Institute of Meteorology, l'exposition de la population belge à la radioactivité consécutive à l'accident de Tchernobyl fut négligeable {1, 2}. La contamination de la Belgique débuta quelques jours après l'explosion, principalement par quelques vagues de pluie d'intensité variable sur le centre et l'Est du pays. Alors que l'essentiel de la contamination provint de l'ingestion d'aliments contaminés par l'iode 131, la zone Est, la plus contaminée ne reçut que 10 000 Becquerels d'Iode 131 par m<sup>2</sup> {Fig. Th1}{3}, soit des doses 500 à 1000 fois inférieures à celles reçues par les enfants au Nord de l'Ukraine et au Sud de la Biélorussie, touchés par l'épidémie de cancers thyroïdiens.

Pourtant, diverses publications parues ces dernières années en Europe font état d'une augmentation localisée du nombre de cancers de la thyroïde et certains d'entre eux soulèvent l'hypothèse d'une influence probable de l'accident de Tchernobyl sur cette augmentation {4, 5}. Aucune étude générale n'a cependant encore été faite à ce propos en Belgique et le Registre National du cancer, notant de sérieuses lacunes d'enregistrement pour cette période, ne permet pas de répondre à cette question très médiatisée.

Notre étude exploite la base nationale des données du Résumé Clinique Minimum (RCM) pour essayer de lever les doutes sur les effets éventuels de la contamination.

### Matériel et méthode

Notre étude d'incidence a été réduite à celle des patients opérés d'un cancer de la thyroïde, le concept de diagnostic, enregistré sur les données hospitalières, ne nous permettant pas de distinguer un nouveau diagnostic d'un diagnostic posé antérieurement.

Nous avons sélectionné, parmi les données des RCM des hôpitaux accueillant les patients en séjours aigus pendant la période 1993 à 1998, tous les séjours pour lesquels avaient été enregistrés au moins un des diagnostics ou une des prestations médicales correspondants aux codes ICD-9-CM suivants :

- pour les diagnostics : 193 Tumeur maligne de la thyroïde
- pour les prestations médicales : 06.2 Lobectomie thyroïdienne unilatérale, 06.3 Thyroïdectomie partielle, 06.4 Thyroïdectomie totale, 06.5 Thyroïdectomie substernale, 06.6 Excision de la thyroïde linguale.

Les codes morphologiques pour identifier le carcinome papillaire étaient:

M805.0, carcinome papillaire, M826.0 adénocarcinome papillaire et M834.0 adénocarcinome papillaire et folliculaire.

Le calcul d'incidence concerne les patients définis dans chaque hôpital par le numéro anonyme, l'âge et le sexe et l'arrondissement du domicile et pour lesquels ont été enregistrés au moins une fois dans l'année le code diagnostique 193 ainsi qu'un des codes opératoires ci-dessus.

La sélection ne se limite pas au diagnostic principal; elle inclut également les diagnostics secondaires. Tous les séjours hospitaliers ont été retenus quelle que soit leur durée.

Pour mettre en évidence les variations géographiques de l'incidence, nous avons calculé, pour chaque arrondissement, un rapport d'incidence standardisée (incidence observée / incidence standardisée) suivant la méthode de standardisation indirecte et nous avons appliqué la correction de Bonferroni pour interpréter la signification des résultats (correction pour tests multiples).

## Résultats

L'analyse de la base de données met en évidence un nombre croissant de patients atteints d'un cancer de la thyroïde traités chirurgicalement entre 1993 et 1998. L'ensemble totalise 1920 patients pour les six années. Le taux standardisé d'incidence par âge (population standard européenne) passe de 2,1 (1.8 - 2.4) pour 100 000 personnes en 1993 à 3,1 en 1995 puis à 3.5 (3.2 - 3.9) en 1998 (Fig. Th2). Ces taux correspondent à une augmentation de 1,4 à 2,4 par 100.000 hommes et à une augmentation de 2,9 à 4,7 par 100.000 femmes.

L'âge médian lors de la découverte de la maladie se situe pour les six années, entre 45 et 49 ans. Le nombre de cancers opérés chez les jeunes de moins de 30 ans représente 9 à 12% du nombre total. Il n'y a pas d'augmentation particulière de la proportion des interventions chez ces jeunes, notamment pour traiter les cancers mentionnant le code morphologique des cancers papillaires (Fig. Th3). L'enregistrement des codes morphologiques n'est cependant pas systématique (44% en moyenne) et varie beaucoup d'un hôpital à l'autre.

Le sex ratio (femmes/hommes) varie entre 2,0 à 3,0 selon les années ; il semble similaire pour les cancers papillaires.

Dans 13% des cas, le patient opéré présenterait des métastases ganglionnaires. Cette fréquence semble deux fois plus élevée chez l'homme (21% versus 9% chez la femme), particulièrement chez les jeunes de moins de 30 ans (28% chez l'homme et 13% chez la femme).

La distribution géographique des patients opérés de cancers thyroïdien en Belgique est deux fois plus dense dans le Sud du pays (la Wallonie et la capitale) (Fig. Th 4). Il en est de même pour les cancers opérés chez les jeunes de moins de 30 ans. Cette différence s'accroît entre 1993 et 1998 : pour le Sud du pays, l'incidence moyenne passe de 3.0 à 5.4 alors que pour le Nord, elle passe de 1.4 à 2.4 en 1995 et reste stable ensuite (Fig. Th5).

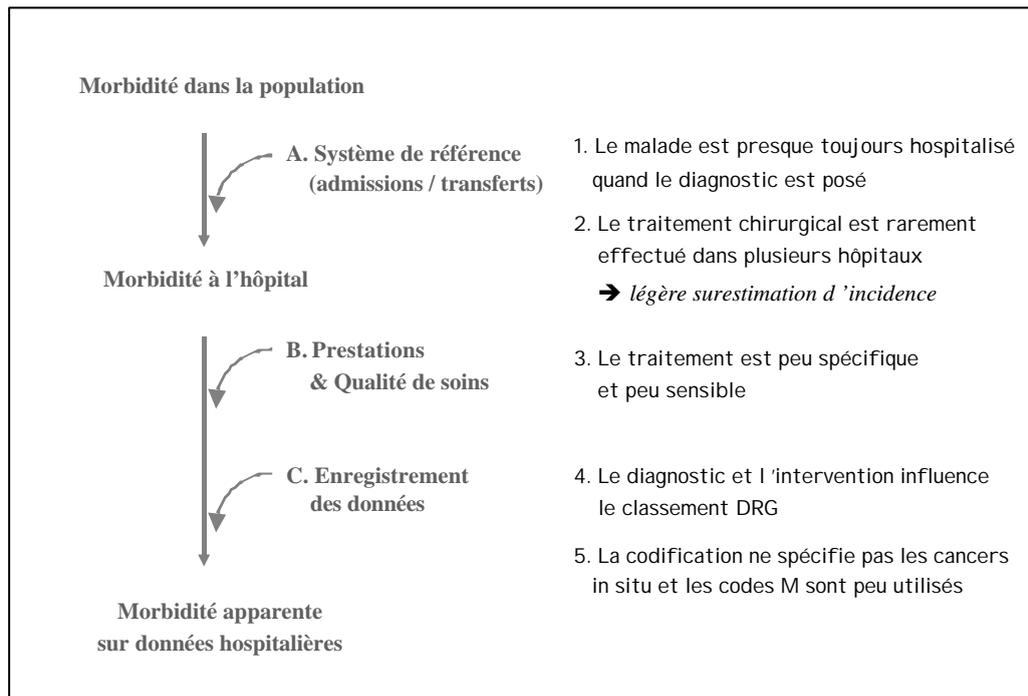
Les patients atteints de cancers représentent 6.9 à 7.9% des patients opérés d'une thyroïdectomie. Les 2/3 de ces cancers sont opérés ou décelés par une ablation totale (Fig. Th6). Le type d'intervention associée au diagnostic de cancer varie peu pendant les 6 années, sauf pour le Sud du pays. La proportion de thyroïdectomie totale y étant nettement plus élevée, s'accroît encore entre 1993 et 1998, passant de 65% à 73% des interventions.

## Discussion

Avant toute interprétation des résultats, il est important de poser les limites théoriques de la validité de notre analyse sur les données hospitalières. Celles-ci sont résumées dans la figure Th7 présentée ci-dessous

D'une part, l'incidence a pu être un peu surestimée par le fait de quelques patients opérés plusieurs fois. En effet, l'anonymisation des données empêche le suivi thérapeutique des patients d'une année à l'autre ou dans d'un hôpital à l'autre. Un patient opéré par exemple dans deux hôpitaux a donc été comptabilisé deux fois dans l'analyse.

D'autre part, une légère sous-évaluation de l'incidence peut être due à des lacunes dans l'enregistrement des données, par exemple, pour certains diagnostics de cancer révélés tardivement après une opération pour goitre. De même, les défauts et l'imprécision de la codification invalident certains résultats. Ces carences sont fréquentes surtout pour la codification des diagnostics secondaires qui n'a souvent aucune implication sur le traitement régulier des données hospitalières dont l'objectif actuel est uniquement financier. C'est pour cette raison que la prévalence des métastases ganglionnaires au moment de l'opération a probablement été nettement sous-évaluée dans notre étude, la plupart des publications faisant mention d'une fréquence plus élevée {6,7}. Pour cette même raison, l'enregistrement des codes histologiques n'est pas systématique. Ce défaut empêche de tirer des conclusions définitives sur l'incidence des cancers papillaires, alors que c'est celle-ci qui est plus spécialement corrélée aux effets des irradiations ionisantes.



**Fig. Th 7 Limites de validité des données hospitalières pour l'étude d'incidence du cancer de la thyroïde opéré**

Enfin, l'enregistrement des données hospitalières ne s'est généralisé en Belgique qu'en 1990. Il est donc probable que cette base de données ne soit pas encore tout à fait exhaustive dans les quelques années consécutives.

Cependant, même en considérant cette dernière remarque importante, les résultats montrent une augmentation du nombre de patients opérés d'un cancer thyroïdien durant les dernières années, comme d'ailleurs le montrent plusieurs études récentes dans d'autres pays {8, 10}.

Il existe de grandes variations dans les incidences relevées dans les différents registres européens. En 1998 par exemple, les taux standardisés par âge variaient de 1.10 par 100 000 personnes en Irlande à 6.44 en Italie. Le taux relevé pour la Belgique se situait alors à 2.62, en dessous du taux moyen européen de 3.99 {9}

Les taux standardisés par âge calculé sur les données RCM est beaucoup plus élevé que celui relevé par le Registre National du Cancer mais est cependant très proche de celui qui est enregistré par le récent registre flamand sur lequel est basé les données publiées par le registre européen IARC pour la Belgique. En 1998, le taux pour la population flamande était sur RCM de 1.7 chez les hommes et 3.4 chez les femmes tandis qu'il était de 1.4 chez les hommes et 3.5 chez les femmes sur le registre flamand.

Les caractéristiques de ces patients semblent similaires à celles découvertes dans diverses études internationales : l'âge médian se situe entre 45 et 50 ans et le sex-ratio (femmes/hommes) varie entre 2.1 et 3.0 selon les années {11, 12}

Les études épidémiologiques notamment sur les conséquences de l'accident de Tchernobyl dans les régions les plus exposées, ont montré que ce sont les enfants qui constituent la population la plus radio sensible {13, 14}. En dehors des épisodes aigus, l'incidence du cancer de la thyroïde chez l'enfant est très faible. En Belgique, le risque d'augmentation d'incidence est mineur et une faible augmentation ne pouvait être pratiquement détectable à partir des données disponibles dans les registres qui ne couvraient pas adéquatement l'ensemble du territoire. Dans notre analyse, l'évolution de l'incidence

des cancers ne montre pas d'augmentation particulière chez les personnes qui étaient âgées de moins de 15 ans au moment de l'accident.

La comparaison de la répartition géographique de l'incidence des cancers opérés en Belgique et de la carte de la contamination en iode radioactif provenant de l'accident de Tchernobyl, plaide aussi contre l'hypothèse d'une influence directe et prioritaire de cet accident sur l'augmentation d'incidence du cancer thyroïdien opéré en Belgique entre 1993 et 1998.

En examinant les pratiques chirurgicales, une interprétation s'impose en priorité: l'augmentation de l'incidence du cancer de la thyroïde diagnostiqué et opéré ne serait-elle pas seulement la conséquence de l'évolution des pratiques diagnostiques et thérapeutiques liées aux maladies thyroïdiennes ? En effet, les cancers thyroïdiens « occultes », de petit diamètre, ont une prévalence élevée, de l'ordre de 3 à 36 % sur les pièces d'autopsie {15}, et certaines publications récentes mettent en évidence l'accroissement de la proportion des stades plus précoce de la maladie {16}.

Il est dès lors probable que ce soit l'intensification des investigations diagnostiques qui ait conduit à une augmentation de l'incidence des cancers détectés. Cette hypothèse a été récemment étayée dans une étude française qui a fait la corrélation entre l'augmentation d'incidence à celle des pratiques diagnostiques notamment par échographies {17}.

D'autre part, l'analyse de la base de données de l'Institut National pour l'Assurance Maladie-Invalidité de 1988 à 1997 révèle que les pratiques opératoires relatives à l'ensemble des affections thyroïdiennes ont évolué durant la dernière décennie. Le nombre total des interventions est assez stable, mais à partir de 1995, la proportion des ablations totales s'accroît et leur nombre représente 59% des actes chirurgicaux sur la thyroïde dès 1996, alors qu'il représentait seulement 40% des actes durant la dizaine d'années précédant cet accroissement {Fig. Th8}. Cette augmentation semble parallèle à celle de l'incidence des patients opérés d'un cancer, calculée sur la base des données hospitalières. On peut donc émettre l'hypothèse que cette évolution de pratique vers l'ablation totale a conduit à la découverte fortuite d'un nombre croissant de cancers thyroïdiens non symptomatiques, contribuant ainsi à l'augmentation de l'incidence. La pratique de l'ablation totale, nettement plus fréquente dans le Sud du pays expliquerait alors partiellement, l'incidence deux fois plus élevée dans cette région.

## Conclusion

Pour répondre à la question médiatique des conséquences de l'accident de Tchernobyl, cette étude tente de mesurer l'incidence du cancer de la thyroïde opéré en Belgique une dizaine d'années après l'explosion. L'analyse a été faite sur les données hospitalières de 1993 et 1998. Malgré ses limites de validité, la recherche de l'incidence de ce cancer sur les données hospitalières montre pour la région flamande, un taux d'incidence similaire au taux publié par le registre flamand ainsi que des caractéristiques épidémiologiques similaires à celles présentées dans d'autres études internationales. Contrairement à la majorité de celles-ci, notre étude offre l'avantage d'une couverture nationale.

Les résultats révèlent une augmentation apparente de l'incidence des cancers opérés mais celle-ci ne concerne pas spécialement la population jeune plus susceptible au moment de l'accident. De plus, la distribution géographique de l'incidence, plus marquée dans le Sud du pays, n'est pas influencée par la contamination en iode radioactif en 1986, plus importante dans l'Est du pays. Ces deux affirmations permettent d'exclure une influence prépondérante de cet accident sur l'augmentation d'incidence du cancer thyroïdien opéré en Belgique entre 1993 et 1998. L'analyse du nombre et du type d'interventions sur la thyroïde suggère que l'augmentation observée puisse être imputée, au moins partiellement, à l'évolution des pratiques médicales

Le cancer de la thyroïde est rare, il représente moins d'1% des tumeurs malignes humaines. De plus, le pronostic de la maladie est en général assez bon. Cette maladie peut cependant être prioritaire en santé publique lors d'un risque de contamination majeure, pour établir un traitement préventif urgent. Un intérêt important de notre étude réside dans la mise en évidence des potentialités des données

hospitalières pour contribuer à la surveillance épidémiologique de ce cancer. Dans les années futures, un complément d'analyse pourrait nous permettre de corroborer les résultats de celle-ci.

## Références bibliographiques

1. De Muer D, De Dycker E, Malcorps H, Trullemans L, Van Der Aowera L, Vandiependeek M. Meteorological aspects of the Chernobyl nuclear accident: consequences for Belgium. Brussels: Belgian Royal Institute of Meteorology, 1990. Publication Series No. 123 {ISSN 0020-255-X}.
2. Uyttenhove J, Pommé S, Van Waeyenberge B, Hardeman F, Buysse J, Culot JP. Survey of the <sup>137</sup>Cs contamination in Belgium by in situ gamma spectrometry, a decade after the Chernobyl accident. *Health Physics*, 1997, 73(4), 644-646.
3. Vandecasteele CM, Coughtrey PJ, Kirchmann R. Impact of the Chernobyl accident on the environment and the management of contaminated areas. *Annales de l'Association belge de Radioprotection*, 1997, 22(1), 59-81.
4. Cotterill SJ, Pearce MS, Parker L. Thyroid cancer in children and young adults in the North of England. Is increasing incidence related to the Chernobyl accident? *Eur J Cancer*, 2001, 37(B), 1020-1026.
5. Blackburn DJ, Michel LA, Rosière A, Trigaux J-P, Donckier JE. Occurrence of Thyroid Papillary Carcinoma in Young Patients. A Chernobyl Connection? *J Pediatr Endocrinol Metab*, 2001, 14(5), 503-506.
6. Tsugawa K, Ohnishi I, Nakamura M, Miwa K, Yokoyama K, Michigishi T, Noguchi M, Nonomura A. Intraoperative lymphatic mapping and sentinel lymph node biopsy in patients with papillary carcinoma of the thyroid gland. *Biomed Pharmacother* 2002, 56 (Suppl 1), 100s-103s.
7. Mazzaferri EL. Thyroid carcinoma: papillary and follicular. In: Mazzaferri EL, Samaan N, eds. *Endocrine tumors*. Cambridge, MA: Blackwell Scientific 1993, 278.
8. Nagataki S, Nyström E. Epidemiology and primary prevention of thyroid cancer. *Thyroid* 2002, 12(10), 889-896.
9. J. Ferlay, F. Bray, R. Sankila and D.M. Parkin. EUCAN: Cancer Incidence, Mortality and Prevalence in the European Union.1998, version 5.0. in: IARC CancerBase No. 4. Lyon, IARC Press, 1999.
10. Black RJ, Bray F, Ferlay J, Parkin DM. Cancer incidence and mortality in the European Union: cancer registry data and estimates of national incidence for 1990. *Eur J Cancer* 1997, 33(7), 1075-107.
11. Görges R. The changing epidemiology of thyroid cancer. In: *Thyroid cancer*. Eds Biersack H-J, Grünwald F. Springer Verlag Berlin Heidelberg New York 2001, 3-25.
12. Schlumberger MJ. Papillary and follicular thyroid carcinoma. *N Engl J Med* 1998, 338, 297-306.
13. Kazakov VS, Demidchik EP, Astakhova LN. Thyroid cancer after Chernobyl. *Nature* 1992, 359, 21.
14. Williams ED, Becker DV, Dimidchik EP, Nagataki S, Pinchera A, Tronko ND. Effects on the thyroid in populations exposed to radiation is a result of the Chernobyl accident . In: *One Decade after Chernobyl: Summing up the consequences of the accident*. International Atomic Energy Agency, Vienna, 1996, 207-230.
15. Nasir A, Chaudhry AZ, Gillespie J, Kaiser HE. Papillary microcarcinoma of the thyroid: a clinico-pathologic and prognostic review. *In Vivo* 2000, 14, 367-376. Review.

16. Schwartz C, Pochart JM, Merle C, Patey M. Thyroid carcinoma: temporal trends for incidence and tumor characteristics from 1975 to 2001, about 1462 cases. *J Endocr Invest* 2002, 25, (Suppl. 7), 52.
17. Leenhardt L, Bernier MO, Boin-Pineau MH et al. Medical practice evolution and thyroid cancer incidence in France. *J Endocr Invest* 2002, 25, (Suppl. 7), 53.

## **Collaborateurs**

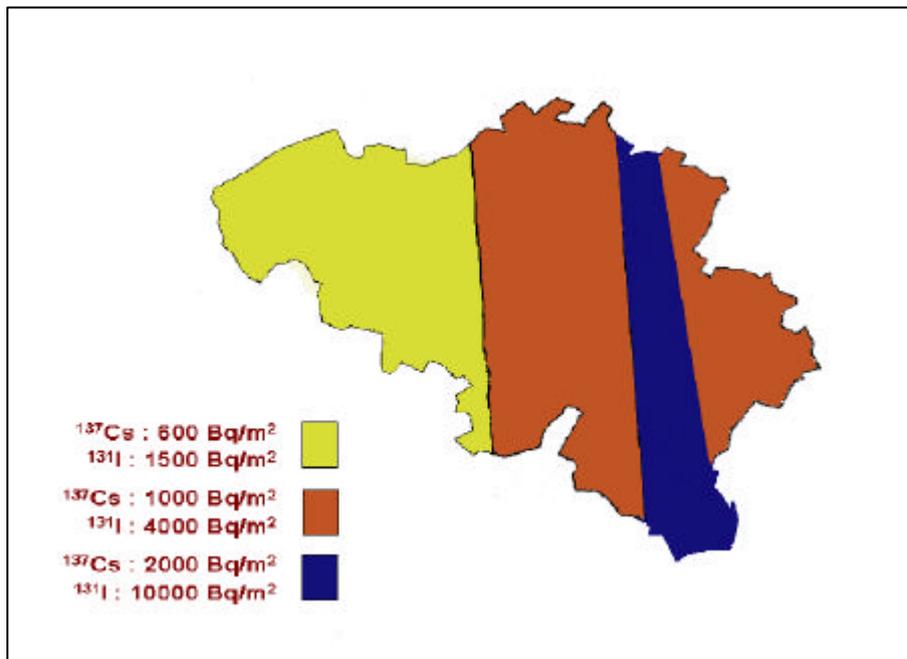
Ch. Daumerie, Département d'endocrinologie, Cliniques Universitaires St-Luc

L. Van Eycken, responsable du Registre flamand

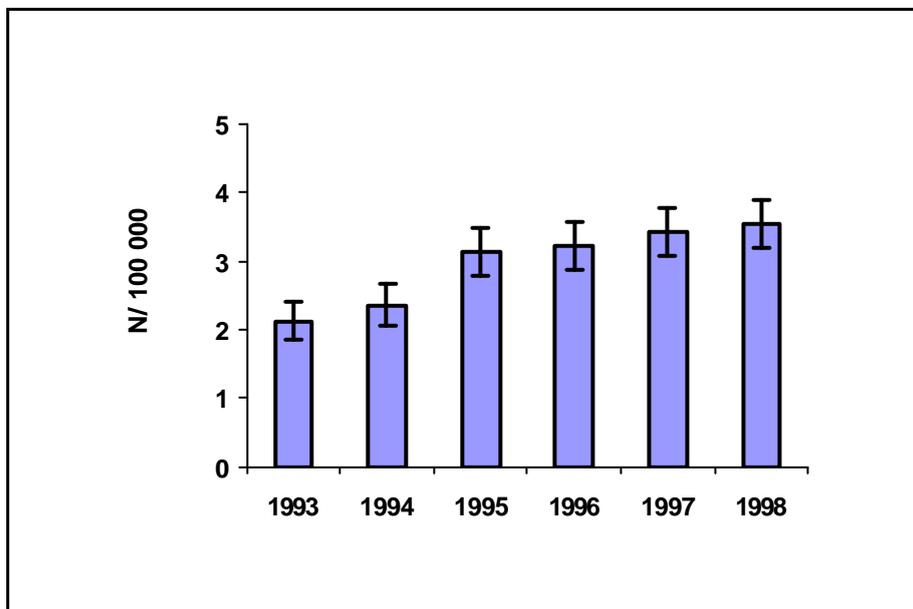
D. Deliège, Ecole de Santé Publique, Université Catholique de Louvain - pour l'étude des données de l'INAMI (Fig. Th 7)

C.M. Vandecasteele (SCK-CEN)- - pour l'utilisation de la carte géographique de la contamination belge en 1986 (Fig. Th1)

**Figures illustrant les résultats**



*Fig.Th1 Contamination de la Belgique après l'accident de Tchernobyl (laboratoire de radioécologie du SCK-CEN)*



*Fig. Th2 Evolution du taux standardisé par âge du cancer de la thyroïde opéré en Belgique entre 1993 et 1998 (Population standard européenne, intervalle de confiance de 95%) – Etude nationale sur RCM*

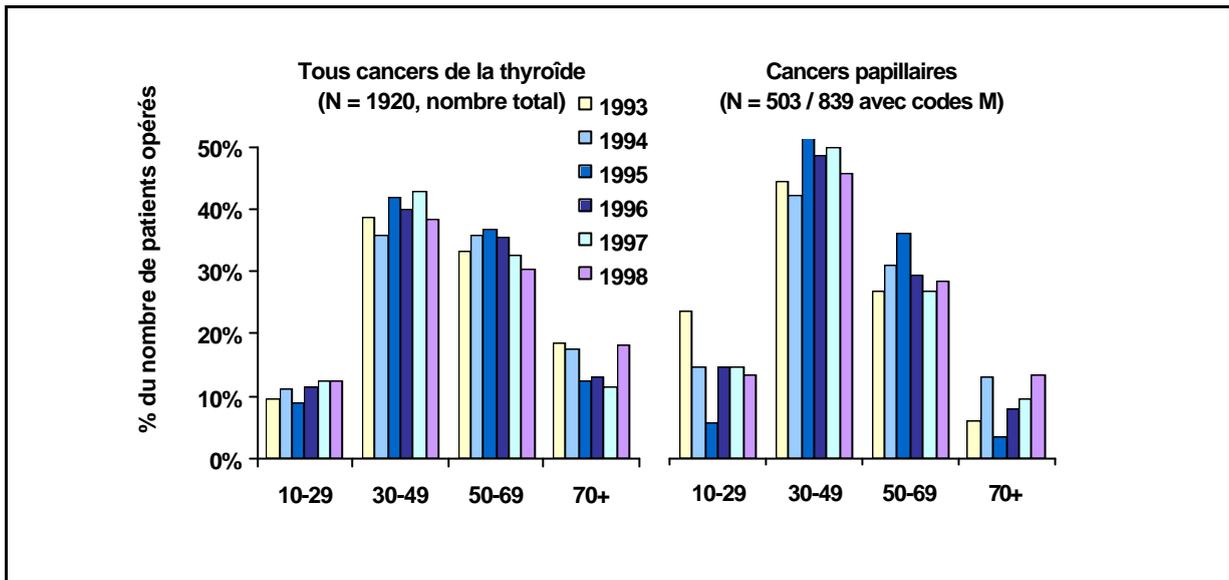


Fig. Th3 Distribution de la proportion par âge des patients opérés d'un cancer de la thyroïde, évolution par année pour l'ensemble des cancers opérés et pour les cancers papillaires (parmi les cas enregistrés avec codes M) – Etude nationale sur RCM.

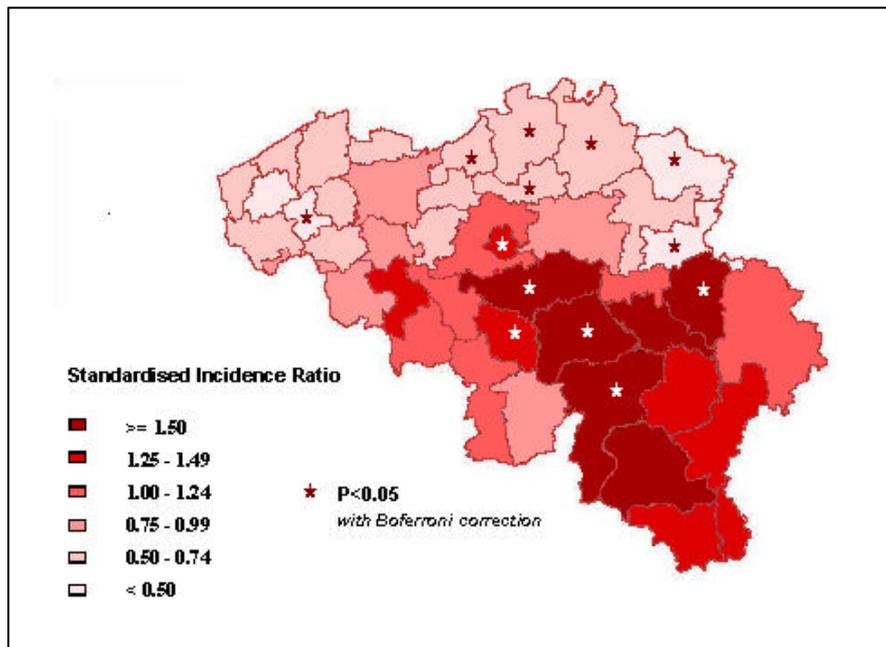


Fig. Th4 Cancer de la thyroïde opéré en Belgique entre 1993 et 1998, par arrondissement: rapport d'incidence standardisée (méthode de standardisation indirecte)\* – Etude nationale sur RCM.

\* La différence est significative si l'intervalle de confiance du SIR (pour un seuil de signification de 95%) ne contient pas la valeur 1 La correction de Bonferroni est appliquée pour réduire l'effet du hasard dans la signification lors de tests multiples

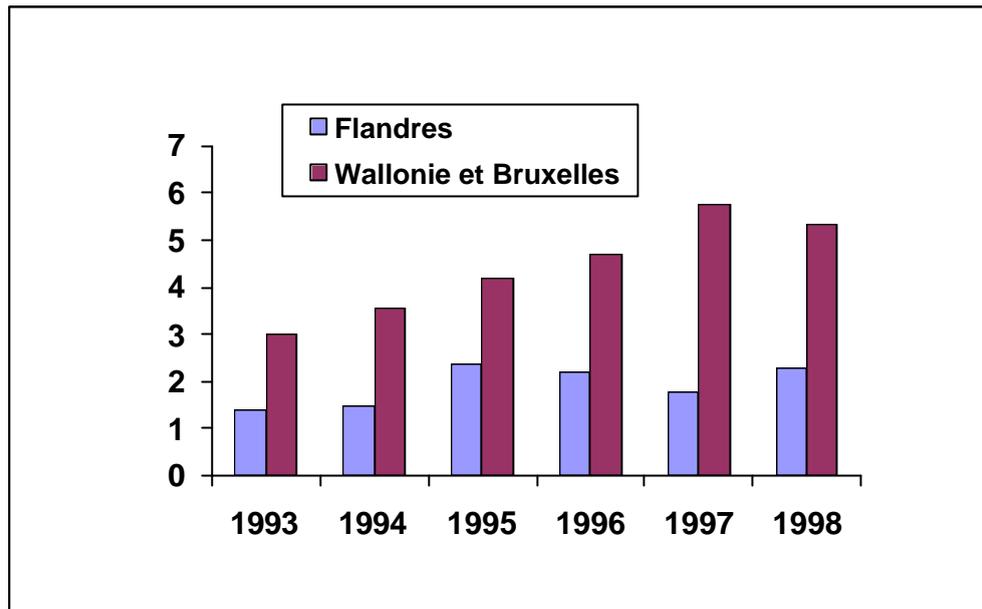


Fig. Th 5 Evolution de l'incidence standardisée par région de 1993 à 1998 – Etude nationale sur RCM.

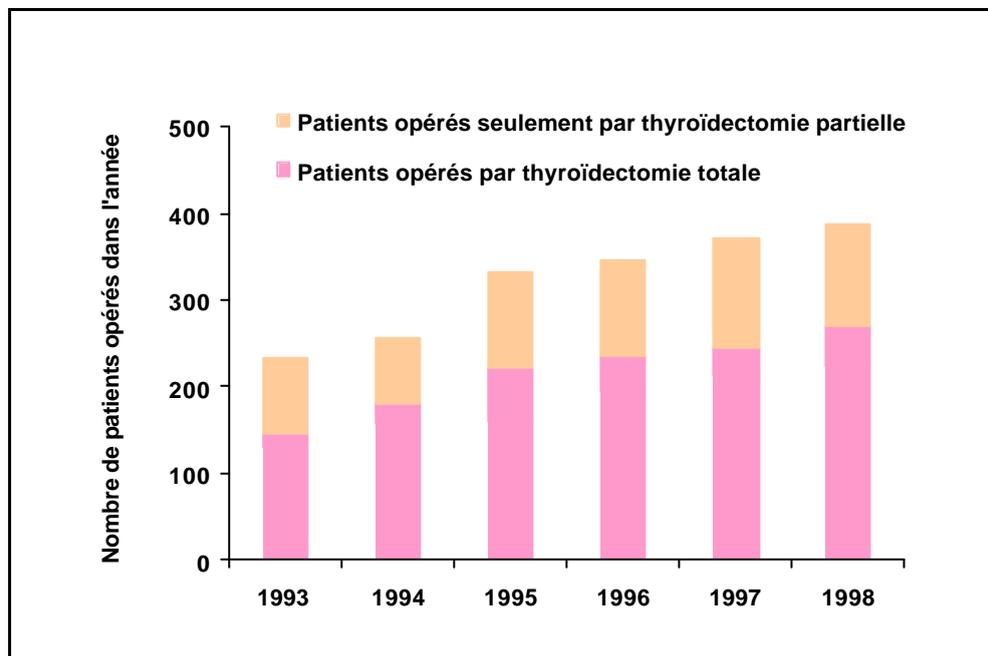
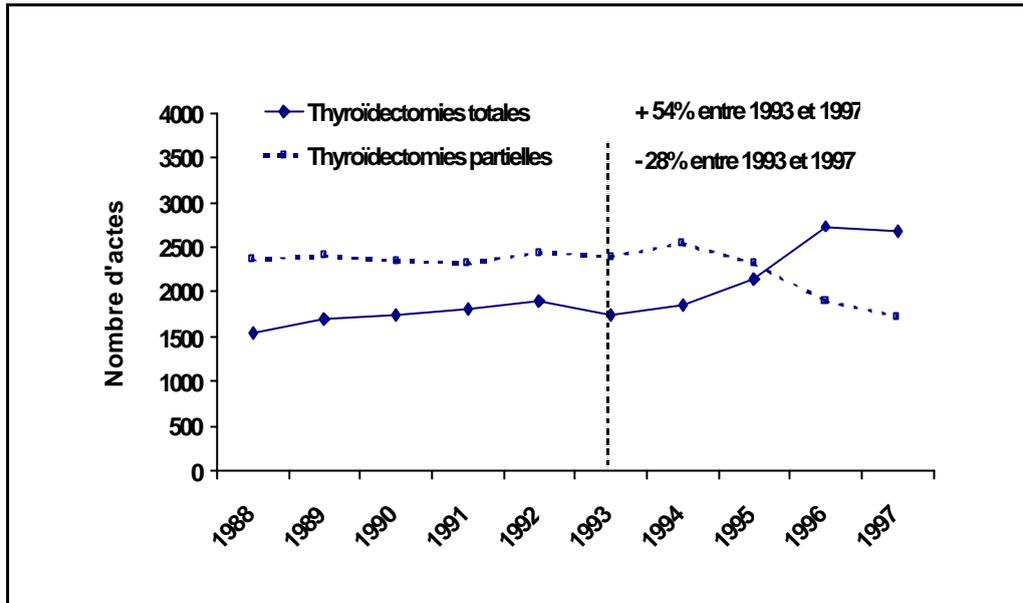


Fig. Th6 Evolution du nombre de cancers de la thyroïde opérés et du type de pratique chirurgicale en Belgique de 1993 à 1998 – Etude nationale sur RCM.



*Fig. Th8 Evolution des types de pratiques opératoires associée aux maladies thyroïdiennes en Belgique - Etude nationale sur les données de l'INAMI*

## IV. Les infections à méningocoques

### Objectif particulier de l'étude

La méningite à méningocoques est actuellement la plus fréquente des méningites bactériennes en Belgique comme en Europe. C'est la seule aussi qui provoque des épidémies. En l'absence de traitement, la létalité peut atteindre 50%.

Depuis une dizaine d'années, la Belgique connaît une recrudescence de la maladie et, depuis 1996, notre pays est classé parmi les pays européens présentant une haute incidence {1}. Depuis un an, après l'augmentation récente des infections à méningocoques C, la vaccination des enfants contre cette maladie est préconisée.

En Belgique, les infections à méningocoques sont des «maladies à déclaration obligatoire», qui doivent être notifiées par les médecins et les laboratoires à l'Inspecteur d'Hygiène de la Province dans laquelle est domicilié le patient. L'inspection provinciale envoie les données aux communautés mais celles-ci ne sont malheureusement pas centralisées au niveau national.

Depuis 1983, la section d'épidémiologie de l'Institut Scientifique de Santé Publique (ISP) coordonne un réseau sentinelle de laboratoires (laboratoire vigies) qui permet d'estimer l'incidence de certaines maladies infectieuses, dont les infections invasives à méningocoques. Grâce à sa répartition géographique, celui-ci couvre 50% du pays et permet de détecter d'éventuelles épidémies {2}. D'autre part, le laboratoire de référence confirme le diagnostic des souches envoyées, assure le typage de celles-ci, analyse l'évolution des sérogroupes et étudie la résistance des germes aux antibiotiques. Tous les laboratoires du pays peuvent envoyer leurs souches au laboratoire de référence mais cet envoi n'est pas systématique: il est estimé à 75%.

Notre étude a pour but de déterminer dans quelle mesure l'analyse des données hospitalières pourrait améliorer la surveillance nationale des maladies infectieuses comme l'infection à méningocoques. Nous avons dès lors comparé les résultats de notre étude sur les Résumés Cliniques Minimum (RCM) aux résultats d'analyse des deux bases de données de l'ISP.

### Matériel et méthode

L'étude porte sur les données du RCM des hôpitaux ayant accueilli les patients en séjours aigus pendant la période 1993 à 1998. Nous avons sélectionné tous les séjours pour lesquels avaient été enregistrés au moins un des diagnostics correspondants aux codes ICD9-CM suivants :

036.0 Méningite à méningocoque, 036.1 Encéphalite à méningocoque, 036.2 Méningococcémie, 036.3 Syndrome de Waterhouse-Friderichsen sur méningocoque, 036.4 Cardite sur méningocoque, 036.8 Autres infections à méningocoque, 036.9 Infection à méningocoque, non spécifiée.

La sélection inclut tous les diagnostics, principaux ou secondaires. Tous les séjours hospitaliers ont été retenus quelle que soit leur durée.

Le calcul d'incidence concerne les patients pour lesquels a été enregistré au moins une fois dans l'année un code mentionné ci-dessus. Les patients sont définis dans chaque hôpital par le numéro anonyme, l'âge, le sexe et l'arrondissement du domicile. Les données envoyées au ministère sont anonymes et il n'est donc pas possible de suivre les patients d'une année à l'autre ou d'un hôpital à l'autre.

La comparaison des cas entre la base de données RCM, celle des laboratoires vigies et celle du laboratoire de référence ne fut pas possible car l'ensemble des variables communes entre ces bases n'est pas assez discriminant. Elles sont au nombre de quatre (âge et le sexe du patient, l'arrondissement

de son domicile et l'année du diagnostic de la maladie) et pour deux d'entre elles, les deux bases de données de l'ISP avaient trop de valeurs manquantes.

Nous avons dès lors comparé, pour les trois sources de données, les résultats d'incidence dans le temps (annuelle et saisonnière), dans l'espace (répartition géographique) et selon les caractéristiques des patients ou de leur maladie (âge, sexe, létalité).

## Résultats

Pour la période étudiée, l'incidence annuelle moyenne des infections à méningocoques, calculée sur les RCM est de 3.2 patients par 100 000 habitants, soit deux fois plus importante que celle enregistrée par le réseau des laboratoires vigies (1.6 cas / 100 000 habitants) et 1,7 fois plus que celle du laboratoire de référence qui est de 1,9 cas / 100 000 habitants (Fig. M1).

Les données hospitalières nous présente le type clinique de l'infection : la majorité de celles-ci sont enregistrées comme méningite (55%) ou comme méningococcémie (30%).

Les infections à méningocoques sont classées en diagnostic principal dans 88% des cas. Quand ce diagnostic apparaît seulement en diagnostic secondaire, le diagnostic principal désigne directement cette affection seulement dans 2% des cas totaux: septicémie pour 1% et méningite bactérienne pour 1%.

Entre 1993 et 1998, le nombre de cas enregistrés sur les données hospitalières a pratiquement doublé (+ 88%). Le laboratoire de référence et les laboratoires vigies enregistrent aussi une nette augmentation (pour le laboratoire de référence + 59%; pour les laboratoires vigies + 49% entre 1993 et 1998 mais +70% entre 1993 et 1997).

La répartition du pourcentage des cas par classes d'âge est restée relativement constante. Elle montre un profil semblable sur les RCM et sur les deux bases de l'ISP (Fig. M2). Sur les données RCM, la proportion semble cependant un peu moins importante chez les adolescents entre 15 et 19 ans (14% sur RCM versus 18% sur les bases de l'ISP). La proportion d'hommes reste toujours un peu plus élevée pendant les six années, elle est en moyenne de 55% sur les RCM et de 54% dans l'analyse des données des laboratoires vigies.

L'évolution saisonnière montre un nombre de cas plus faible en été. Le profil est identique sur les RCM et sur les données des laboratoires vigies (Fig. M3) ainsi que sur les données du laboratoire de référence.

Pour 67%, les patients sont envoyés à l'hôpital par un médecin généraliste, 14% viennent de leur propre initiative tandis que 11% ont été transférés d'un autre hôpital, la provenance des autres patients n'étant pas enregistrée.

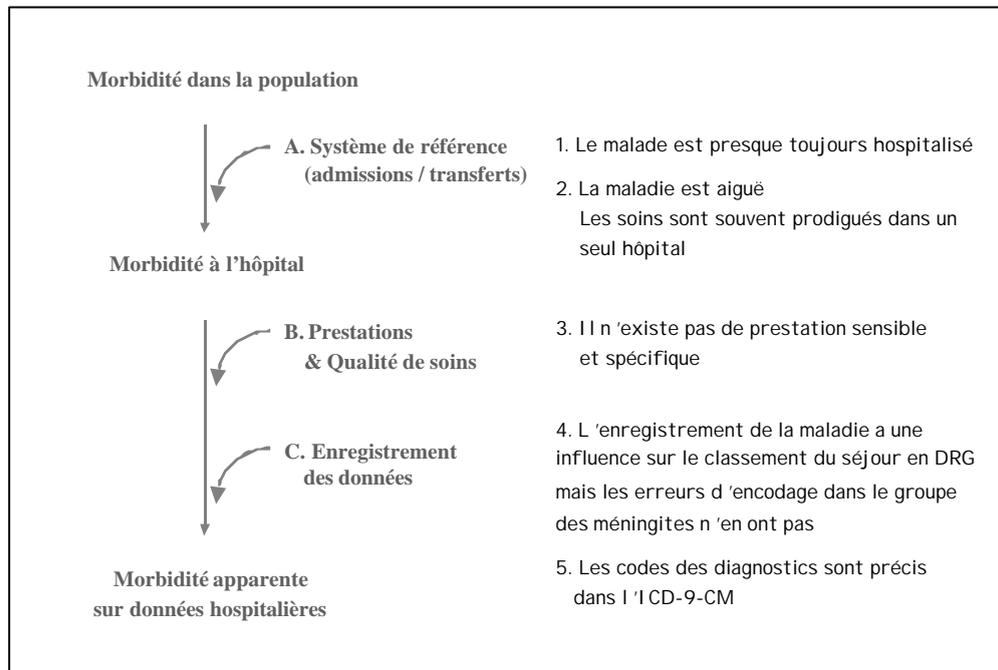
La létalité moyenne à l'hôpital est de 7% (10% en 1993, 6% en 1998), pour le laboratoire de référence, elle est de 5.3%. Le décès n'est pas enregistré dans la base de données des laboratoires vigies.

Hormis la province de Luxembourg, pour laquelle l'incidence importante représente un nombre très faible de cas (3 à 12 cas par an relevés sur les RCM), les provinces les plus touchées sont identiques pour la période étudiée d'après l'analyse des trois bases de données (Fig. M4) {3}. Dans les données des laboratoires vigies, l'enregistrement de l'arrondissement du patient manque dans 11% des cas pour l'ensemble des 6 années étudiées.

La province d'Anvers et en particulier, l'arrondissement d'Anvers présenterait la plus forte incidence: de 3.9 cas pour 100 000 habitants (4.1 pour l'arrondissement) sur les données hospitalières. Le Brabant serait ensuite le plus concerné avec sur les RCM, une incidence moyenne de 3.8 cas, atteignant 4.6 cas pour 100 000 habitants dans l'arrondissement de Bruxelles. Et enfin, le taux d'incidence dans la Flandre occidentale s'élèverait à 3.6 pour 100 000 habitants.

## Discussion

Les limites théoriques de validité des données hospitalières pour l'étude de l'incidence de l'infection à méningocoques sont résumées dans la figure M5 ci-dessous



**Fig. M5** Les limites théoriques de validité des données hospitalières pour l'étude de l'incidence de l'infection à méningocoques

Sauf pour quelques rares exceptions, tous les patients chez qui le diagnostic de méningite ou d'infection à méningocoques est posé ou suspecté, sont hospitalisés. L'incidence de la maladie à l'hôpital représente donc bien l'incidence dans la population générale.

Quelques rares patients sont transférés dans un second hôpital et si le diagnostic a bien été enregistré dans les 2 hôpitaux, ces patients ont donc été comptabilisés deux fois dans l'analyse. En effet, les données envoyées par les hôpitaux sont rendues anonymes avant leur envoi au Ministère. D'après notre analyse, parmi les 218 patients (11%) transférés d'un autre hôpital, 41 patients seulement ont peut-être été enregistrés deux fois si l'on considère les caractéristiques suivantes de ces patients: l'âge, le sexe, l'arrondissement du domicile, le mois de l'admission dans l'hôpital (même mois ou mois suivant) L'incidence n'a probablement été surévaluée de ce fait que de 1 à 3%.

La codification ICD-9-CM employée sur RCM est assez fournie pour décrire les méningites à méningocoques mais quelques erreurs de classification dans le groupe des méningites sont probables car elles n'ont aucune influence sur le classement du séjour pour leur traitement financier.

On peut par exemple, se poser la question de l'exactitude du diagnostic du syndrome de Waterhouse-Friderichsen enregistré en diagnostic secondaire quand le diagnostic principal encodé est une cholangite ou une myocardite idiopathique, ou de l'étiologie d'une méningite atteignant un patient deux mois après l'enregistrement d'une infection à HIV.

En résumé, il est probable qu'une légère surestimation de l'incidence soit due à quelques rares transferts de patients d'un hôpital à l'autre et principalement à des erreurs de codification.

La comparaison des résultats de l'analyse sur RCM montre cependant une bonne concordance avec ceux calculés sur les données des laboratoires vigies et du laboratoire de référence pour la majorité des paramètres mesurés : âge, sexe et évolution saisonnière.

## **Conclusion**

Pour la période étudiée, l'incidence des infections à méningocoques relevée sur les RCM est deux fois plus importante que celle relevée sur la base des laboratoires vigies et 1,7 fois plus élevée que celle calculée sur les données du laboratoire de référence.

On peut évaluer à moins de 10% la surestimation de l'incidence de l'infection sur les RCM, due principalement à quelques transferts de patients d'un hôpital à l'autre et quelques erreurs de codifications.

Dans les trois bases de données, les caractéristiques mesurées sont néanmoins comparables: l'évolution saisonnière, la létalité et la répartition par sexe et groupes d'âge, les plus exposés étant les enfants en bas âge et les jeunes de 15 à 20 ans.

La répartition géographique des cas montre les mêmes tendances, mais suggère peut-être une sous-estimation plus importante de l'incidence par le réseau pour les provinces du Hainaut et de Liège.

Etant donné les délais d'enregistrement et de traitement des données RCM, celles-ci ne peuvent être utilisées pour contribuer à la surveillance épidémiologique rapprochée de maladies aiguës et contagieuses telles que l'infection à méningocoques. De plus, ces données sont exclusivement cliniques et ne peuvent donc mettre en évidence l'évolution des sérogroupes, indispensables pour prévoir d'éventuelles campagnes de vaccination.

Notre étude offre cependant l'avantage d'une couverture nationale et les résultats nous montrent l'intérêt d'exploiter cette base de données de manière courante pour contribuer à la validation des données du réseau de surveillance actuellement en place.

## **Références bibliographiques**

1. F. Van Loock, G. Ducoffre, and F. Carion. 2001. Meningokokkeninfecties: staan we voor een nieuwe epidemie? [Dutch]. *Epidemiologisch Bulletin van de Vlaamse Gemeenschap*. 36:4-7.
2. Ducoffre G. Surveillance des maladies infectieuses par un réseau des laboratoires de microbiologie 1999 + Tendances épidémiologiques 1983 - 1998 Institut Scientifique de Santé Publique D/2000/2505/30.
3. Hanquet, G., F. Carion, and F. Van Loock. 2001. Evolution des infections à méningocoques en Belgique [French]. *Perceptile* 6(5):128-132.

## **Collaborateurs**

G. Ducoffre, G. Hanquet, Institut Scientifique de Santé Publique, section d'épidémiologie

F. Carion, Institut Scientifique de Santé Publique, section de bactériologie

Figures illustrant les résultats

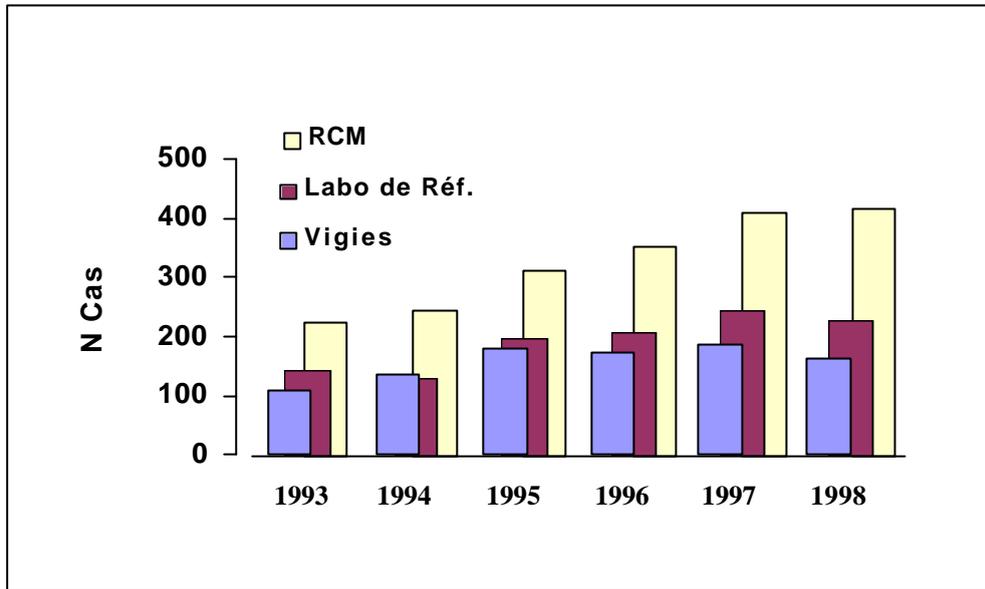


Fig. M1 Evolution du nombre de cas d'infections à méningocoques relevés annuellement entre 1993 et 1998- Comparaison des nombres enregistrés par - le réseau des laboratoires vigies, - le laboratoire de référence et - sur les données hospitalières RCM

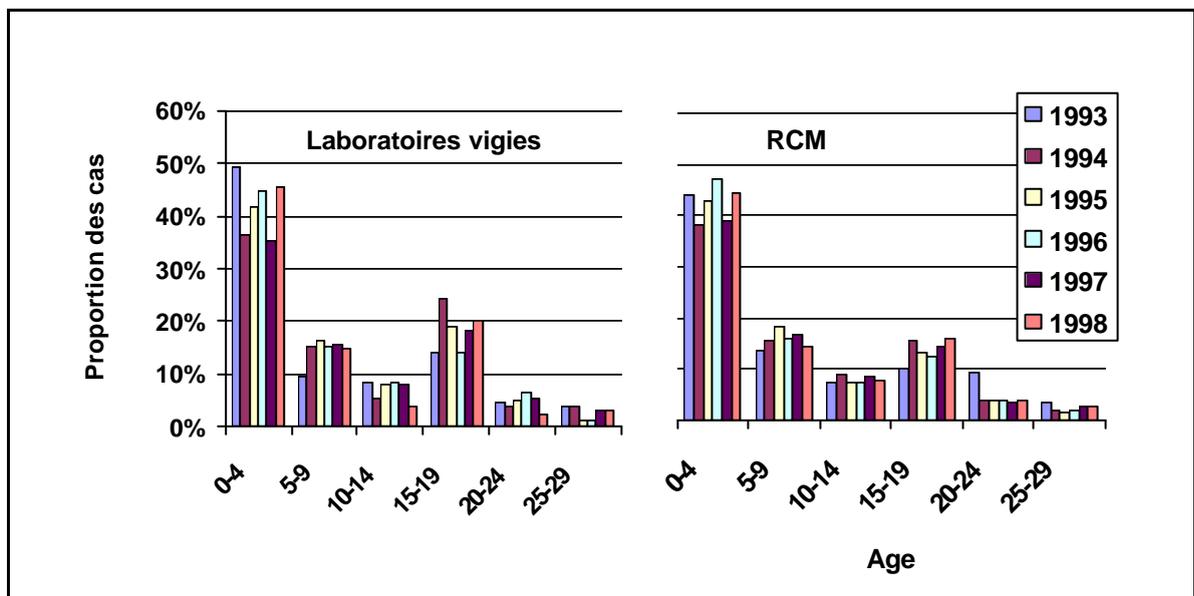


Fig. M2 Proportion par tranches d'âge, des cas d'infections à méningocoques de 1993 à 1998, - Comparaison entre les cas enregistrés par les laboratoires vigies et sur le RCM.

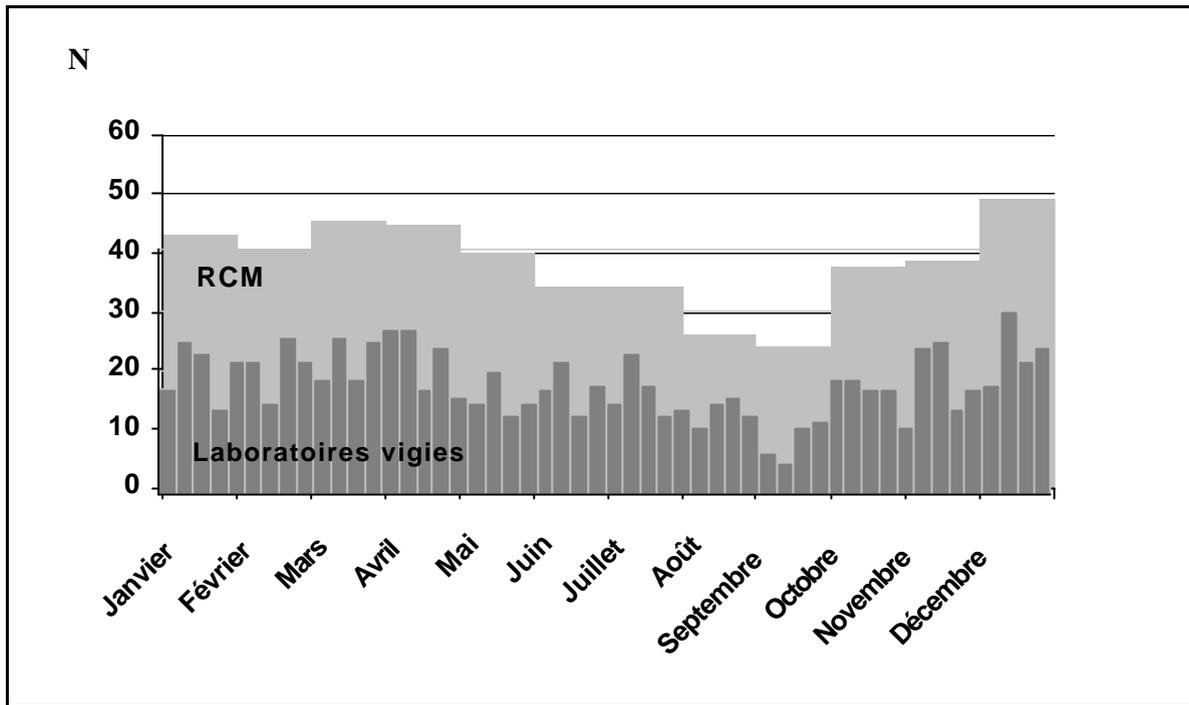


Fig.M3 Evolution saisonnière du nombre moyen de cas d'infections à méningocoques de 1993 à 1998 - Comparaison entre les cas enregistrés par les laboratoires vigies (incidence hebdomadaire) et sur le RCM (incidence mensuelle)

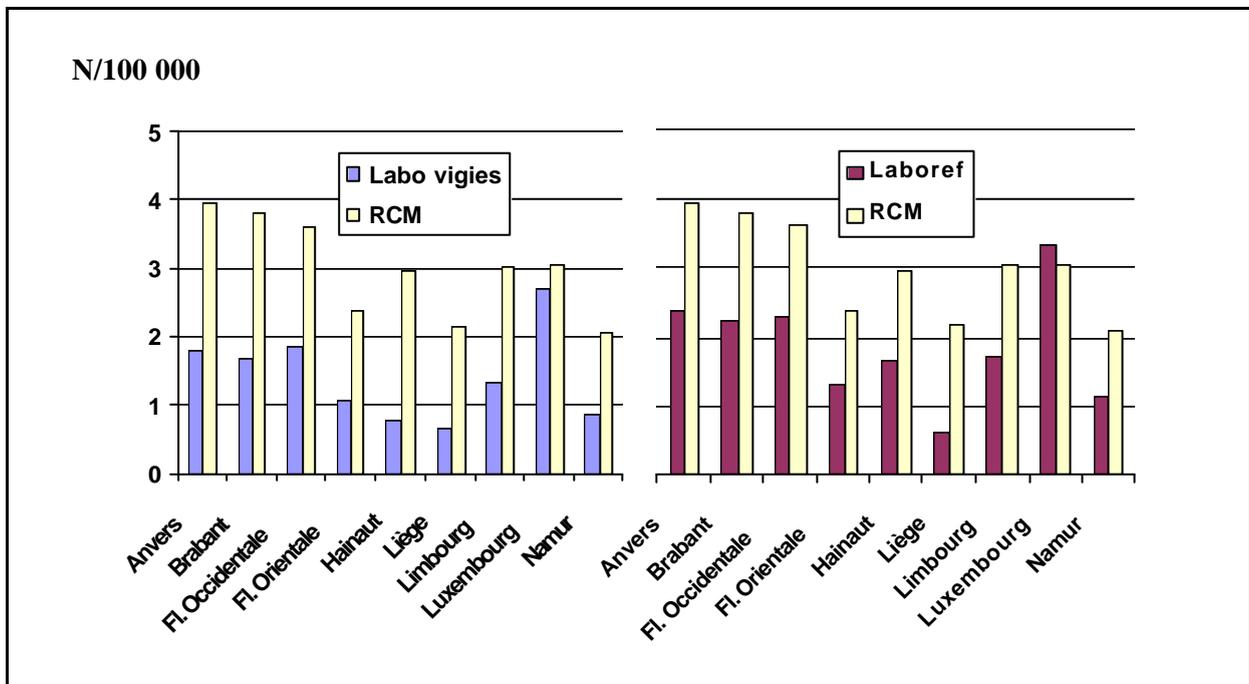


Fig. M4 Répartition géographique des cas d'infections à méningocoques entre 1993 à 1998 - Comparaison entre les incidences enregistrées par province - par les laboratoires vigies, le laboratoire de référence et sur le RCM

## VII. L'hypospadias

### Objectifs particuliers de l'étude.

Les techniques dans les soins périnataux ont fait des progrès considérables dans les dernières décennies, diminuant fortement la mortalité et la morbidité périnatale. Dans ce contexte, les malformations congénitales sont apparues aux premiers rangs des préoccupations médicales en pédiatrie. Actuellement, on considère que 2.5 à 3% des enfants naissent avec une de ces anomalies {1}.

L'hypospadias est une malformation très fréquente de l'appareil génito-urinaire chez le garçon, qui peut être invalidante ou bénigne, découverte parfois à un âge avancé. Des études épidémiologiques semblent indiquer un lien entre cette anomalie et l'exposition à certains polluants agricoles ou industriels.

Le programme européen d'étude sur les maladies congénitales (European Registration of Congenital Anomalies EUROCAT) a débuté en 1979. En Belgique, deux registres régionaux y participent: celui de la province d'Anvers et celui des provinces de Namur et Hainaut-oriental {2}. Notre étude essaie de déterminer dans quelle mesure l'analyse des données hospitalières pourrait contribuer à la surveillance des maladies congénitales comme l'hypospadias à un niveau national.

### Matériel et méthodes

La classification ICD-9-CM ne distingue pas l'hypospadias de l'épispadias ni pour le diagnostic, ni pour le traitement chirurgical de cette maladie. Par contre, la prestation enregistrée dans la nomenclature INAMI est spécifique de l'hypospadias et classe même les interventions en : cure totale, cure partielle au temps opératoire principal ou cure partielle au temps opératoire complémentaire. Le rapport de l'incidence annuelle des cas opérés sur le nombre de naissances par an, nous a semblé une bonne approche de la prévalence de la maladie à la naissance.

Pour la période 1993 à 1998, nous avons sélectionné, parmi les données des Résumé Cliniques Minimum (RCM) des hôpitaux accueillant les patients en séjours aigus, tous les séjours pour lesquels avaient été enregistrés au moins un des codes suivants :

- Codes du diagnostic ICD-9-CM: 752.6 hypospadias ou épispadias
- Codes des prestations
  - ICD9-CM : 5845 Réparation d'hypospadias ou épispadias
  - INAMI : 261310, 261321 Opération pour hypospadias, cure radicale;  
261332, 261343 Opération pour hypospadias, temps principal.

La sélection inclut tous les diagnostics et tous les séjours hospitaliers, quelle que soit leur durée.

Le calcul d'incidence concerne les patients pour lesquels ont été enregistrés au moins une fois dans l'année, le code diagnostique 752.6 ainsi qu'un des codes opératoires ci-dessus. Les patients sont définis dans chaque hôpital par le numéro anonyme, l'âge et l'arrondissement de leur domicile.

### Résultats

Le nombre de patients opérés d'hypospadias augmente de 272 en 1993, à 453 en 1995 puis reste relativement stable dans les années suivantes (Fig. H1). Pour 13 % de l'ensemble des patients subissant cette intervention, le diagnostic enregistré sur le RCM n'était pas l'hypospadias mais d'autres anomalies de la sphère uro-génitale

Avant d'interpréter ces résultats en terme de prévalence, il est intéressant de considérer l'âge des patients au moment de l'intervention (Fig. H2) : 25% seulement ont été opérés avant l'âge de 2 ans,

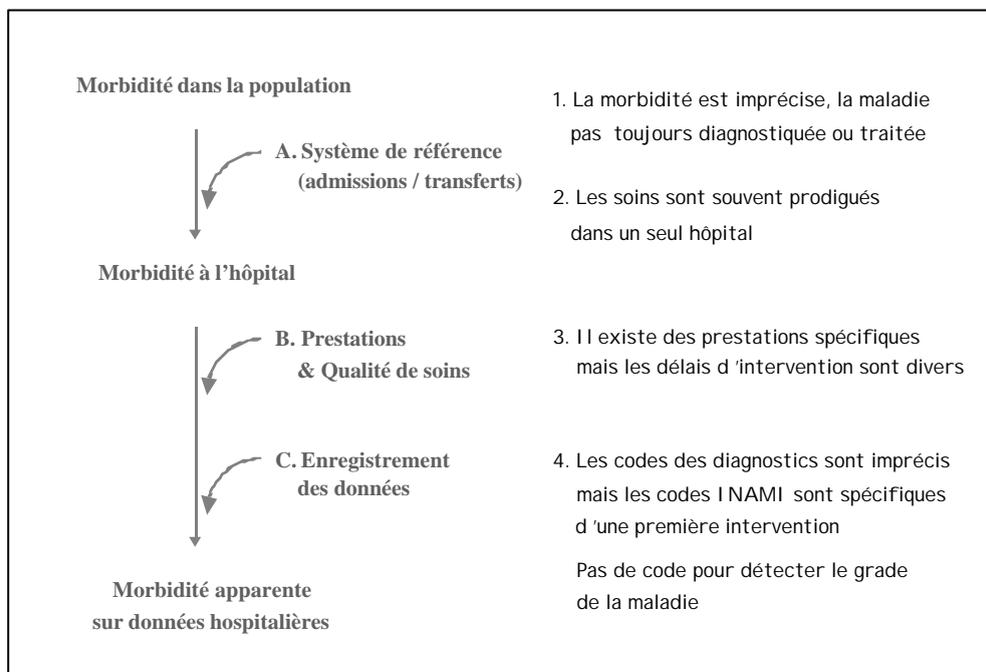
43% entre l'âge de 2 et 4 ans et 32% après cet âge, quelques patients étant même âgés de plus de 65 ans lors de l'opération.

Si nous divisons nos résultats d'incidence annuelle des patients opérés par le nombre de naissances, nous avons une approche de la prévalence à la naissance. Ce calcul ne représente pas une vraie prévalence car il comptabilise des patients nés parfois 10 ou 20 ans auparavant et néglige les jeunes patients qui ne sont pas opérés dans l'année (Fig. H2). On ne peut dès lors comparer les différentes périodes annuelles de l'étude en terme de prévalence. Globalement, pour l'ensemble de la période étudiée, l'incidence des patients opérés de cette maladie serait de 23 pour 10 000 naissances en 1993, 30 en 1994, puis entre 35 et 41 pour les années suivantes.

Pour la période de l'étude, la distribution géographique montre quelques disparités d'une province à l'autre (Fig. H3): l'incidence est plus élevée en Flandre occidentale et dans le Hainaut, plus basse dans la province d'Anvers et de Liège.

## Discussion

Avant toute discussion, il est important de poser les limites théoriques de la validité de notre analyse sur les données hospitalières. Celle-ci est résumée dans la figure H4 ci-dessous



**Fig. H4 Limites de validité des données hospitalières pour l'étude de prévalence de l'hypospadias**

Pour cette étude de prévalence, la validité des données hospitalières est bridée principalement par

- L'imprécision du diagnostic de la maladie qui désigne un défaut morphologique de morbidité très variable. De ce fait, le traitement de la maladie devient aléatoire quant au délai d'intervention et peut-être quant à la prestation elle-même.
- la pauvreté de la classification ICD-9-CM qui ne répertorie pas de code spécifique pour cette maladie. De même, il n'existe pas de code déterminant le grade de la maladie. Pour détecter les patients atteints, nous avons eu recours aux codes de prestations répertoriés dans la codification INAMI.

Du fait de ces limites d'enregistrement, nous n'avons pu détecter tous les cas diagnostiqués à la naissance, mais seulement les patients lorsqu'ils sont opérés, parfois à l'âge adulte.

Par contre, les codes INAMI sont assez spécifiques pour permettre la seule sélection des nouveaux cas. Nous avons donc pu éviter les "faux positifs" dus aux patients ayant subi une deuxième intervention dans un hôpital différent et qui n'auraient pas été discriminés à cause de l'anonymisation des données à la sortie de l'hôpital.

En comparant le nombre total de patients pour lesquels un code de l'intervention avait été enregistré sur RCM (avec ou sans diagnostic d'hypospadias), et le nombre d'actes enregistrés par l'INAMI sur une base de données indépendante, nous avons remarqué une discordance nette pour les années 1993 et 1994 (Fig. H5): les données RCM ne paraissent pas exhaustives pour ces deux premières années

Les chiffres présentés par notre analyse pour les années suivantes sont néanmoins intéressants pour déterminer l'importance de cette maladie dans la population tant au niveau national qu'à un niveau plus régional:

- Le nombre global de patients opérés se situe entre 35 et 40 par an et par 10 000 naissances, ce chiffre est donc au moins trois fois plus important que la prévalence relevée par EUROCAT en province de Namur (10.8 / 10 000 naissances) mais ce registre ne relève pas les cas présentant une forme légère de la maladie.
- Sur le RCM, l'incidence moyenne dans la province de Namur (32 patients opérés par 10 000 naissances), diffère peu de la moyenne nationale. Le registre régional de Namur donne donc une prévalence représentative de la prévalence dans notre pays pour la période étudiée.

## Conclusion

Au niveau méthodologique, cette étude a mis en évidence l'intérêt particulier des données de prestations et de la codification INAMI. Elle a toutefois aussi montré qu'il n'était pas possible de suivre l'évolution de l'incidence annuelle d'une maladie sur les données hospitalières si celle-ci n'est pas répertoriée par un code précis pour le diagnostic et si le traitement chirurgical n'est pas relativement urgent et indispensable.

Pour l'étude de la maladie, nous avons pu néanmoins indirectement quantifier l'importance de la morbidité liée à l'hypospadias. Nos résultats montrent des chiffres trois fois plus élevés que la prévalence enregistrée par EUROCAT en province de Namur mais la notion de morbidité étudiée n'est pas identique dans les deux analyses.

Au niveau géographique, nous pouvons affirmer que la prévalence relevée par le registre semble représentative de la prévalence nationale pour la période de 1993 à 1998. Des analyses géographiques trop fines seraient aléatoires car trop liées à la qualité des soins.

## Références bibliographiques

1. [http://web.wanadoo.be/ggolfb/20010217\\_GILLEROT.htm](http://web.wanadoo.be/ggolfb/20010217_GILLEROT.htm)
2. 15 years of surveillance of congenital anomalies in Europe 1980-1994. EUROCAT Report 7

## Collaborateurs:

Mme D. Deliège, Ecole de Santé Publique, Université Catholique de Louvain - pour l'étude des données de l'INAMI

Y. Gillerot, Institut de génétique – Lovreval, Responsable Eurocat pour les provinces de Namur et Hainaut oriental

Figures illustrant les résultats

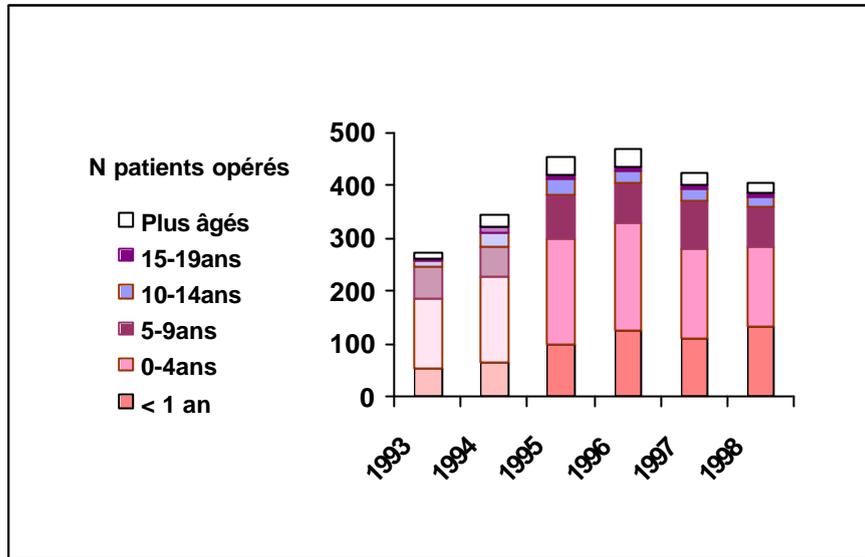


Fig. H1 Nombre de patients opérés d'hypospadias en Belgique (avec diagnostic d'hypospadias), par classes d'âge entre 1993 et 1998 - Etude nationale sur RCM -

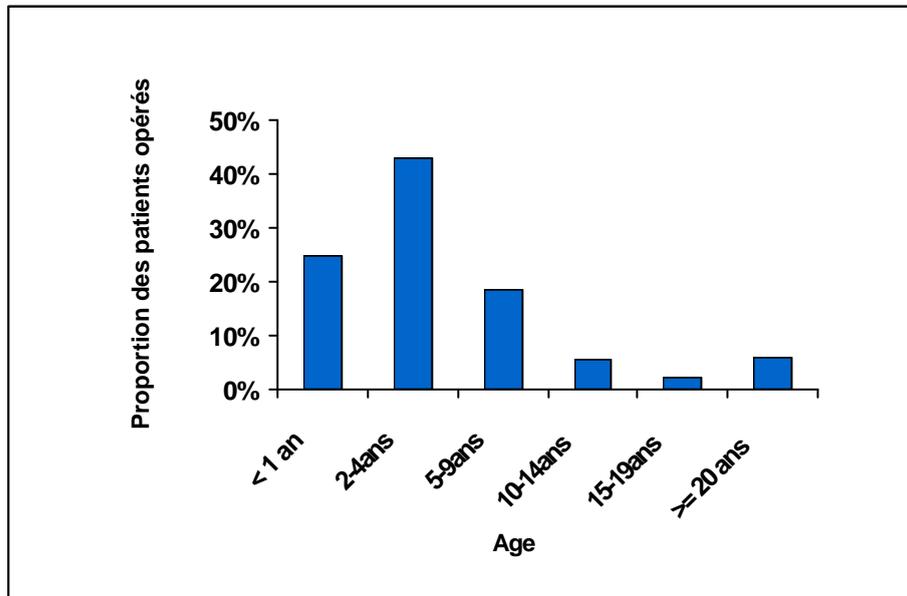


Fig. H2 Patients opérés d'hypospadias en Belgique entre 1993 à 1998, proportion des cas par classes d'âge - Etude nationale sur RCM -

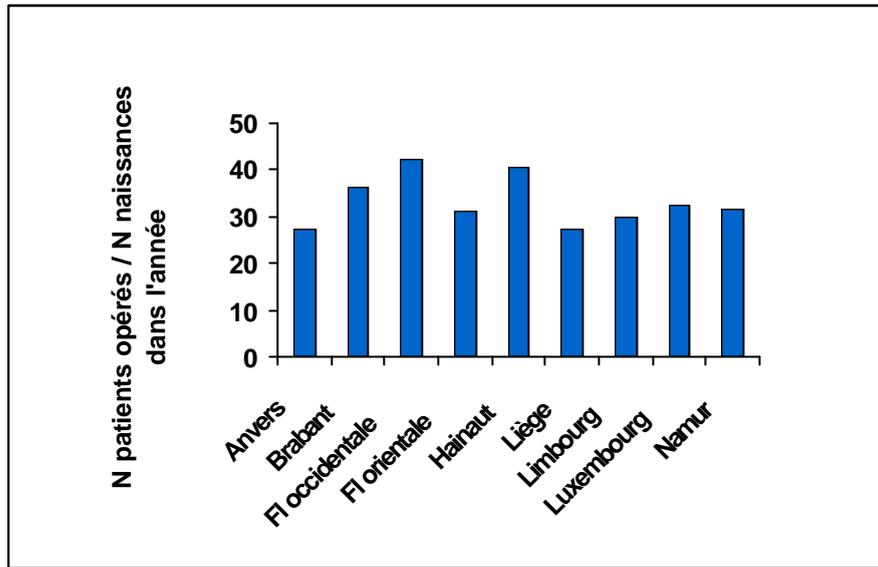


Fig. H3 Rapport de l'incidence des hypospadias opérés par province, sur le nombre de naissances, par province pour la période 1993 à 1998 - Etude nationale sur RCM

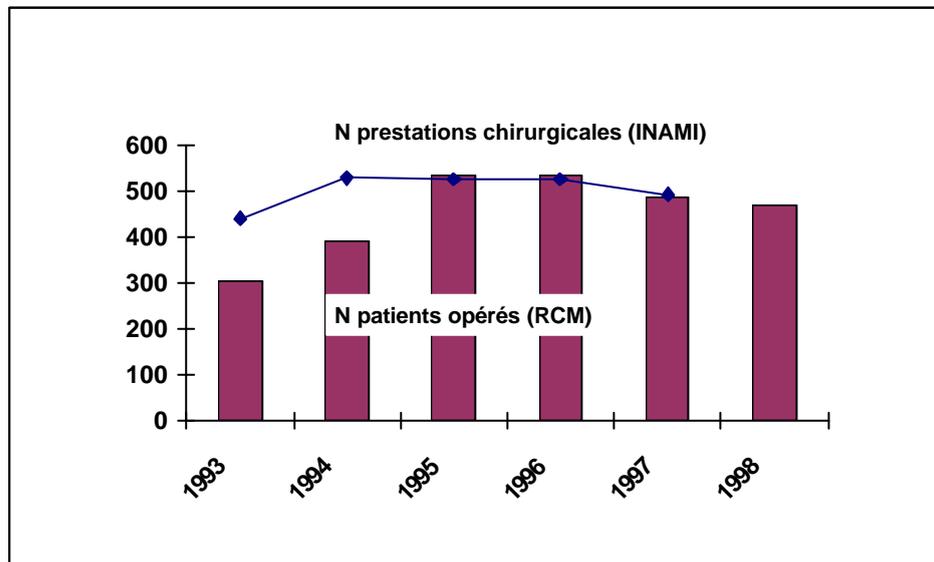


Fig. H5 Evolution du nombre d'interventions pour hypospadias de 1993 à 1998 - Confrontation des résultats d'analyses sur les données RCM (nombre de patients opérés avec ou sans diagnostic d'hypospadias) et du nombre d'actes enregistrés par l'INAMI (cure radicale pour hypospadias ou cure partielle, 1ere intervention)

## VIII. Le diabète sucré

### Objectifs particuliers de l'étude.

Le diabète est une maladie prévalente dans la population occidentale. En Belgique, la dernière enquête de santé par interview évaluait à 2.6% la proportion des personnes ambulantes reconnaissant en être atteintes {1}. L'analyse des données du réseau des médecins vigies, révélait qu'en 2000, 3.3% de leurs patients étaient diabétiques {2}. Une étude antérieure a cependant montré qu'un grand nombre de malades ignoraient leur état {3}. En effet, le dépistage de la maladie, pourtant très important dans la prévention de ses complications, n'est pas systématique dans la pratique médicale générale. C'est par contre le cas en milieu hospitalier ; le diabète est dès lors parfois découvert fortuitement lorsqu'un patient séjourne à l'hôpital pour une toute autre raison. L'enjeu d'une exploitation épidémiologique des données hospitalières pour étudier cette maladie est donc considérable. Mais cette exploitation est-elle possible pour les maladies chroniques telles le diabète?

Certaines études ont déjà évalué la validité de l'enregistrement de la pathologie diabétique sur RCM, en comparant les diagnostics enregistrés à ceux figurant dans le dossier médical {4}. Notre analyse essaye de déterminer la faisabilité de l'étude du diabète sur les données hospitalières, principalement en testant la validité des résultats pour les études de la prévalence hospitalière de la maladie, de son traitement et de ses complications cliniques.

L'étude de l'incidence des maladies chroniques telles le diabète n'est pas possible sur les RCM pour les raisons suivantes (Fig. D1) :

- le patient diabétique n'est pas toujours hospitalisé pour diagnostiquer ou traiter sa maladie
- les patients sont souvent traités successivement dans différents hôpitaux mais l'anonymisation des données avant leur envoi au ministère empêche le suivi de ces patients d'une année à l'autre ou d'un hôpital à l'autre
- Les données RCM ne mentionnent pas si le diagnostic est récent et il n'existe pas de prestation spécifique du diabète marquant l'incidence de la maladie

### Matériel et méthodes

L'analyse a été faite sur la base des données des Résumés Cliniques Minimum (RCM) et des Résumés Financiers Minimum (RFM) récoltées à l'Ecole de Santé Publique de l'UCL pour l'année 1997, comprenant 940 942 séjours de patients accueillis dans 62 hôpitaux. Cette base représentait 40% du nombre des séjours ayant eu lieu dans l'ensemble des hôpitaux généraux belges cette année-là.

L'examen détaillé de la classification ICD-9-CM avant la sélection du fichier permet de retenir tous les séjours pour lesquels a été enregistré au moins un «code spécifique du diabète», soit un code contenant le libellé de la maladie, soit celui d'une maladie ou d'un traitement spécifiquement associé au diabète. Nous avons donc sélectionné tous les séjours pour lesquels avaient été enregistré au moins

- sur RCM, un des diagnostics ou procédures médicales correspondant aux codes ICD-9-CM suivants : 250.\*<sup>4</sup> Diabète sucré, 251.8 Troubles de la sécrétion pancréatique, autres<sup>5</sup>, 648.0 Diabète et grossesse, 775.1 Diabète néonatal, 362.0 Rétinopathie diabétique, 366.41 Cataracte diabétique, 357.2 Polyneuropathie diabétique, Z 99.17 Injection d'insuline
- ou sur RFM, un des médicaments correspondant aux codes ATC suivants : A10A Insuline ou A10 B Médicaments antidiabétiques oraux

Le calcul de prévalence concerne les patients définis dans chaque hôpital par le numéro anonyme, l'âge, le sexe et le code de l'arrondissement de son domicile. Dans ce calcul, nous avons considéré

<sup>4</sup> Le 4<sup>e</sup> digit détermine s'il existe une complication et la définit, le 5<sup>e</sup> digit détermine le type de diabète I ou II.

<sup>5</sup> Selon les directives d'enregistrement du Ministère de la Santé, ce code doit être utilisé pour le diabète secondaire

comme patients diabétiques, tous les patients pour lesquels a été enregistré pour au moins un de ses séjours dans l'année, un des codes ci-dessus.

La sélection ne se limite pas au diagnostic principal; elle inclut également les diagnostics secondaires. Tous les séjours hospitaliers ont été retenus quelle que soit leur durée.

Pour étudier les complications chroniques du diabète, nous avons considéré, dans l'ensemble des séjours du patient, identifié comme diabétique par un « code spécifique », toute maladie habituellement associée au diabète même si le code 250\* ne mentionne pas un digit pour ces complications (Cfr. Tableau. D1).

Enfin, nous avons étudié parmi les patients diabétiques, la prévalence de l'obésité (code 278.0\*) et celle de l'incidence de l'infarctus aigu du myocarde (code 410\*)

|  |
|--|
| <ul style="list-style-type: none"> <li>• Diabète avec manifestations rénales : <ul style="list-style-type: none"> <li>- 250.4</li> <li>- ou « code spécifique »<sup>6</sup> du diabète et un des codes ci-dessous chez le même patient : 581.81 Syndrome néphrotique avec maladie classifiée ailleurs, 583.81 Néphropathie avec maladie classifiée ailleurs, 585 Insuffisance rénale chronique, 586 Insuffisance rénale, non spécifiée, V42.0 Transplantation rénale, V56*Dialyse, Z 39.27 Artériostomie pour dialyse rénale, Z 39.95 cathétérisme veineux pour dialyse rénale.</li> </ul> </li> <li>• Diabète avec manifestations ophtalmiques : <ul style="list-style-type: none"> <li>- 250.5</li> <li>- ou 362.0 Rétinopathie diabétique</li> <li>- ou 366.41 Cataracte diabétique,</li> <li>- ou « code spécifique » du diabète et un des codes ci-dessous chez le même patient : 362.83 Œdème de la rétine, 366.41 Amaurose, Z 1424 Photocoagulation de la rétine par laser.</li> </ul> </li> <li>• Diabète avec manifestations neurologiques : <ul style="list-style-type: none"> <li>- 250.6</li> <li>- ou 357.2 polyneuropathie diabétique</li> <li>- ou « code spécifique » du diabète et un des codes ci-dessous chez le même patient : 337.1 Neuropathie du système autonome périphérique, 536.3 Gastroparésie</li> </ul> </li> <li>• Diabète avec désordres circulatoires périphériques : <ul style="list-style-type: none"> <li>- 250.7</li> <li>- ou « code spécifique » du diabète et un des codes ci-dessous chez le même patient : 4402 artériosclérose des artères des extrémités, 443.81 Angiopathie périphérique avec maladie classifiée ailleurs, 997.62 Complication infectieuse sur moignon, V49.7 Status post-amputation d'un membre inférieur, Z 84.1 amputation d'un membre inférieur</li> </ul> </li> <li>• Diabète avec autres complications : <ul style="list-style-type: none"> <li>- 250.8</li> <li>- ou « code spécifique » du diabète et un des codes ci-dessous chez le même patient : 607.1 Balanite, 607.82 Désordre vasculaire du pénis, 682.5, 682.6, 682.7 Abcès au membre inférieur, 707.1 Ulcère des membres inférieurs, 707.11 Mal perforant plantaire, 713.5 Arthropathie neurogénique (Charcot), 731.8 Ostéomyélite associée, 785.4 Gangrène, 997.62 Complication infectieuse sur moignon.</li> </ul> </li> </ul> |
|--|

**Tableau D1. Codes ICD-9-CM identifiant les séjours mentionnant les complications chroniques du diabète**

<sup>6</sup> Les codes identifiant les diabétiques dans notre analyse sont appelés « codes spécifiques » de la maladie, à savoir un code contenant le libellé de la maladie ou d'un diagnostic ou un traitement spécifique de celle-ci.

## Résultats

Sur l'ensemble des séjours hospitaliers, nous avons défini 684 176 patients et parmi eux, 48 333 soit 7% étaient identifiés comme diabétiques. La prévalence hospitalière du diabète varie fortement avec l'âge (Fig. D1). Un « code spécifique »<sup>7</sup> du diabète a été enregistré, chez 1% des patients de moins de 40 ans, 6% des patients âgés de 40 à 59 ans et plus de 15% des patients plus âgés. La maladie semble plus fréquente chez les hommes hospitalisés avant l'âge de 65 ans, et plus fréquente chez les femmes plus âgées.

Après avoir regroupé l'ensemble des séjours des patients diabétiques, soit des patients pour lesquels un « code spécifique » du diabète avait été identifié sur un des séjours, nous avons constaté que (Fig. D2):

- Pour près d'un quart des patients (24 %), le diagnostic de la maladie n'avait pas été enregistré pendant l'année. Ceux-ci n'auraient donc pas été identifiés comme diabétiques sans l'enregistrement de la prescription des médicaments spécifiques du diabète
- Pour 21% de l'ensemble des séjours, aucun médicament ni aucun diagnostic de diabète n'avait été enregistré. Pour l'hospitalisation classique, ce chiffre était de 13% mais pour les hospitalisations de jour, il s'élevait à 66%.

Selon la classification ICD-9-CM, le type de diabète doit être enregistré par le 5<sup>e</sup> digit du code 250\*. Ce code n'est cependant présent que dans 61% des séjours. Le regroupement des séjours par patient permet d'identifier le type de diabète chez 75% d'entre eux. Parmi ceux-ci, les 3/4 sont de type II (76%), 20% sont de type I et 4% sont définis comme I et II, selon les séjours (Fig. D3). Les patients avec un diabète de type I représentent 58% des patients diabétiques à l'hôpital avant l'âge de 40 ans, 22% entre 40 et 59 ans et ils ne sont plus que 16% chez les patients plus âgés.

A l'hôpital, 35% des patients étaient traités uniquement par insuline, 23% recevaient des médicaments antidiabétiques oraux, tandis que les 2 types de traitement étaient administrés à 16% des malades au cours de l'année et 25% n'avaient reçu aucune médication antidiabétique dans le service (Fig. D3)

Parmi les facteurs de risque de la maladie diabétique, seule l'obésité apparaissait comme une maladie clinique susceptible d'être enregistrée sur les données hospitalières. Ce diagnostic n'était cependant encodé que pour 3.5 % des patients adultes hospitalisés (Fig. D4). Cette sous-évaluation franche discréditait toute analyse ultérieure de corrélation avec la pathologie diabétique.

L'analyse de tous les séjours regroupés des patients diabétiques nous a permis aussi de détecter certaines complications chroniques du diabète que n'auraient pas décelées une analyse séparée de ces séjours. Parmi les patients hospitalisés, 30% présenteraient des complications chroniques de la maladie. Les complications microvasculaires soit - rénales (12%), neurologiques (9%) et oculaires (7%) - sont enregistrées à un âge plus précoce que les complications vasculaires périphériques (8%) (Fig.D5).

Les complications aiguës de la maladie - acidocétose, coma hyperosmolaire ou autre coma - sont beaucoup moins fréquentes sur le RCM. Près de 3% des patients diabétiques sont admis dans un hôpital avec ces complications au moins une fois au cours de l'année. La proportion atteint 7.4% quand il s'agit de patients atteints de diabète type I et seulement 2.7% pour les patients de diabète type II.

La seule maladie associée au diabète, dont l'incidence est susceptible d'être étudiée sur RCM est l'infarctus du myocarde. En effet, celui-ci nécessite une hospitalisation, et l'enregistrement du diagnostic semble relativement fiable (sensibilité 87%, spécificité 55%)<sup>8</sup>. Nous avons estimé que la validité de l'enregistrement de ce diagnostic était suffisante pour permettre l'étude de l'association des deux maladies. A l'hôpital, l'incidence de l'infarctus du myocarde avant l'âge de 50 ans, serait 3 fois

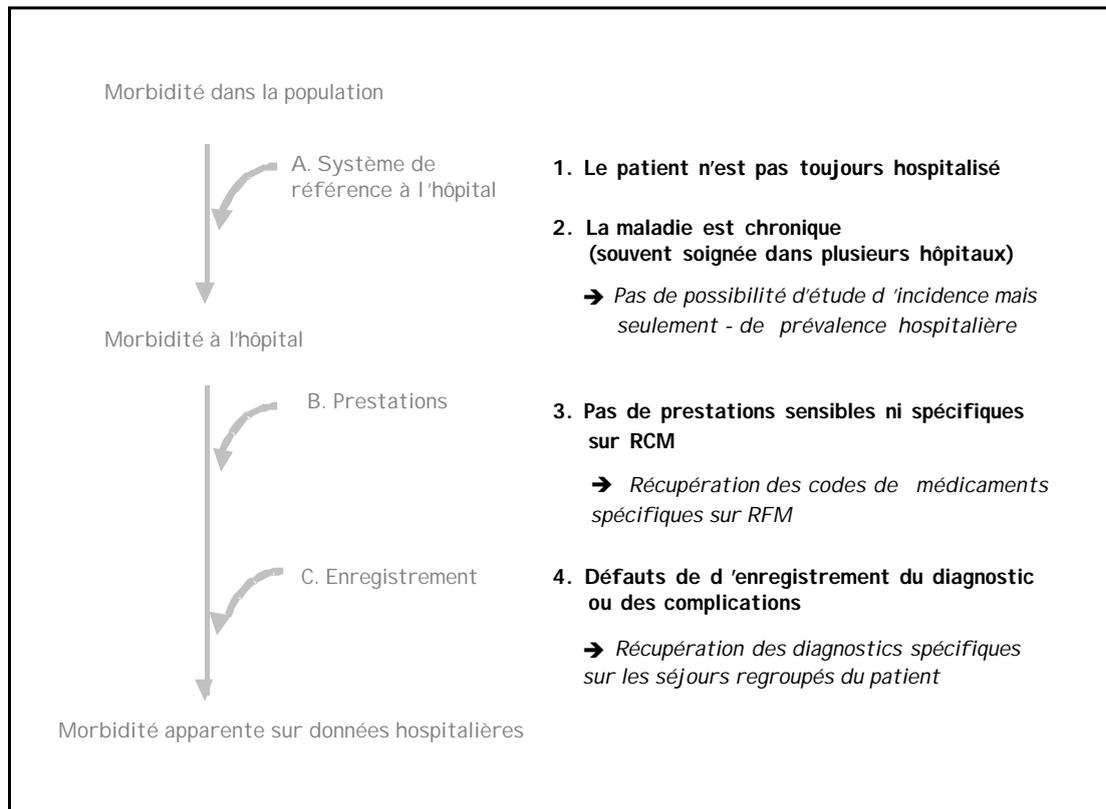
<sup>7</sup> Les codes identifiant les diabétiques dans notre analyse sont appelés « codes spécifiques » de la maladie, à savoir un code contenant le libellé de la maladie ou d'un diagnostic ou un traitement spécifique de celle-ci.

<sup>8</sup> Cfr. L'infarctus du myocarde – chapitre 3

plus élevée quand le patient est diabétique. Globalement, au moins 21% des patients hospitalisés présentant un infarctus sont diabétiques.

## Discussion

Les limites de validité de l'étude de la prévalence hospitalière du diabète sur les données hospitalières et les procédures pour surmonter les difficultés de ces analyses sont présentées dans la figure D6 ci-dessous.



**Fig. D6** Limites de validité de l'étude de la prévalence hospitalière du diabète sur les données hospitalières et les procédures pour surmonter les difficultés de ces analyses

Pour notre étude de prévalence, la validité des données RCM est bridée principalement par l'objectif usuel de cet enregistrement, qui est surtout financier, et par la procédure d'enregistrement qui en résulte.

En effet, le traitement actuel de ces données RCM impose le choix d'un diagnostic principal, déterminant dans la classement du séjour en Diagnosis Related Groups (DRG)<sup>1</sup>. Les maladies chroniques sont souvent refoulées au rang des diagnostics secondaires en faveur d'une maladie plus aiguë en relation avec le motif de l'admission du patient. Parfois même l'enregistrement de la maladie est négligé, quand le nombre de diagnostics secondaires est important comme c'est le cas chez beaucoup de patients diabétiques.

<sup>1</sup> Système de classification des séjours, utilisé pour tous les séjours dans les hôpitaux généraux, afin de permettre la comparer des dépenses et de la durée de l'hospitalisation dans chacun de ces hôpitaux.

Pour détecter le diabète présent chez un plus grand nombre de patients, nous avons vu l'importance de considérer à la fois les données diagnostiques et les données de prestations médicales, en l'occurrence dans cette analyse, la consommation en médicaments. Pour caractériser la maladie diabétique, il semble aussi essentiel de regrouper l'ensemble des séjours des patients pour lesquels un code spécifique du diabète avait été identifié sur un des séjours.

Nous avons comparé nos résultats de prévalence avec ceux présentés dans le dernier rapport du réseau de surveillance par médecins vigies, traitant un échantillon représentatif de 1.4% de la population belge. Le rapport mentionnait que 3.3% des patients des médecins généralistes étaient diabétiques (95% I.C. 3.2 –3.4). D'après notre analyse, la proportion des patients diabétiques à l'hôpital est deux fois plus élevée que dans la population fréquentant les médecins généralistes (Fig. D7). Le ratio est surtout élevé chez les patients plus jeunes, il atteindrait 3.4 avant l'âge de 45 ans tandis qu'il ne serait que de 1.6 pour les patients plus âgés. La prévalence globale plus importante à l'hôpital est due principalement à la moyenne d'âge nettement plus élevée des patients hospitalisés.

Par ailleurs, lors de l'enquête de santé par interview en 2001, 2.6 % de la population (95% IC 2.3-2.9) se déclaraient atteints de diabète. En comparant les résultats de cette enquête avec ceux de notre étude, on peut estimer que pour les personnes âgées de moins de 45 ans, le risque d'être dans un hôpital serait 5.6 fois plus élevé pour les diabétiques. Pour les personnes plus âgées, ce risque est seulement doublé.

Dans les trois analyses, le diabète semble plus fréquent chez les hommes plus jeunes, (au moins jusqu'à l'âge de 65 ans) et chez les femmes plus âgées. Dans l'enquête de santé par interview, la taille réduite de l'échantillon (10 000 personnes interrogées) ne permet pas d'affirmer que cette différence est significative dans la population générale.

La distinction du type de diabète (type I et II) est mal définie dans la codification ICD-9-CM et par ailleurs, l'enregistrement de celle-ci sur les RCM n'a aucun impact financier pour l'hôpital. Dès lors, la distinction présentée dans les résultats de notre analyse semble peu fiable.

Le rapport des médecins vigies montre que les traitements ambulatoires des patients diabétiques avérés sont médicamenteux pour les  $\frac{3}{4}$  d'entre eux (76.7%) : insuline pour 15.2%, médicaments antidiabétiques oraux pour 61.5% et les 2 types de médicaments dans 8% des cas {5}. A l'hôpital, on constate une proportion plus importante de patients pour lesquels on n'enregistre pas de médicament ainsi que de patients traités par insuline. La situation de déséquilibre de la maladie lors de l'hospitalisation du patient pourrait expliquer cette dernière constatation.

Les complications aiguës de la maladie sont probablement bien enregistrées. En effet, cet enregistrement ne pâtit pas des limites de validité énoncées dans la figure D6 : dans ces situations, le malade est presque toujours hospitalisé ; l'enregistrement de ces complications est important pour la classement du séjour en DRG ; le diagnostic et la codification de celles-ci sont assez précis. Peut-être les résultats sont-ils légèrement sous-évalués car l'urgence de ces situations amène sans doute plus souvent un même patient dans différents hôpitaux au cours de l'année.

Les défauts d'enregistrements limitent spécialement l'étude des maladies chroniques associées au diabète. Ainsi, d'après notre analyse, le diagnostic d'obésité était enregistré chez seulement 3.5 % des patients adultes hospitalisés alors que dans l'enquête nationale de santé en 2001, 11.7% des personnes interrogées, de plus de 18 ans étaient déclarées obèses {1}. Dans les deux distributions, la prévalence de l'obésité augmente jusqu'à l'âge de 65 ans, puis diminue progressivement après cet âge. Chez la femme cependant, la fréquence de l'obésité augmente encore légèrement jusqu'à un âge plus avancé (Fig. D8) {1}.. Il est possible que cet état de fait explique en partie la fréquence également plus élevée des femmes âgées atteintes de diabète.

La fréquence des complications chroniques du diabète augmente avec l'ancienneté de la maladie. On pourrait dès lors s'attendre à une proportion plus importante de complications chez les patients hospitalisés dans les tranches d'âge plus élevées. D'après notre étude, cette proportion est plus faible. Deux hypothèses peuvent expliquer partiellement cette constatation :

- La fréquence des complications du diabète relevées sur les données hospitalières serait sous-estimée. En effet, celles-ci figurent souvent en diagnostics secondaires et leur enregistrement serait donc parfois incomplet.
- Les patients diabétiques présentant des complications de la maladie ont une moindre espérance de vie. On peut dès lors supposer qu'une part non négligeable des patients âgés, identifiés comme diabétiques, est atteinte de la maladie depuis relativement peu de temps et ne présente pas ou peu de complications.

Selon notre étude, le risque d'être atteint d'un infarctus du myocarde est trois fois plus important quand on est diabétique. Nos résultats semblent être corroborés par plusieurs études internationales, dans lesquelles ce risque est a été évalué entre 2 et 6 fois plus important selon les antécédents du patient {6}. Avant d'attribuer un certain crédit à nos résultats sur les données hospitalières, il est indispensable de considérer les limites de validité des deux types d'étude, prévalence du diabète et incidence de l'infarctus.

## Conclusion

L'intérêt de l'utilisation des données hospitalières pour l'étude du diabète réside dans les avantages suivants :

- la maladie est dépistée presque systématiquement en milieu hospitalier
- l'étude ne se confine pas à celle d'un échantillon, elle permet une photographie de l'ensemble de la population belge hospitalisée
- la collecte des données se fait en routine et ne nécessite donc pas de moyens supplémentaires.

Les résultats de notre étude révèlent que pour 7% des patients hospitalisés, un code identifiant la maladie a été enregistré au moins une fois dans l'année. Si nous comparons ces résultats avec ceux de l'analyse d'autres sources de données, nous pouvons estimer assez probant ce calcul de prévalence de la maladie à l'hôpital bien que celle-ci reste sans doute légèrement sous-estimée. Par contre, le type de diabète ou les maladies associées semblent mal enregistrés sur les données hospitalières.

Les limites importantes de validité de l'analyse de cette maladie chronique sur les données RCM ont pu être partiellement levées en adoptant les stratégies suivantes dans l'analyse :

- En préalable à la sélection du fichier, l'examen détaillé de la classification ICD-9-CM permet de sélectionner non seulement les séjours mentionnant le code de la maladie mais aussi tout séjour incluant le code d'une maladie ou d'un traitement spécifiquement associé au diabète.
- Tout diagnostic (principal ou secondaire) doit être inclus dans l'analyse car la maladie chronique apparaît souvent en diagnostic secondaire.
- Les médicaments spécifiques, enregistrés sur RFM, permettent parfois de détecter la maladie quand son enregistrement fait défaut sur RCM.
- L'analyse de tous les séjours regroupés par patient permet de récupérer certains diagnostics ou procédures médicales qui ne seraient pas détectés sur une analyse séparée de ces séjours.

Les principaux problèmes de validité de notre analyse proviennent des directives ministérielles d'enregistrement de ces données dont l'objectif de traitement est financier. Le classement des séjours dans le système des APR-DRG qui prennent plus en compte les diagnostics secondaires, améliorera probablement le niveau de l'enregistrement de cette pathologie sur RCM.

## Références bibliographiques

1. Gisle L, Buziarsist J, Van der Heyden J, Demarest S, Miermans PJ, Sartor F, Van Oyen H, Tafforeau J. Enquête de Santé par Interview, Belgique, 2001. Section d'Epidémiologie, 2002; Bruxelles. Institut Scientifique de Santé Publique, IPH/EPI REPORTS N° 2002 – 22
2. Puddu M, Van Casteren V. Health information on chronic diseases : incidence and prevalence of diabetes mellitus in Belgium, 2000 from the Belgian Sentinel General Practice Network. Annual Eupha Meeting 2001. Abstract book
3. Hortulanus-Becq D, Lefèvre PJ, Jeanjean MF. Le diabète dans la province belge du Luxembourg : fréquence, importance de l'épreuve de surcharge glucosée orale et d'une glycémie à jeûn discrètement accrue. Diabète et métabolisme 1990 ; 16 :311-7
4. Newton Katherine M., Wagner Edward H, Ramsey Scott D., McCullooch David, Evans Rhian, Sandhu Nirmala and Davis Connie The use of automated Data to identify complications and comorbidities of Diabetes : a validation studyJ. Clin. Epidemiol Vol. 52, No. 3, pp. 199-207, 1999
5. Puddu M. Diabète sucré type I et II en 2000 : prévalence et prise en charge thérapeutique <http://www.iph.fgov.be/epidemio/epien/medven/sentibul/sentibul2.pdf>  
Puddu, Van Casteren V. Diabetes in 2000: Regionale verschillen in incidentie, prevalentie en behandeling. Resultaten van de huisartsen Gezondheidsindicatoren 2000. Eds. Cloots H, Hooft P, Smets H. Ministerie van de Vlaamse Gemeenschap, 2002:87-91.
6. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med. 1998;339:229-234

## Collaborateur:

F. Capet - Institut Scientifique de Santé Publique, section d'épidémiologie

M. Puddu, Institut Scientifique de Santé Publique, section d'épidémiologie, pour son analyse des données du réseau des médecins 2000.

Figures illustrant les résultats

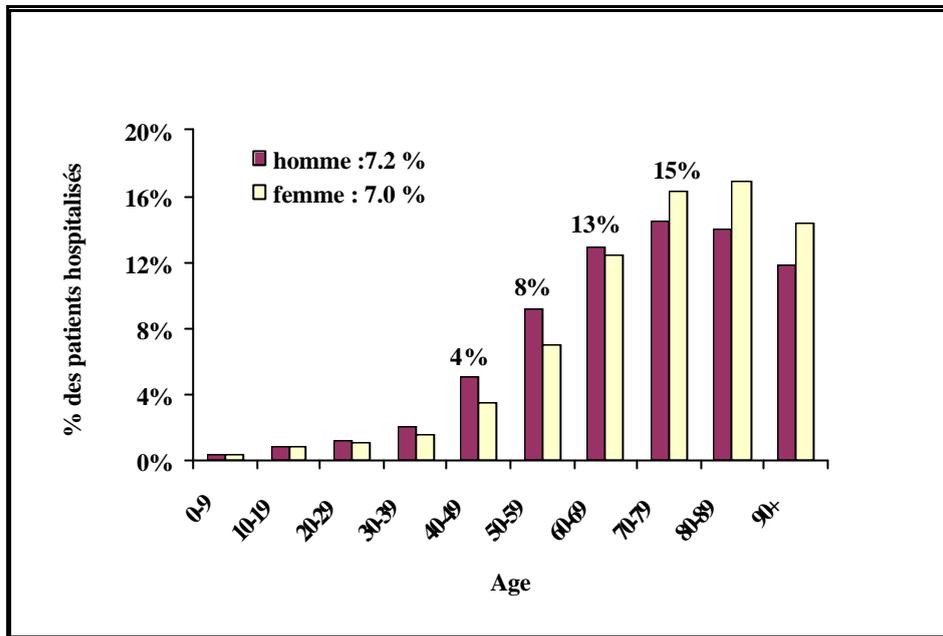


Fig. D1 Prévalence du diabète à l'hôpital – Etude sur Résumé Clinique minimum + Résumé Financier Minimum- 1997

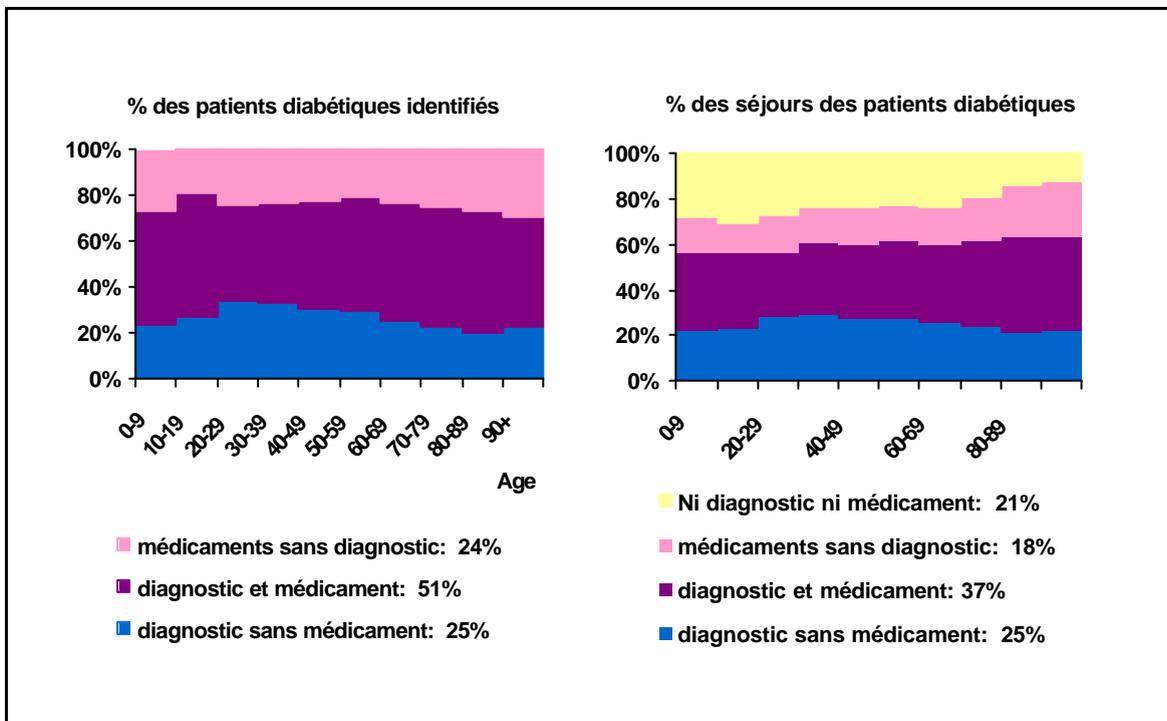


Fig. D2 Apport de la méthodologie d'analyse des données hospitalière dans l'étude du diabète : utilisation des RCM + RFM et regroupement des séjours par patients - L'enregistrement de la prescription médicamenteuse permet une bonne part de l'identification des patients diabétiques

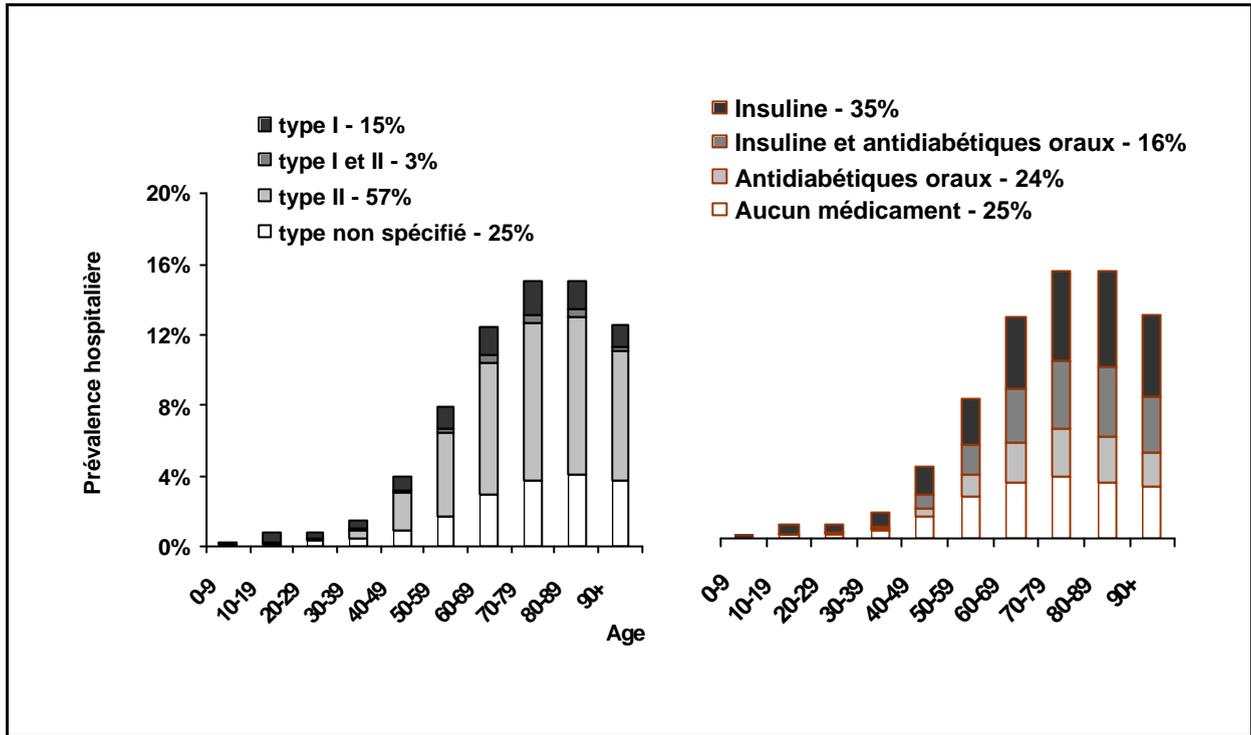


Fig. D3 Type de diabète et type de traitement médicamenteux chez les patients diabétiques hospitalisés – sur Résumé Clinique minimum + Résumé Financier Minimum – 1997 .

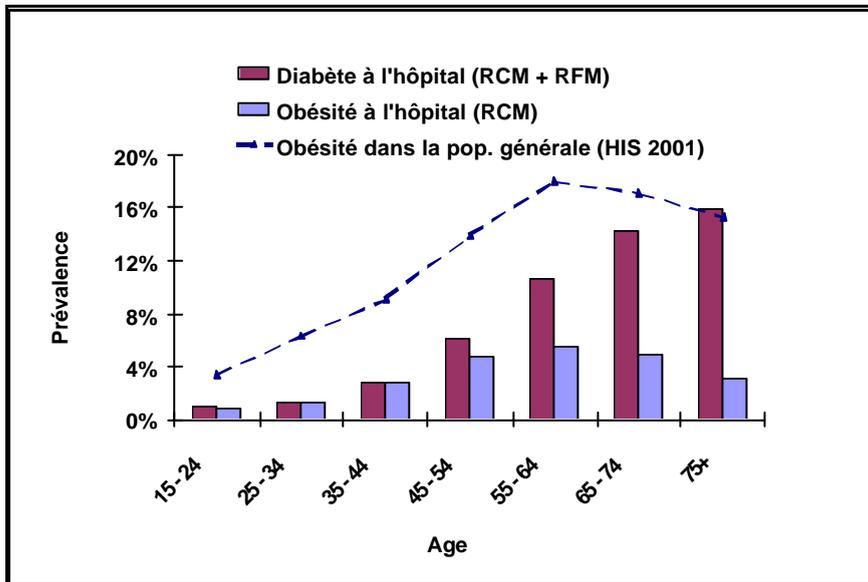


Fig. D4 Prévalence du diabète et de l'obésité chez les patients hospitalisés – sur Résumé Clinique minimum + Résumé Financier Minimum – 1997 , comparée à la prévalence de l'obésité selon l'enquête de santé par interview en 2001

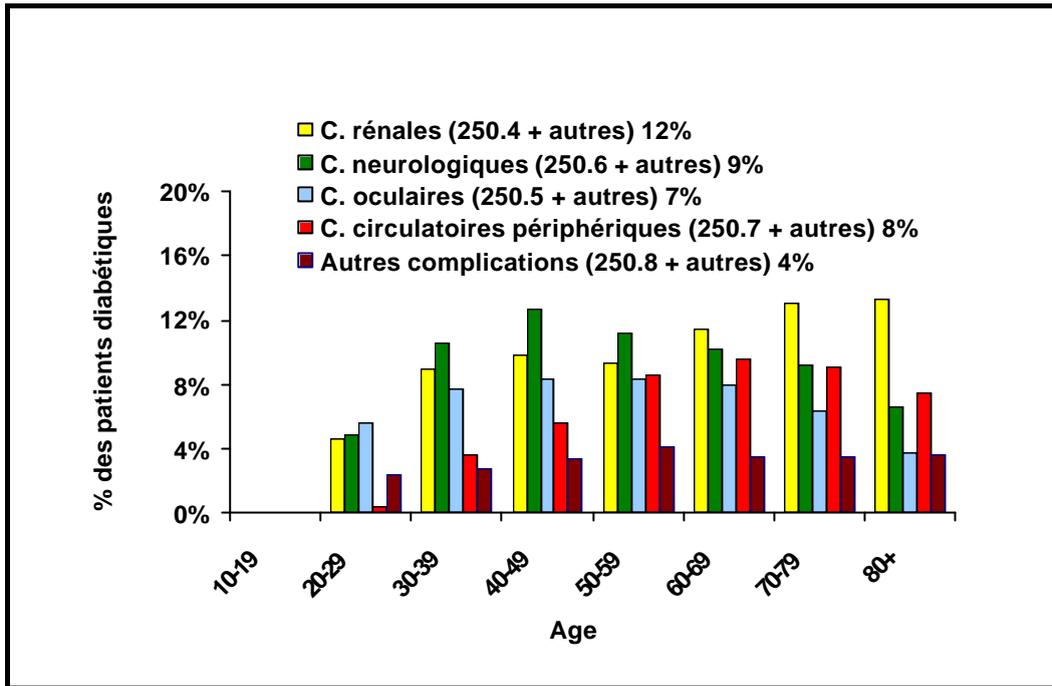


Fig. D5 Prévalence des complications chroniques enregistrées parmi les patients diabétiques hospitalisés - Etude du Résumé Clinique Minimum et du Résumé Financier Minimum 1997

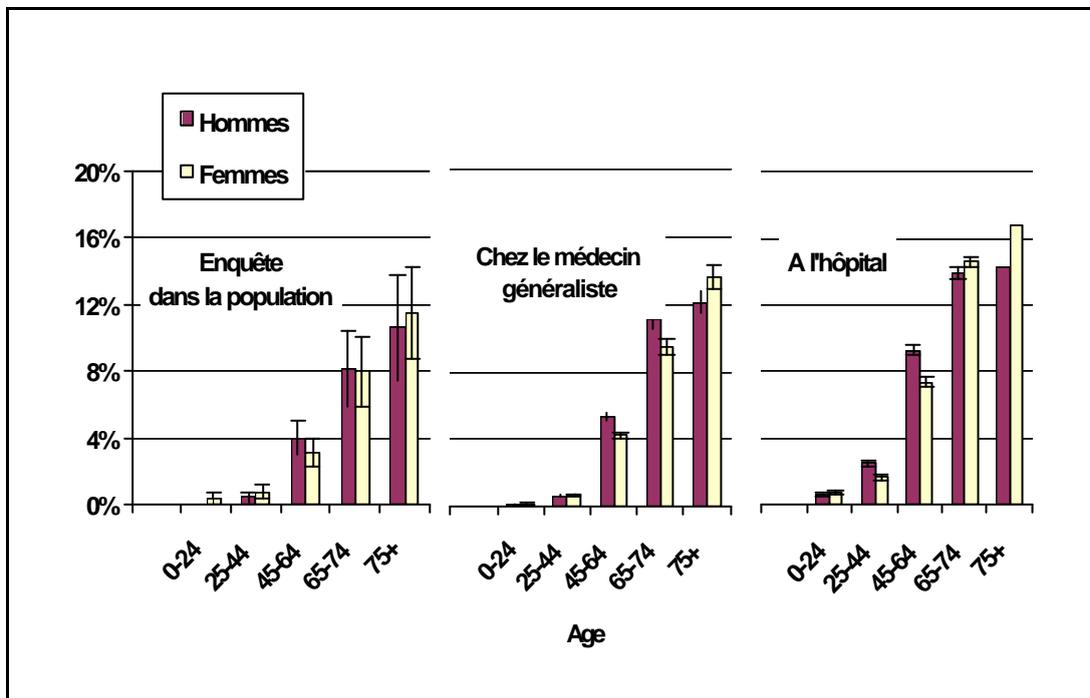
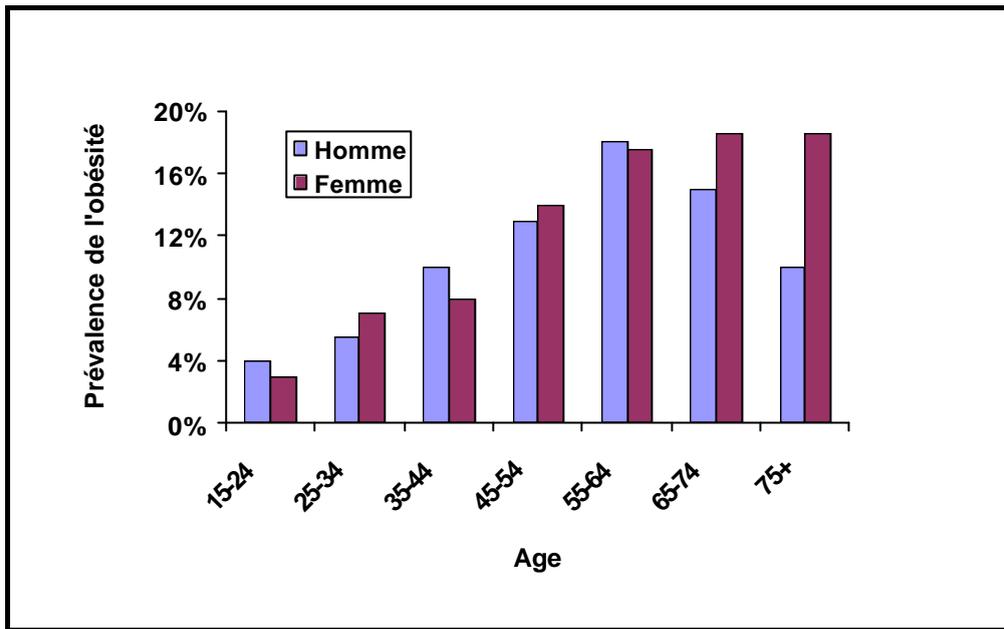


Fig. D7 Comparaison des prévalences du diabète - dans la population générale (selon la dernière enquête de santé par interview en 2001) - chez les patients traités par le médecin généraliste (selon le rapport des médecins vigies 2000) - et chez les patients hospitalisés (selon l'analyse des Résumés Cliniques Minimum et des Résumés Financiers minimum 1997)



*Fig.D8 Prévalence de l'obésité dans la population générale selon l'enquête de santé par interview en 2001*

## IX. Conclusions et perspectives

Les résultats de notre étude montrent que ce système d'information sanitaire, actuellement encore axé vers un objectif financier, peut être un outil très précieux pour la surveillance épidémiologique des maladies dans certaines limites de validité.

Nous avons vu la contribution que l'analyse de ces données hospitalières pouvait apporter à différents registres ou réseaux de surveillance (incidence des cancers, des maladies infectieuses...) et même l'importance que celle-ci pouvait avoir à défaut d'autres résultats d'analyses nationales (prévalence hospitalière du diabète, ..)

En effet, le problème principal rencontré dans la réalisation de notre étude résidait plus dans la difficulté de trouver d'autres sources de données comparables plutôt que dans la difficulté des analyses. Les analyses plus sophistiquées auraient été d'ailleurs incongrues avant d'avoir validé l'enregistrement des données pour de calculs simples d'incidence ou de prévalence.

Outre l'intérêt de ces analyses pour la santé publique, les avantages principaux de l'exploitation de la base des données hospitalières à des fins épidémiologiques sont les suivants:

- L'étude ne se confine pas à celle d'un échantillon, elle permet une photographie de l'ensemble de la population belge.
- Certaines maladies sont diagnostiquées ou traitées principalement en milieu hospitalier et pas en milieu ambulatoire
- La collecte des données se fait en routine et ne nécessite donc pas de moyens supplémentaires.

La moitié de nos analyses repose principalement sur les données diagnostiques. Cependant nous ne pouvons négliger l'apport essentiel des données de prestations médicales, aussi bien dans les études de prévalence (par ex. - du diabète) que dans les études d'incidence. Pour ces dernières, la confrontation de nos résultats sur les données RCM (séjours des malades opérés) et sur les données de l'INAMI (nombre d'interventions) laisse supposer que l'enregistrement des données RCM pour les années 1993 et 1994 n'était pas exhaustif.

Les principales limites de validité de notre étude sont expliquées et résumées dans le chapitre "Méthodologie". A ce stade, notre analyse a pu partiellement les surmonter. Cependant, afin que ce système d'information sanitaire contribue de manière effective aux études épidémiologiques courantes, quelques modifications seront ou devront être apportées dans l'avenir:

- Au niveau de l'enregistrement, un changement mineur dans le format actuel des données (marquant les nouveaux diagnostics) permettrait d'éviter les biais dans les analyses d'incidence basées sur les diagnostics corroborés par les prestations, cette incidence calculée dépendant dans ce cas de la qualité des soins.
- Au niveau du traitement des données, l'évolution vers le système APR-DRG, favorisera certainement l'enregistrement plus exhaustif des diagnostics.
- Au niveau de l'analyse, un traitement récent des données hospitalières lie les données RCM et les données de prestations INAMI par un numéro unique pour le patient, quelle que soit la structure de soins fréquentée, dans le secteur hospitalier. Le traitement de ces données liées permettra notamment de surmonter le problème des enregistrements multiples d'un même patient dans différents hôpitaux.

## **x. Annexe : Programmes SAS – Utilisation de la base de données hospitalières**

---

La base de données de travail se construit à partir des bases de données nationales du RCM. Elle comprend les séjours hospitaliers correspondant à la pathologie traitée. Ces séjours sont sélectionnés à l'aide d'une série de codes ICD-9-CM (de diagnostics et d'interventions) et de codes INAMI. Cette partie du travail est réalisée au niveau du Ministère de la Santé.

Le programme SAS suivant permet de faire le lien entre les trois bases de départ : la base contenant les informations sur le séjour hospitalier et sur le patient, la base des diagnostics et celle des interventions. Comme les données sont anonymes, ne sont pas connus: le nom, la date de naissance et le code postal du patient ainsi que l'hôpital et le jour d'admission à l'hôpital. La date d'anniversaire des patients est fixée arbitrairement au 30 juin.

L'exemple suivant est donné pour le traitement des données sur le cancer de la thyroïde

- **1<sup>ère</sup> base : les caractéristiques du séjour**

```
libname adress "/workspace/thimus/thimus_thyr";
data base;
set adress.CCSZH97;

keep zhtv zhstpj zhstps no zhdtop zhvbnr
    patsex patlft catage patnr patarr klarr
    dghdg thyroid1;

sub3=substr(dghdg,1,3); /* diag principal */
if sub3='193' then thyroid1=1; else thyroid1=0;

if patsex in (1 2); /* 0 = sexe inconnu */

if patlft<5 then catage='0-4ans ';
else if patlft<10 then catage='5-9ans';
else if patlft<15 then catage='10-14ans';
else if patlft<20 then catage='15-19ans';
else if patlft<25 then catage='20-24ans';
else if patlft<30 then catage='25-29ans';
else if patlft<35 then catage='30-34ans';
else if patlft<40 then catage='35-39ans';
else if patlft<45 then catage='40-44ans';
else if patlft<50 then catage='45-49ans';
else if patlft<55 then catage='50-54ans';
else if patlft<60 then catage='55-59ans';
else if patlft<65 then catage='60-64ans';
else if patlft<70 then catage='65-69ans';
else if patlft<75 then catage='70-74ans';
else if patlft<80 then catage='75-79ans';
else if patlft<85 then catage='80-84ans';
else if patlft<90 then catage='85-89ans';
else catage='>90 ans';

proc sort;
by zhtv zhstpj zhstps no zhdtop zhvbnr; run;
```

**2<sup>ème</sup> base : les diagnostics du séjour**

```

data base2;
set adress.CISDG97; /* 1 ligne par diagnostic */

keep zhtv zhstpj zhstps no zhdtop zhvbnr
    dgcodb;

proc sort;
by zhtv zhstpj zhstps no zhdtop zhvbnr;
run;

proc transpose out=trans1;
by zhtv zhstpj zhstps no zhdtop zhvbnr;
var dgcodb;
run;

data trans1; set trans1; /* 1 ligne par séjour */

keep zhtv zhstpj zhstps no zhdtop zhvbnr
    thyroid2 papil vesic medul anapla mixt
    codeM gangl chimio radio;

/* création de nouvelles variables */

array arr{137} coll-coll137;
array a{137} a1-a137;
array b{137} b1-b137;
array c{137} c1-c137;
array d{137} d1-d137;
array e{137} e1-e137;
array f{137} f1-f137;
array g{137} g1-g137;
array h{137} h1-h137;
array i{137} i1-i137;
array j{137} j1-j137;
do x=1 to 137;
sub1=substr(arr{x},1,1);
sub3=substr(arr{x},1,3);
sub4=substr(arr{x},1,4);
sub5=substr(arr{x},1,5);
if sub3='193' then a{x}=1; else a{x}=0;
if sub1='M' then b{x}=1; else b{x}=0;
if sub3='196' then c{x}=1; else c{x}=0;
if sub5 in ('M8050' 'M8260') then d{x}=1; else d{x}=0;
if sub4='M833' then e{x}=1; else e{x}=0;
if sub4='M851' then f{x}=1; else f{x}=0;
if sub5 in ('M8020' 'M8021') then g{x}=1; else g{x}=0;
if sub4='M834' then h{x}=1; else h{x}=0;
if sub4='V581' then i{x}=1; else i{x}=0;
if sub4='V580' then j{x}=1; else j{x}=0;
end;
if sum(of a1-a137) ne 0 then thyroid2=1; else thyroid2=0;
if sum(of b1-b137) ne 0 then codeM=1; else codeM=0;
if sum(of c1-c137) ne 0 then gangl=1; else gangl=0;
if sum(of d1-d137) ne 0 then papil=1; else papil=0;
if sum(of e1-e137) ne 0 then vesic=1; else vesic=0;
if sum(of f1-f137) ne 0 then medul=1; else medul=0;
if sum(of g1-g137) ne 0 then anapla=1; else anapla=0;
if sum(of h1-h137) ne 0 then mixt=1; else mixt=0;
if sum(of i1-i137) ne 0 then chimio=1; else chimio=0;
if sum(of j1-j137) ne 0 then radio=1; else radio=0;

```

- **3<sup>ème</sup> base : les interventions du séjour**

```

data base2;
set adress.CISIC97; /* 1 ligne par intervention */

keep zhtv zhstpj zhstps no zhdtop zhvbnr
    iccodb;

proc sort;
by zhtv zhstpj zhstps no zhdtop zhvbnr;
run;

proc transpose out=trans2;
by zhtv zhstpj zhstps no zhdtop zhvbnr;
var iccodb;
run;

data trans2; set trans2; /* 1 ligne par séjour */

keep zhtv zhstpj zhstps no zhdtop zhvbnr
    lobect partiel total subnal subnal lingual;

/* création de nouvelles variables */

array brr{317} coll-col317;
array c{317} c1-c317;
array d{317} d1-d317;
array e{317} e1-e317;
array f{317} f1-f317;
array g{317} g1-g317;
do x=1 to 317;
sub3=substr(brr{x},1,3);
if sub3='062' then c{x}=1; else c{x}=0;
if sub3='063' then d{x}=1; else d{x}=0;
if sub3='064' then e{x}=1; else e{x}=0;
if sub3='065' then f{x}=1; else f{x}=0;
if sub3='066' then g{x}=1; else g{x}=0;
end;
if sum(of c1-c317) ne 0 then lobect=1; else lobect=0;
if sum(of d1-d317) ne 0 then partiel=1; else partiel=0;
if sum(of e1-e317) ne 0 then total=1; else total=0;
if sum(of f1-f317) ne 0 then subnal=1; else subnal=0;
if sum(of g1-g317) ne 0 then lingual=1; else lingual=0;

```

- **La fusion des trois bases**

La clé de la fusion est composée des variables

- zhtv = le type de séjour
- zhstpj = année
- zhstps = semestre
- no = numéro anonyme de l'hôpital
- zhdtop = année et mois d'admission
- zhvbnr = numéro de séjour

```

data base;
merge base(in=in1) trans1(in=in2) trans2(in=in3);
by zhtv zhstpj zhstps no zhdtop zhvbnr;
if in1 and in2 and in3;

/* création de nouvelles variables */

```

```

if thyroid1=1 or thyroid2=1 then thyroid=1; else thyroid=0;

if thyroid1=1 then diag=1;
else if thyroid2=1 then diag=2;
else diag=3;

```

- **Le problème des doublons de patient**

Le numéro de patient est normalement unique dans un même hôpital. Or certains numéros de patients regroupent dans un même hôpital des séjours dont le sexe et l'âge sont différents (plus d'un an de différence). Avant de mettre la base sous format patient, une nouvelle variable patlft8 est créée pour différencier les doublons. Pour l'analyse sur le cancer de la thyroïde jusqu'à huit patients différents ont été retrouvés pour un même numéro de patient dans le même hôpital.

```

/* CREATION VARIABLE PATLFT8 */

proc sort;
by no patnr patsex;
run;

data baselb; set base;
by no patnr patsex;
if first.patsex;      /* 1er patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data basel;
merge base(in=in1) baselb(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft2=1;      /* meme patient */
else patlft2=2;      /* autre patient */
run;

data data1; set basel;
if patlft2=1;
patlft8=1;
drop age patlft2;
run;

data base2; set basel;
if patlft2=2;
drop age patlft2;
run;

data base2b; set base2;
by no patnr patsex;
if first.patsex;      /* 2eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base2;
merge base2(in=in1) base2b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft3=2;      /* meme patient */

```

```

else patlft3=3; /* autre patient */
run;

data data2; set base2;
if patlft3=2;
patlft8=2;
drop age patlft3;
run;

data base3; set base2;
if patlft3=3;
drop age patlft3;
run;

data base3b; set base3;
by no patnr patsex;
if first.patsex; /* 3eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base3;
merge base3(in=in1) base3b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft4=3;
else patlft4=4;
run;

data data3; set base3;
if patlft4=3;
patlft8=3;
drop age patlft4;
run;

data base4; set base3;
if patlft4=4;
drop age patlft4;
run;

data base4b; set base4;
by no patnr patsex;
if first.patsex; /* 4eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base4;
merge base4(in=in1) base4b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft5=4; /* meme patient */
else patlft5=5; /* autre patient */
run;

data data4; set base4;
if patlft5=4;
patlft8=4;
drop age patlft5;
run;

data base5; set base4;
if patlft5=5;
drop age patlft5;
run;

```

```

data base5b; set base5;
by no patnr patsex;
if first.patsex;      /* 5eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base5;
merge base5(in=in1) base5b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft6=5;      /* meme patient */
else patlft6=6;      /* autre patient */
run;

data data5; set base5;
if patlft6=5;
patlft8=5;
drop age patlft6;
run;

data base6; set base5;
if patlft6=6;
drop age patlft6;
run;

data base6b; set base6;
by no patnr patsex;
if first.patsex;      /* 6eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base6;
merge base6(in=in1) base6b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft7=6; /* meme patient */
else patlft7=7; /* autre patient */
run;

data data6; set base6;
if patlft7=6;
patlft8=6;
drop age patlft7;
run;

data base7; set base6;
if patlft7=7;
drop age patlft7;
run;

data base7b; set base7;
by no patnr patsex;
if first.patsex;      /* 7eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base7;
merge base7(in=in1) base7b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1

```

```

then patlft8=7; /* meme patient */
else patlft8=8; /* autre patient */
run;

data data7; set base7;
if patlft8=7;
drop age;
run;

data base8; set base7;
if patlft8=8;
drop age patlft8;
run;

data base8b; set base8;
by no patnr patsex;
if first.patsex; /* 8eme patient */
rename patlft=age;
keep no patnr patsex patlft;
run;

data base8;
merge base8(in=in1) base8b(in=in2);
by no patnr patsex;
if in1;
if patlft=age or patlft=age+1 or patlft=age-1
then patlft8=8; /* meme patient */
else patlft8=9; /* autre patient */
run;

data base;
set data1 data2 data3 data4 data5 data6 data7 base8;
drop age;

proc freq; tables patlft8; run;

/* ECRITURE DE LA BASE */

data adress.base97;
set base;

proc contents;
run;

```

- **Base de données sous format patient**

Les séjours d'un même patient dans le même hôpital sont regroupés. Les codes de diagnostic ou d'intervention sont recherchés sur l'ensemble des séjours du patient. Les patients changeant d'hôpital durant la même année sont comptés double. La clé pour différencier les patients est la suivante :

numéro anonyme de l'hôpital, numéro de patient, sexe et âge plus ou moins 1.

```

libname adress "/workspace/thimus/thimus_thyr";
data base;
set adress.base97;

proc sort; by no patnr patsex patlft8;
run;

proc transpose data=base out=trans1;
by no patnr patsex patlft8; var thyroid;
run;

```

```
proc transpose data=base out=trans2;
by no patnr patsex patlft8; var thyroid1;
run;

proc transpose data=base out=trans3;
by no patnr patsex patlft8; var thyroid2;
run;

proc transpose data=base out=trans10;
by no patnr patsex patlft8; var total;
run;

proc transpose data=base out=trans11;
by no patnr patsex patlft8; var partiel;
run;

proc transpose data=base out=trans12;
by no patnr patsex patlft8; var subnal;
run;

proc transpose data=base out=trans13;
by no patnr patsex patlft8; var lobect;
run;

proc transpose data=base out=trans14;
by no patnr patsex patlft8; var lingual;
run;

proc transpose data=base out=trans15;
by no patnr patsex patlft8; var chimio;
run;

proc transpose data=base out=trans16;
by no patnr patsex patlft8; var radio;
run;

proc transpose data=base out=trans17;
by no patnr patsex patlft8; var papil;
run;

proc transpose data=base out=trans18;
by no patnr patsex patlft8; var vesic;
run;

proc transpose data=base out=trans19;
by no patnr patsex patlft8; var medul;
run;

proc transpose data=base out=trans20;
by no patnr patsex patlft8; var anapla;
run;

proc transpose data=base out=trans21;
by no patnr patsex patlft8; var mixt;
run;

proc transpose data=base out=trans22;
by no patnr patsex patlft8; var codeM;
run;

proc transpose data=base out=trans23;
by no patnr patsex patlft8; var gangl;
run;

data trans1; set trans1;
if sum(of coll-col175) ne 0 then thyroid=1; else thyroid=2;
```

```

keep no patnr patsex patlft8 thyroid;
run;

data trans2; set trans2;
if sum(of coll-col74) ne 0 then thyroid1=1; else thyroid1=2;
keep no patnr patsex patlft8 thyroid1;
run;

data trans3; set trans3;
if sum(of coll-col74) ne 0 then thyroid2=1; else thyroid2=2;
keep no patnr patsex patlft8 thyroid2;
run;

data trans10; set trans10;
totalp=sum(of coll-col74);
keep no patnr patsex patlft8 totalp;
run;

data trans11; set trans11;
partielp=sum(of coll-col74);
keep no patnr patsex patlft8 partielp;
run;

data trans12; set trans12;
subnalp=sum(of coll-col74);
keep no patnr patsex patlft8 subnalp;
run;

data trans13; set trans13;
lobectp=sum(of coll-col74);
keep no patnr patsex patlft8 lobectp;
run;

data trans14; set trans14;
lingualp=sum(of coll-col74);
keep no patnr patsex patlft8 lingualp;
run;

data trans15; set trans15;
if sum(of coll-col74) ne 0 then chimio=1; else chimio=0;
keep no patnr patsex patlft8 chimio;
run;

data trans16; set trans16;
if sum(of coll-col74) ne 0 then radio=1; else radio=0;
keep no patnr patsex patlft8 radio;
run;

data trans17; set trans17;
if sum(of coll-col74) ne 0 then papil=1; else papil=0;
keep no patnr patsex patlft8 papil;
run;

data trans18; set trans18;
if sum(of coll-col74) ne 0 then vesic=1; else vesic=0;
keep no patnr patsex patlft8 vesic;
run;

data trans19; set trans19;
if sum(of coll-col74) ne 0 then medul=1; else medul=0;
keep no patnr patsex patlft8 medul;
run;

data trans20; set trans20;
if sum(of coll-col74) ne 0 then anapla=1; else anapla=0;
keep no patnr patsex patlft8 anapla;
run;

```

```

data trans21; set trans21;
if sum(of coll-col74) ne 0 then mixt=1; else mixt=0;
keep no patnr patsex patlft8 mixt;
run;

data trans22; set trans22;
if sum(of coll-col74) ne 0 then codeM=1; else codeM=0;
keep no patnr patsex patlft8 codeM;
run;

data trans23; set trans23;
if sum(of coll-col74) ne 0 then gangl=1; else gangl=0;
keep no patnr patsex patlft8 gangl;
run;

data ajout; set base;
keep no patnr patsex patlft8 patlft catage zhdtop;
rename zhdtop=zhdtop1; /* lere admission de l'annee */
by no patnr patsex patlft8;
if first.patlft8;
run;

data base;
merge trans1(in=in1) trans2(in=in2) trans3(in=in3) trans10(in=in10)
trans11(in=in11)
      trans12(in=in12) trans13(in=in13) trans14(in=in14) trans15(in=in15)
      trans16(in=in16) trans17(in=in17) trans18(in=in18) trans19(in=in19)
      trans20(in=in20) trans21(in=in21) trans22(in=in22) trans23(in=in23)
      ajout(in=in25);
by no patnr patsex patlft8;
if in1 and in2 and in3 and in10 and in11 and in12 and in13 and in14 and in15
   and in16 and in17 and in18 and in19 and in20 and in21 and in22 and in23 and
in25;

if thyroid1=1 then diag=1;
else if thyroid2=1 then diag=2;
else diag=3;
run;

/**** CREATION BASE PAR PATIENTS *****/

data adress.basept97;
set base;

proc contents;
run;

```

---