

Maladies infectieuses

Surveillance des infections invasives à méningocoque en France métropolitaine en 2005

Évaluation quantitative par la méthode
de capture-recapture à trois sources

Sommaire

Abréviations	2
1. Introduction	3
2. Objectifs	4
3. Matériels et méthodes	5
3.1 Description des systèmes de surveillance des infections invasives à méningocoque (IIM) en France	5
3.2 Application de la méthode capture-recapture aux infections invasives à méningocoque	5
3.3 Outils d'analyse statistique	8
4. Résultats	9
4.1 Cas retenus pour l'analyse	9
4.2 Caractéristiques des cas déclarés par source	9
4.3 Distribution des cas selon leur(s) source(s) d'appartenance	11
4.4 Étude de la dépendance entre les sources	11
4.5 Analyse stratifiée	12
4.6 Évolution de l'exhaustivité selon les sources [2,8]	13
4.7 Représentativité des sources	14
4.8 Évaluation des taux d'incidence à partir des effectifs estimés (tableau 16)	14
4.9 Estimation de l'exhaustivité de la déclaration obligatoire par une analyse capture-recapture restreinte aux laboratoires participants à Epibac	14
5. Discussion	16
5.1 Résultats	16
5.2 Méthode	17
6. Conclusion	19
Références bibliographiques	20
Annexe 1 – Principe et conditions de la méthode capture-recapture	21
Annexe 2 – Illustrations de l'analyse capture-recapture globale	25
Annexe 3 – Exhaustivité de la surveillance des IIM par le réseau Epibac	32
Annexe 3 bis – Illustrations de l'analyse capture-recapture restreinte aux laboratoires du réseau Epibac	38

Surveillance des infections invasives à méningocoque en France métropolitaine en 2005

Évaluation quantitative par la méthode de capture-recapture à trois sources

Auteurs du rapport

Franck Berger, Département des maladies infectieuses (DMI), Institut de veille sanitaire (InVS), Saint-Maurice ; Département d'épidémiologie et de santé publique, école du Val-de-Grâce, Paris
Isabelle Parent du Châtelet, DMI, InVS, Saint-Maurice
Pascale Bernillon DMI, InVS, Saint-Maurice
Anne Gallay DMI, InVS, Saint-Maurice

Contribution

Muhamed-Kheir Taha Centre national de référence (CNR) des méningocoques, Institut Pasteur, Paris
Daniel Lévy-Bruhl, DMI, InVS, Saint-Maurice
Scarlett Georges, DMI, InVS, Saint-Maurice

Remerciements

Les auteurs remercient l'ensemble des cliniciens et biologistes et les professionnels des Ddass qui ont contribué à un ou plusieurs systèmes de surveillance des infections invasives à méningocoque (IIM) en 2005, Pascale Vienne du CNR des méningocoques, Catherine Maine et Edith Laurent à l'InVS ainsi que Bruno Hubert (Cellule de l'InVS en région (Cire) Pays-de-la-Loire) pour sa relecture et ses conseils.

Abréviations

Cire	Cellule de l'InVS en région
CNR	Centre national de référence des méningocoques
Ddass	Direction départementale des affaires sanitaires et sociales*
DGS	Direction générale de la santé
DO	Déclaration obligatoire
IC	Intervalle de confiance
IIM	Infection invasive à méningocoque
InVS	Institut de veille sanitaire
MLST	Multi Locus Sequence Typing
LCR	Liquide céphalo-rachidien
PCR	Polymerase Chain Reaction

* Depuis le 1^{er} avril 2010, les Ddass ont été intégrées dans les Agences régionales de santé (ARS), sous le nom de Délégation territoriale de l'ARS.

1. Introduction

Les infections invasives à méningocoque (IIM) sont des infections transmissibles graves, potentiellement mortelles, qui affectent le plus souvent des personnes jeunes. En 2005, l'incidence des IIM déclarées était de 1,22 pour 100 000 habitants en France métropolitaine [1]. Les sérogroupe B et C étaient les sérogroupe les plus fréquemment retrouvés dans respectivement 66 % et 25 % des cas. Les sérogroupe W135 et Y représentaient respectivement 5 % et 3 % des cas. D'autres sérogroupe (A, 29E...) étaient impliqués dans 3 % des cas.

La surveillance des IIM est prioritaire en raison de la gravité de la maladie pour l'individu et la collectivité du fait du risque épidémique. En France, cette maladie est à déclaration obligatoire (DO). Tout cas doit être signalé à la Direction départementale des affaires sanitaires et sociales (Ddass) du département de résidence du patient qui organise la prévention des cas secondaires. Les Ddass transmettent les signalements et les fiches de DO à l'Institut de veille sanitaire (InVS). Une analyse hebdomadaire des signalements est effectuée afin de détecter des cas groupés ou des augmentations d'incidence. Une synthèse des cas déclarés est publiée chaque année dans le Bulletin épidémiologique hebdomadaire.

Le Centre national de référence (CNR) des méningocoques reçoit les souches de méningocoques de l'ensemble des laboratoires de bactériologie en France pour confirmation du sérogroupe et de l'antibiogramme ainsi que pour le typage phénotypique (sérotypage et sous-typage) et moléculaire (MLST). Les résultats de ces analyses sont transmis chaque mois à l'InVS.

L'InVS coordonne également une surveillance des méningites et septicémies bactériennes à travers le réseau Epibac (www.invs.sante.fr). Les laboratoires de bactériologie volontaires contribuant à cette surveillance sentinelle transmettent à l'InVS des données sur les souches de méningocoque une fois par an.

Depuis 1996, on observe une augmentation constante du nombre de cas déclarés par la déclaration obligatoire liée en partie à l'amélioration de l'exhaustivité de la DO [2]. Plusieurs études sur l'exhaustivité de la surveillance des IIM ont été menées depuis une vingtaine d'années par des études de capture-recapture à deux ou trois sources (DO, CNR, Epibac). La dernière étude à trois sources a été réalisée sur les données collectées pour l'année 2000. Elle estimait l'exhaustivité de la DO à 73 % [IC 95 % = 69-77] [2].

2. Objectifs

L'objectif principal de cette étude était d'estimer le nombre total de cas d'IIM (méningites et/ou septicémies) confirmées par isolement de méningocoque (culture positive) survenus en France en 2005.

Les objectifs secondaires étaient :

- estimer le taux d'exhaustivité de la déclaration obligatoire des IIM en 2005 ;
- estimer les taux d'exhaustivité de la surveillance des IIM par le CNR et le réseau Epibac en 2005 ;
- identifier les facteurs influençant l'exhaustivité de la DO et estimer des taux d'exhaustivité des systèmes de surveillance pour des variables déterminant l'épidémiologie de la maladie (séro groupe, âge, localisation géographique) ;
- comparer les taux d'exhaustivité de la DO estimés à partir d'une analyse portant sur l'ensemble des cas recensés par au moins une des trois sources à une analyse restreinte aux cas pour lesquels le diagnostic a été confirmé par un laboratoire participant au réseau Epibac.

3. Matériels et méthodes

3.1 DESCRIPTION DES SYSTÈMES DE SURVEILLANCE DES INFECTIONS INVASIVES À MÉNINGOCOQUE (IIM) EN FRANCE

3.1.1 Déclaration obligatoire (DO)

La surveillance épidémiologique des IIM est nationale, elle intéresse la France métropolitaine et les départements d'outre-mer (DOM). Chaque cas diagnostiqué doit être signalé à l'InVS par les Ddass, soit du lieu de résidence du cas, soit du lieu d'hospitalisation. Une fiche standardisée est remplie par le médecin qui a pris en charge le cas et est adressée au médecin inspecteur de santé publique de la Ddass. Ce dernier vérifie les critères de déclaration, complète les éventuelles données manquantes et transmet la fiche à l'InVS.

Les critères de notification d'un cas d'IIM en 2005 sont les suivants :

- isolement d'une souche de méningocoques dans un site normalement stérile (sang, liquide céphalo-rachidien (LCR), liquide articulaire, pleural, péricardique) ou à partir d'une lésion cutanée purpurique ;
- présence de diplocoque à gram négatif à l'examen microscopique du LCR ;
- LCR évocateur de méningite bactérienne purulente (à l'exclusion de l'isolement d'une autre bactérie) et soit :
 - présence d'éléments purpuriques cutanés quel que soit leur type,
 - présence d'antigènes solubles méningococciques dans le LCR, le sang ou les urines,
 - PCR positive à partir du LCR ou du sérum ;
- présence d'un *purpura fulminans*.

3.1.2 Centre national de référence

Le Centre national de référence des méningocoques (CNR), situé à l'Institut Pasteur de Paris, travaille en étroite collaboration avec environ 600 laboratoires de bactériologie hospitaliers qui envoient les souches de méningocoques isolées d'infections invasives (sang, LCR, liquide articulaire, pleural, péricardique), de prélèvements pharyngés, pulmonaires ou gynécologiques. Le CNR reçoit également des échantillons cliniques pour des diagnostics par Polymerase Chain Reaction (PCR), ou des extraits d'ADN.

Le CNR confirme ou identifie les sérogroupes, teste la sensibilité des souches aux antibiotiques, réalise le typage phénotypique (sérotypage et sous-type) et/ou génétique (séquençage *PorA*, MLST) pour les laboratoires demandeurs et assure un suivi de l'épidémiologie moléculaire des méningocoques en France.

Les laboratoires de bactériologie envoient les souches responsables d'infections invasives au CNR, à la demande de la Direction générale de la santé (DGS)¹.

3.1.3 Réseau de laboratoires hospitaliers : Epibac

Le réseau Epibac repose sur des laboratoires hospitaliers de microbiologie couvrant les services de médecine prenant en charge les infections invasives bactériennes (*Haemophilus influenzae*, *Listeria monocytogenes*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Streptococcus agalactiae* (B) et *Streptococcus pyogenes* (A)).

Un cas d'IIM est défini par la présence d'une souche de méningocoque dans le sang ou le LCR.

Pour chaque laboratoire, un recueil mensuel est réalisé sur une fiche qui est adressée à l'InVS. Pour chaque isolement sont recueillis : le site et la date de prélèvement, l'âge et le sexe du patient, le sérotype si approprié, les nom et adresse du laboratoire de microbiologie.

L'objectif de ce réseau est d'estimer l'incidence de ces infections en France, de suivre leur évolution dans le temps et de décrire les principales caractéristiques épidémiologiques des patients hospitalisés.

Le réseau Epibac est coordonné par l'InVS. En 2005, 308 laboratoires volontaires ont participé. La proportion d'admissions en médecine des établissements dont la microbiologie est assurée par les laboratoires Epibac représentait 79% du nombre total d'admissions en médecine des établissements hospitaliers susceptibles de prendre en charge les pathologies étudiées au niveau national. Le réseau Epibac couvrait de manière satisfaisante l'ensemble des régions de France métropolitaine.

Les laboratoires du réseau Epibac, comme l'ensemble des laboratoires de bactériologie en France, sont censés envoyer les souches de méningocoques au CNR.

3.2 APPLICATION DE LA MÉTHODE CAPTURE-RECAPTURE AUX INFECTIONS INVASIVES À MÉNINGOCOQUE

La méthode capture-recapture a été utilisée dans un premier temps en zoologie pour estimer la taille d'une espèce animale dans un endroit donné [3]. Elle consiste à attraper de façon aléatoire des animaux, ceux-ci sont alors marqués puis relâchés. L'étape suivante vise à "recapter" un ou plusieurs échantillons d'animaux. La taille de la population est alors estimée à partir du nombre d'animaux capturés plusieurs fois.

L'utilisation de cette méthode en épidémiologie est intéressante pour les études d'incidence [4], de prévalence [5] et pour estimer l'exhaustivité d'un système de surveillance [6-9]. Les principes, conditions d'application et limites de cette méthode ont été décrits dans la littérature [10-13] et sont présentés brièvement en annexe 1.

¹ Circulaire n° DGS/5C/2006/458 du 23 octobre 2006 relative à la prophylaxie des infections invasives à méningocoque.

3.2.1 Définition des cas

Nous avons retenu une définition de cas commune aux trois sources : "tout patient chez lequel une souche de *N. meningitidis* a été isolée dans le sang ou le LCR" (le réseau Epibac ne recensant pas les cas d'IIM diagnostiqués par PCR).

Les cas retenus étaient les patients résidant en France métropolitaine, hospitalisés en 2005. Les dates utilisées pour inclure les cas variaient selon les sources (date d'hospitalisation pour la DO, date de prélèvement pour Epibac, date d'isolement pour le CNR).

DO : les cas hospitalisés entre le 1^{er} janvier et le 31 décembre 2005 ont été inclus. Pour un cas sans date d'hospitalisation (décès à domicile), la date utilisée pour l'analyse a été la date de notification.

Epibac : les cas pour lesquels les méningocoques ont été isolés sur des prélèvements effectués entre le 1^{er} janvier 2005 et le 1^{er} janvier 2006 ont été inclus (pour tenir compte du délai de prélèvement par rapport à la date d'hospitalisation).

CNR : Les cas pour lesquels les méningocoques ont été isolés entre le 1^{er} janvier 2005 et le 3 janvier 2006 ont été inclus (pour tenir compte du délai d'isolement à partir de la date de prélèvement).

Les cas notifiés par le CNR et Epibac pour des dates de prélèvement ou d'isolement début 2005 et début 2006 pour lesquels existaient une DO fin 2004 et début 2006 ont été exclus.

3.2.2 Identification des cas communs

3.2.2.1 Variables disponibles

Les variables utilisées pour identifier les cas communs aux trois sources sont présentées dans le tableau 1. Parmi les 741 cas déclarés par DO, le sérotype n'était connu que chez 635 (85,7%) d'entre eux. De la même façon, le sérotype de seulement 78,4% des 430 cas identifiés par Epibac était connu. La date de naissance n'était pas précisée chez 38,1% des cas identifiés par le CNR (tableau 3).

TABLEAU 1

Variables utilisées pour mettre en évidence les cas communs entre les trois sources et taux de complétude, 2005

	DO (n=741)		Epibac (n=430)		CNR (n=496)	
	N	%	N	%	N	%
Date de naissance	741	100	420	98	307	62
Âge	741	100	430	100	496	100
Sexe	739	100	429	100	496	100
Département du lieu de résidence/laboratoire ^a	741	100	430	100	496	100
Date d'hospitalisation	734	99	-	-	-	-
Date de prélèvement	-	-	430	100	-	-
Date d'isolement	-	-	-	-	496	100
Sérotype	635	86	337	78	496	100
Site d'isolement	589	79	430	100	484	98

^a Le département de résidence est disponible dans la DO, celui du laboratoire dans les données du CNR et d'Epibac.

3.2.2.2 Création d'un identifiant

Trois variables ont été créées afin d'identifier les cas communs.

Création de la variable n° 1 par concaténation des variables suivantes :

- DO : date de naissance/département de résidence/sexe/sérotype/date d'hospitalisation ;
- CNR : date de naissance/département du laboratoire/sexe/sérotype/date d'isolement ;
- Epibac : date de naissance/département du laboratoire/sexe/sérotype/date de prélèvement.

Création de la variable n° 2 par concaténation des variables suivantes :

- DO : âge/département de résidence/sexe/sérotype/date d'hospitalisation ;
- CNR : âge/département du laboratoire/sexe/sérotype/date d'isolement ;
- Epibac : âge/département du laboratoire/sexe/sérotype/date de prélèvement.

Création de la variable n° 3 par concaténation des variables suivantes :

- DO : date de naissance/département de résidence/sexe/sérotype ;
- CNR : date de naissance/département du laboratoire/sexe/sérotype ;
- Epibac : date de naissance/département du laboratoire/sexe/sérotype.

3.2.2.3 Recherche des doublons

Au sein de chacune des trois sources, les doublons (cas enregistrés plusieurs fois dans la même source) ont été recherchés à l'aide des trois variables créées puis ont été retirés de la base d'analyse.

3.2.2.4 Stratégie de recherche des cas communs

- La variable n° 2 utilisait l'âge pour identifier les cas communs. L'âge a été recalculé lorsque les dates de naissance étaient disponibles (des différences avec l'âge rapporté en années ont été observées chez 30 à 45% des sujets mais la différence était minime, ne dépassant pas un an dans la majorité des cas).

- La variable n° 3 était composée des mêmes éléments que la variable n° 1 mais avec la date de l'évènement (hospitalisation, isolement, ou prélèvement) en moins. Les cas communs identifiés ont été vérifiés, lorsque c'était possible, par rapport à la date d'hospitalisation (source DO), de prélèvement (source Epibac) ou d'isolement (source CNR).
- À l'issue de la recherche effectuée avec les 3 variables créées, les cas communs ont été identifiés manuellement parmi les cas restants.
- Les cas étaient considérés comme étant communs à deux sources si les données concordait et si, au maximum, une variable différait d'une source à l'autre, cette variable pouvait être :
 - la date de naissance, la différence devait être compatible avec une erreur de saisie,
 - l'âge (si dates de naissance non disponibles), qui ne pouvait différer que d'un an maximum,
 - le département, mais il devait rester limitrophe (le département de résidence et le département du laboratoire pouvaient être différents),
 - le sexe.
- En revanche, le sérotype ne devait pas différer mais pouvait être manquant dans une ou deux sources (DO ou Epibac). Si, malgré la concordance de plusieurs variables, une distorsion concernait plus d'une variable, les 2 cas n'étaient pas considérés comme étant communs aux deux sources. Certains laboratoires ont été contactés pour des cas suspectés communs mais dont le sérotype différait.
- Les paramètres suivants ont été pris en compte :
 - le taux d'incidence dans le département de survenue (plus ce dernier était bas, plus la probabilité qu'un cas soit considéré à tort comme étant commun aux deux sources nous semblait faible),
 - la taille de l'hôpital : plus l'hôpital était "petit", plus le risque de considérer à tort un cas comme étant commun nous semblait faible,
 - la fréquence du sérotype, plus ce dernier était rare, plus forte était la probabilité qu'il s'agisse d'un même individu.

› Cas communs entre DO et CNR

Un délai de trois jours au maximum a été choisi entre la date d'hospitalisation (DO) et la date d'isolement (CNR) pour retenir les cas communs à ces deux sources.

› Cas communs entre CNR et Epibac

Un délai de trois jours au maximum a été choisi entre la date de prélèvement (Epibac) et la date d'isolement (CNR) pour retenir les cas communs à ces deux sources.

› Cas communs entre DO et Epibac

Entre la date d'hospitalisation (DO) et la date de prélèvement (Epibac), un délai de deux jours maximum a été choisi pour retenir les cas communs à ces deux sources.

Parfois, l'hôpital différait entre le CNR et Epibac. Ceci s'expliquait en raison des accords institutionnels pouvant exister entre des hôpitaux dans le cadre des Groupements hospitaliers universitaires (disponible sur : www.aphp.fr/index.php?module=hospital&action=hopitaux_liste&vue=liste_hopitaux_group_fo).

Un patient peut être transféré ou un prélèvement peut être adressé dans un autre centre hospitalier pour être analysé. Par exemple, les prélèvements de patients hospitalisés sur Paris pouvaient être analysés sur Créteil, département possédant un laboratoire "de référence".

Lors des comparaisons des sources deux à deux, la troisième source servait parfois à vérifier s'il s'agissait d'un cas commun ou non. Si les 2 cas étaient retrouvés dans la troisième source, il s'agissait bien alors de deux cas distincts et non d'un cas commun.

3.2.3 Identification et prise en compte des variables d'hétérogénéité de capture

3.2.3.1 Variables étudiées

Les facteurs pour lesquels une hétérogénéité de capture pouvait être suspectée était l'âge des patients pour la DO (les jeunes enfants ou les enfants scolarisés peuvent être mieux déclarés car le signalement entraîne la mise en œuvre rapide par la Ddass des mesures de chimioprophylaxie auprès des contacts, notamment en collectivités), le sérotype pour le CNR (les souches pour lesquelles le laboratoire hospitalier identifie un sérotype rare peuvent être adressées de façon plus systématique au CNR pour confirmation du sérotype) et la région (l'analyse capture-recapture effectuée en 2000 montrait une meilleure exhaustivité de la DO dans le quart nord-est de la France).

Les strates pour ces trois variables ont été les suivantes :

- l'âge : quatre classes d'âge ont été définies (moins de 4 ans, 4 à 15 ans, 16 à 24 ans et 25 ans et plus). Ce choix repose sur les caractéristiques épidémiologiques des IIM en France et le type de contacts à rechercher autour du cas pour la mise en œuvre des mesures de chimioprophylaxie (circulaire DGS) ;
- la région : le grand nombre de départements et le faible nombre de cas dans certains départements ne permettaient pas une stratification sur le département pour l'analyse. Les lieux de résidence/hospitalisation des cas ont été regroupés selon cinq interrégions correspondants aux indicatifs téléphoniques (région 1=Île-de-France, région 2=quart nord-ouest, région 3=quart nord-est, région 4=quart sud-est, région 5=quart sud-ouest) (figure 1) ;
- les sérotypes : 4 catégories de sérotypes ont été créées : sérotype B, sérotype C, sérotypes rares (W135, Y, A, Z, 29E, non groupable) et sérotype inconnu.

| FIGURE 1 |

Carte des indicatifs téléphoniques en France (cinq régions)



3.2.3.2 Représentativité des sources

La proportion de cas observée dans chaque sous-groupe a été comparée à la proportion attendue (test du χ^2).

3.2.4 Dépendance entre les sources

Une **dépendance entre la DO et le CNR** avait été retrouvée lors des précédentes analyses capture-recapture à trois sources [2]. Lorsqu'un cas d'IIM est déclaré à la Ddass, cette dernière peut contacter le déclarant afin de s'assurer de l'envoi du prélèvement au CNR. Les cas recensés par la DO et par le CNR sont centralisés à l'InVS. Si un cas présent dans l'une des deux sources est absent de la seconde, cette dernière peut être contactée par l'InVS afin d'en connaître la raison.

Une **dépendance entre le CNR et Epibac** avait également été retrouvée. Il est probable que les biologistes appartenant à un réseau de surveillance bactériologique comme Epibac suivent mieux les recommandations consistant à envoyer la souche au CNR qui a pour mission d'identifier et d'évaluer la sensibilité de la souche aux antibiotiques.

En revanche, ces analyses n'avaient pas mis en évidence de **dépendance entre la DO et Epibac**.

3.2.5 Approche capture-recapture : analyse globale et hétérogénéité de capture

Dans un premier temps, une analyse globale de l'ensemble des données réparties dans un tableau de contingence récapitulant le nombre de cas en fonction de leur présence ou non dans chacune des sources a été réalisée. Des modèles de régression log-linéaires présentés en annexe (différant par le type d'interactions entre sources considéré) ont été ajustés aux données observées. Après ajustement, chaque modèle conduit à une estimation du nombre total de cas N . Les intervalles de confiance de N ont été calculés par la méthode basée sur la déviance [14].

Le modèle final retenu pour estimer le nombre total de cas était sélectionné en tenant compte de la déviance G^2 , de l'AIC (critère d'information d'Akaike), du BIC (critère d'information Bayésien) (annexe 1).

Dans un second temps, l'existence d'une hétérogénéité de capture a été étudiée en stratifiant successivement les données en fonction de l'âge, de la région, puis du sérotype.

L'exhaustivité des données collectées par chacune des sources a été estimée avec son IC 95 % en rapportant le nombre de cas observés au nombre total de cas estimé par le modèle sélectionné, ainsi qu'aux bornes supérieure et inférieure de son IC 95 %.

Deux analyses distinctes ont été réalisées :

- la première analyse porte sur l'ensemble des cas identifiés par les trois sources. Dans la mesure où Epibac ne couvre pas l'ensemble des laboratoires du territoire national et que des cas rapportés par la DO et/ou le CNR contribuent à l'estimation de N alors qu'ils n'ont aucune chance d'être déclarés dans Epibac, l'analyse conduit à une sous-estimation de l'exhaustivité d'Epibac (les cas de la DO et du CNR qui sont confirmés par des laboratoires non participants au réseau Epibac ne peuvent être capturés par celui-ci) ;
- la seconde analyse a été restreinte aux hôpitaux dont les laboratoires participent au réseau Epibac afin d'estimer l'exhaustivité du réseau Epibac et de vérifier que les rapports entre les nombres observés dans les sources DO et CNR et le nombre estimé étaient proches de ceux calculés dans la première analyse (le nombre observé de cas communs aux trois sources et les nombres de cas communs à DO et Epibac ou à CNR et Epibac restaient inchangés par rapport à la première analyse. Par contre, les nombres des autres cas observés par la DO et le CNR étaient inférieurs à ceux de la première analyse. Le nombre estimé N était donc inférieur à celui de la 1^{re} analyse).

3.3 OUTILS D'ANALYSE STATISTIQUE

La recherche des cas communs, l'estimation du nombre total de cas avec les sources croisées deux à deux et les tests d'indépendance entre deux sources ont été réalisés avec les logiciels MS Excel® et Access®.

Les analyses ont été réalisées avec le logiciel STATA 9.0® (StataCorp, College Station, TX, USA), en utilisant un module développé par l'Institut Robert Koch à Berlin ("Stata module to perform capture-recapture analysis for three sources with goodness-of-fit based confidence intervals; 2007 [revised in July 2009]. Disponible sur <http://ideas.repec.org/c/boc/bocode/s456859.html>)

4. Résultats

4.1 CAS RETENUS POUR L'ANALYSE

Afin de respecter une définition des cas commune aux trois sources, les cas figurant dans chaque base ont été passés en revue et exclus selon les variables prises en compte dans la définition de cas ("tout patient résidant en France métropolitaine et hospitalisé en 2005 chez lequel une souche de *N. meningitidis* a été isolée dans le sang ou le LCR").

4.1.1 DO

4.1.1.1 Lieu de résidence

Six cas résidant dans les DOM ont été exclus.

4.1.1.2 Site d'isolement

Le réseau Epibac ne surveillant que les cas d'infections invasives d'origine bactérienne (méningites et bactériémies), les cas de la DO pour lesquels aucun germe n'a été mis en évidence dans le sang ou le LCR ont été exclus de l'analyse.

L'exclusion de ces cas de la DO a été effectuée après comparaison entre les différentes sources, afin de récupérer dans les autres sources une information éventuellement manquante sur le site d'isolement dans la DO (cela a concerné 3 cas).

Au total, 154 sujets ont été exclus de la DO :

- 139 cas d'IIM déclarés sur des critères autres que l'isolement bactériologique (67 *purpura fulminans* ou association méningite purulente et purpura cutané et non confirmés biologiquement ; 58 cas confirmés par PCR et 14 cas déclarés devant un examen direct positif du LCR) ;
- 10 cas avec culture positive dans un site autre que le sang ou le LCR : cutané (n=6) ; articulaire (n=3) ; pleural (n=1) ;
- 5 cas communs à la DO et au CNR car le site d'isolement était articulaire.

En 2005, parmi les cas déclarés à l'InVS, 588 cas ont été retenus pour l'analyse.

4.1.2 CNR

4.1.2.1 Lieu de résidence

Au total, 17 cas résidant dans les DOM ont été exclus.

4.1.2.2 Site d'isolement

Les seuls cas exclus du CNR sur cette variable étaient les 5 cas communs avec la DO (site d'isolement articulaire).

4.1.2.3 Date d'hospitalisation

Les cas dont la date d'isolement était située en bordure de période d'étude ont été comparés à ceux de la base DO : 9 cas ont été exclus car ils ont été hospitalisés la première semaine de 2006 et un autre cas a été écarté car il avait été hospitalisé le 30 décembre 2004.

4.1.2.4 Doublons

Deux cas étaient enregistrés deux fois, un des deux enregistrements a été supprimé de la base.

Au total, 481 cas ont été inclus dans l'analyse.

4.1.3 Epibac

4.1.3.1 Doublons

Trois (3) cas étaient enregistrés deux fois, un des deux enregistrements a été supprimé de la base.

4.1.3.2 Participation des laboratoires

Seuls les laboratoires ayant participé pendant les 12 mois de l'année 2005 ont été retenus, trois laboratoires ont participé moins de 12 mois à la surveillance et les 2 cas qu'ils avaient déclaré ont été éliminés.

Au total, 425 cas ont été inclus dans l'analyse.

4.2 CARACTÉRISTIQUES DES CAS DÉCLARÉS PAR SOURCE

4.2.1 Âge

La distribution des cas par groupe d'âge (tableau 2) était similaire dans les trois sources ($p=1,0$). Elle correspondait à l'attendu, compte tenu du profil épidémiologique de la maladie.

| TABLEAU 2 |

Distribution des cas déclarés dans les trois sources, selon le groupe d'âge, France, 2005

Groupe d'âge	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
0-3 ans	199 (31)	192	32	158	33	142	33
4-15 ans	118 (19)	110	19	87	18	72	17
16-25 ans	136 (22)	128	22	104	22	92	22
>25 ans	174 (28)	158	27	132	27	119	28
Total	627 (100)	588	100	481	100	425	100

4.2.2 Séro groupe

La distribution des sérogroupe (tableau 3) différait significativement selon la source ($p < 0,001$).

Toutes les souches analysées par le CNR ont été identifiées en termes de sérogroupe, contrairement aux deux autres sources pour lesquelles des souches sont restées de sérogroupe inconnu.

La proportion de sérogroupe inconnu pour Epibac est faible car les biologistes participant au réseau mettent en général tout en œuvre pour caractériser les souches. Cependant, ils peuvent ne pas disposer des techniques de mise en évidence de certains sérogroupe.

Pour la DO, la proportion de près de 5 % de sérogroupe inconnu peut être liée à la non-disponibilité du résultat au moment du signalement à la Ddass et à un défaut de transmission ultérieur. Si l'on exclut les sujets dont le sérogroupe est inconnu, la distribution des cas ne diffère pas significativement ($p = 0,9$).

4.2.3 Trimestre de diagnostic

La distribution des cas par trimestre d'hospitalisation, prélèvement ou isolement (tableau 4) était comparable dans les trois sources ($p = 0,9$). Cette distribution reflète la saisonnalité de la maladie.

4.2.4 Région d'origine

La distribution des cas par région (tableau 5) était comparable dans les trois sources ($p = 0,8$). Cette distribution laisse suspecter une hétérogénéité de capture ou des départements avec des plus forts taux d'incidence. Sur 9 situations de cas groupés en 2008, 7 ont concerné les régions 2, 3 et 4. Si on rapporte le nombre de cas à la population des régions, les taux d'incidence (pour 100 000) de la DO pour 2008 sont de 0,6; 1,1; 1,1; 1,0 et 0,8 pour les régions 1, 2, 3, 4 et 5 respectivement.

| TABLEAU 3 |

Distribution des cas déclarés dans les trois sources selon le sérogroupe, France, 2005

Sérogroupe	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
B	388 (62)	368	63	310	65	271	64
C	145 (23)	138	23	118	24	106	25
Rares	63 (10)	55	9	53	11	42	10
Inconnu	31 (5)	27	5	0	0	6	1
Total	627 (100)	588	100	481	100	425	100

| TABLEAU 4 |

Distribution des cas déclarés dans les trois sources selon le trimestre de diagnostic, France, 2005

Trimestre	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
1 ^{er} trimestre	237 (38)	223	38	190	40	164	39
2 ^e trimestre	148 (24)	140	24	101	21	103	24
3 ^e trimestre	97 (15)	89	15	74	15	61	14
4 ^e trimestre	145 (23)	136	23	116	24	97	23
Total	627 (100)	588	100	481	100	425	100

| TABLEAU 5 |

Distribution des cas déclarés dans les trois sources selon la région de résidence ou de prise en charge, France, 2005

Région téléphonique	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
Région 1	77 (12)	72	12	61	13	52	12
Région 2	147 (23)	140	24	113	24	82	19
Région 3	168 (27)	156	26	131	27	123	29
Région 4	162 (26)	154	26	122	25	113	27
Région 5	73 (12)	66	11	54	11	55	13
Total	627 (100)	588	100	481	100	425	100

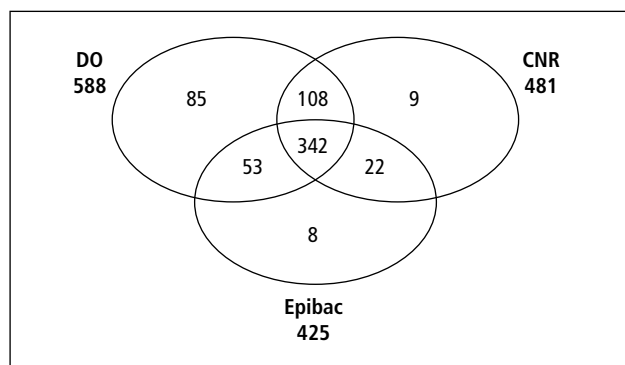
4.3 DISTRIBUTION DES CAS SELON LEUR(S) SOURCE(S) D'APPARTENANCE

En 2005, 627 cas ont été identifiés par les trois sources dont 588 (93,8 %) par la DO.

La distribution des cas selon leur(s) source(s) d'appartenance est présentée à la figure 2.

| FIGURE 2 |

Distribution des 627 cas selon la (les) source(s) de déclaration



83,7 % (525/627) des cas étaient communs à au moins deux sources, 54,5 % (342/627) étaient communs aux trois sources et 16,3 % (102/627) des cas n'ont été déclarés que par une seule source. Parmi ces 102 cas, 85 cas soit 83,3 % ont été déclarés par le système de DO.

4.4 ÉTUDE DE LA DÉPENDANCE ENTRE LES SOURCES

4.4.1 Étude des sources deux à deux (méthode de Wittes)

En croisant les sources 2 à 2, on obtient une estimation du nombre total de cas variant de 562 à 633 (tableau 6). Le nombre de cas estimés était inférieur en croisant les sources CNR et Epibac, suggérant une dépendance positive entre ces deux sources (sous-estimation du nombre de cas par rapport aux deux autres estimations).

4.4.2 Estimation de la dépendance par le rapport de cote (méthode de Wittes)

La recherche de la dépendance entre deux sources peut être effectuée à partir de la troisième source en calculant le rapport de cote. Il existait une dépendance positive entre CNR et Epibac, la dépendance entre DO et CNR était à la limite de la significativité (tableau 7).

| TABLEAU 6 |

Estimation du nombre total de cas en croisant les sources 2 à 2, exhaustivité et intervalle de confiance, France, 2005

Source 1, source 2	n1	n2	n12	n	N _e	IC 95 %	Exh1 (%)	IC 95 %	Exh2 (%)	IC 95 %
DO, CNR	588	481	450	619	629	603-655	94	92-96	76	73-80
DO, Epibac	588	425	395	618	633	587-679	93	91-95	67	63-71
CNR, Epibac	481	425	364	542	562	502-621	86	83-89	76	72-79

n1 : nombre de cas dans la source 1 ; n2 : nombre de cas source 2 ; n12 : nombre de cas communs aux 2 sources ; n : nombre de cas observés dans les deux sources ; N_e : nombre de cas estimés en croisant les deux sources ; Exh1 : taux d'exhaustivité de la source 1 ; Exh2 : taux d'exhaustivité de la source 2.

| TABLEAU 7 |

Rapport de cote de la répartition entre deux sources des cas présents dans la troisième source. Test d'indépendance des sources deux à deux, méthode de Wittes

Sources étudiées	Système dans lequel sont faites les observations	RC	IC 95 %	Indépendance
DO, CNR	Epibac	2,3	0,9-5,9	Pas de dépendance
DO, Epibac	CNR	1,3	0,5-3,1	Pas de dépendance
CNR, Epibac	DO	5,1	3,3-7,8	Dépendance positive

4.4.3 Analyse globale selon les modèles log-linéaires

Les résultats de l'analyse globale (ajustement des huit modèles log-linéaires aux données globales) sont présentés en annexe 2 – tableau 1. Seuls quatre modèles présentent une bonne adéquation

aux données observées : le modèle saturé, le modèle prenant en compte les deux dépendances DO*Epibac et CNR*Epibac, le modèle prenant en compte les deux dépendances DO*CNR et CNR*Epibac, et le modèle prenant en compte uniquement l'interaction entre le CNR et Epibac. Ces modèles conduisent à des estimations du nombre de cas entre 633 et 643.

Le modèle retenu (valeurs AIC et BIC les plus basses) est celui considérant deux dépendances entre les sources, DO*CNR et CNR*Epibac, le nombre total de cas d'IIM estimé était de 639 [IC 95%=630-656]. L'exhaustivité de la DO était de 92,0% [IC 95%=89,6-93,3] (tableau 8).

TABLEAU 8

Exhaustivité des trois sources à partir de l'estimation du nombre total de cas estimé (modèle avec dépendances entre DO et CNR et CNR et Epibac)

N _E =639	Exhaustivité (%)	IC 95 %
DO	92	90-93
CNR	75	73-76
Epibac	66	65-68

4.5 ANALYSE STRATIFIÉE

4.5.1 Sérogroupes

Aucun cas de séro groupe inconnu n'ayant été recensé par le CNR (par définition, le séro groupe des cas pour lesquels la souche était transmise au CNR ne pouvait rester inconnu, ce dernier réalisant le

sérogroupe systématique), l'estimation du nombre d'IIM n'a pas été réalisée dans cette catégorie (figure 1 – annexe 2).

Les modèles ayant conduit aux estimations apparaissent en gras dans les tableaux 2, 3 et 4 en annexe 2. Les trois modèles retenus présentaient une interaction entre les sources CNR et Epibac.

Les trois sources ont identifié 388 cas de séro groupe B, 145 cas de séro groupe C et 63 cas de séro groupes autres et les nombres estimés de cas par les modèles étaient respectivement de 390 [IC 95%=388-395], 145 [IC 95%=145-149] et 63 [IC 95%=63-68] (tableau 9).

L'exhaustivité de la DO était de 94% [IC 95%=93-95] pour le séro groupe B, de 95% [IC 95%=93-95] pour le séro groupe C et de 87% [IC 95%=81-87] pour les autres séro groupes.

L'exhaustivité du CNR était de 79% [IC 95%=79-80] pour le séro groupe B, de 81% [IC 95%=79-81] pour le séro groupe C et de 84% [IC 95%=78-84] pour les autres séro groupes.

L'exhaustivité d'Epibac était de 69% [IC 95%=69-70] pour le séro groupe B, de 73% [IC 95%=71-73] pour le séro groupe C et de 67% [IC 95%=62-67] pour les autres séro groupes.

Les résultats sont en faveur d'une hétérogénéité de capture selon le séro groupe, les cas de séro groupes différents du B et du C étant moins bien recensés par la DO et Epibac.

TABLEAU 9

Exhaustivité des trois sources selon le séro groupe

Trois sources	N _E	DO		CNR		Epibac	
		Exh (%)	IC 95 %	Exh (%)	IC 95 %	Exh (%)	IC 95 %
B	390	94	93-95	79	78-80	69	69-70
C	145	95	92-95	81	79-81	73	71-73
Rares	63	87	81-87	84	78-84	67	62-67

4.5.2 Âge

Les cas âgés de 4 à 15 ans représentaient 18,8% [118/627] des cas alors que ceux âgés de moins de 3 ans représentaient 31,7% (figure 2 – annexe 2).

Les modèles ayant conduit aux estimations apparaissent en gras dans les tableaux 5 à 8 en annexe 2. Les modèles retenus présentaient une interaction entre les sources CNR et Epibac avec en plus pour certains, une interaction entre DO et CNR (0-3 ans et plus de 25 ans).

Pour les cas d'IIM survenant chez des adultes de plus de 25 ans, le tableau 8 en annexe 2 montre qu'aucun modèle ne présente un AIC et BIC de valeurs inférieures à celles observées avec le modèle saturé. L'estimation avec le modèle saturé paraît surestimée. Parmi les autres modèles, le modèle avec une interaction entre les sources DO et CNR et CNR et Epibac avait les valeurs d'AIC et BIC les plus basses (mais positives).

Les trois sources ont identifié 199 cas de 0-3 ans, 120 cas de 4-15 ans, 137 cas de 16-25 ans, 174 cas de plus de 25 ans et les nombres estimés de cas par les modèles étaient respectivement de 199

[IC 95%=199-202], 120 [IC 95%=118-125], 137 [IC 95%=136-141] et 183 [IC 95%=175-202] (tableau 10).

L'exhaustivité de la DO variait entre 96% [IC 95%=94-96] chez les 0-3 ans et 86% [IC 95%=78-90] chez les plus de 25 ans. Dans les autres sources, l'exhaustivité était également supérieure pour les jeunes enfants par rapport aux autres groupes d'âges.

Les résultats sont donc en faveur d'une hétérogénéité de capture en fonction de l'âge des cas.

Compte tenu des interactions prises en compte dans les modèles choisis selon les critères définis dans la méthode, il apparaît que les dépendances entre les sources varient selon les groupes d'âge.

Quand on additionne les nombres estimés d'IIM dans chaque strate, on obtient un total de 639, nombre identique à celui issu de l'estimation par le modèle global. L'exhaustivité globale de la DO serait de 92% (588/639), celle du CNR de 75% (481/639) et celle du réseau Epibac de 66% (425/639).

Dans les données de DO, la proportion de cas liés aux sérogroupes rares était plus élevée chez les adultes de plus de 25 ans, par rapport aux autres groupes d'âge, comme cela est observé habituellement (58 % contre 24 %, $p < 10^{-3}$). Si on effectue l'analyse selon le séro groupe restreinte aux adultes de plus de 25 ans, on a montré que :

- pour les sérogroupes B ou C, les trois sources ont recensé 126 cas (la DO a recensé 119 cas, le CNR 100 cas et Epibac 94 cas). Le modèle présenté en annexe 2 – tableau 8bis, correspond au modèle avec une interaction entre les sources CNR et Epibac. Le nombre de cas estimé par le modèle était de 126 [IC 95 % = 126-130]. L'exhaustivité de la DO était de 94 % [IC 95 % = 91-94], celle du CNR de 79 % [IC 95 % = 77-79] et celle d'Epibac de 75 % [IC 95 % = 72-75];
- pour les sérogroupes rares, les trois sources ont recensé 37 cas (la DO a recensé 31 cas, le CNR 32 cas et Epibac 22 cas). Le modèle présenté en annexe 2 – tableau 8ter, correspond au modèle avec une interaction entre les sources CNR et Epibac. Le nombre de cas estimé par le modèle était de 37 [IC 95 % = 37-41]. L'exhaustivité de la DO était de 83 % [IC 95 % = 76-84], celle du CNR de 86 % [IC 95 % = 78-87] et celle d'Epibac de 59 % [IC 95 % = 54-60].

4.5.3 Régions

Plus de la moitié des cas (53 %) des cas ont été signalés dans les régions 3 et 4 (figure 3 – annexe 2).

Les modèles ayant conduit aux estimations apparaissent en gras dans les tableaux 9 à 13 en annexe 2. Les modèles retenus présentaient tous une interaction entre les sources CNR et Epibac avec, en plus pour certains, une interaction entre DO et Epibac (région 1) ou DO et CNR (région 3).

Les trois sources ont identifié 77 cas dans la région 1 (Île-de-France), 147 cas dans la région 2 (quart nord-ouest), 168 dans la région 3 (quart nord-est), 162 dans la région 4 (quart sud-est) et 73 dans la région 5 (quart sud-ouest). Comme pour les deux autres variables de stratification, les nombres estimés de cas par les modèles étaient égaux ou très proches des nombres observés et étaient respectivement de 78 [IC 95 % = 77-85], de 148 [IC 95 % = 147-153], de 172 [IC 95 % = 168-186], de 163 [IC 95 % = 162-167] et de 73 [IC 95 % = 73-78] (tableau 11).

Pour la DO, les niveaux d'exhaustivité variaient entre 90 et 95 % selon les régions, sans différence significative. Pour le CNR, les exhaustivités étaient également comparables d'une région à l'autre. Il apparaît que pour Epibac, l'exhaustivité de la région 2 (quart nord-ouest) était inférieure aux autres régions.

Si l'on se réfère aux interactions prises en compte dans les modèles choisis selon les critères définis dans la méthode, il apparaît que les dépendances entre les sources varient selon les régions.

Quand on somme les nombres estimés d'IIM dans chaque strate, on obtient un total de 634, nombre très proche de l'estimation par le modèle global (639). L'exhaustivité globale de la DO serait de 93 % (588/634), celle du CNR de 76 % (481/634), et celle du réseau Epibac de 67 % (425/634).

4.6 ÉVOLUTION DE L'EXHAUSTIVITÉ SELON LES SOURCES [2,8]

L'exhaustivité de la DO a augmenté progressivement depuis 15 ans de 53 % à 92 % et de manière significative entre 1991-1992 et 1996, entre 1996 et 2000 et entre 2000 et 2005. L'exhaustivité du CNR est stable depuis 1996 et celle d'Epibac a également augmenté progressivement entre 1996 et 2005.

| TABLEAU 10 |

Exhaustivité des trois sources selon l'âge

Trois sources	N _E	DO		CNR		Epibac	
		Exh (%)	IC 95 %	Exh (%)	IC 95 %	Exh (%)	IC 95 %
0-3 ans	199	96	95-96	79	78-79	71	70-71
4-15 ans	120	91	87-92	72	69-73	60	58-61
16-25 ans	137	93	91-94	76	74-77	67	65-68
>25 ans	183	86	78-90	72	65-75	65	59-68
Total	639	92	-	75	-	66	-

| TABLEAU 11 |

Exhaustivité des trois sources selon la région de résidence

Trois sources	N _E	DO		CNR		Epibac	
		Exh (%)	IC 95 %	Exh (%)	IC 95 %	Exh (%)	IC 95 %
Région 1	78	92	85-94	78	72-79	67	61-67
Région 2	148	95	91-95	76	74-77	55	54-56
Région 3	172	91	84-93	76	70-78	71	66-73
Région 4	163	94	92-95	75	73-75	69	68-70
Région 5	73	90	85-90	74	69-74	75	70-75
Total	634	93	-	75	-	67	-

Évolution de l'exhaustivité des sources DO, CNR et Epibac – 1989-2005

Années	DO		CNR		Epibac	
	Exh (%)	IC 95 %	Exh (%)	IC 95 %	Exh (%)	IC 95 %
1989-1990	53	50-56	55	52-58	-	-
1991-1992	56	54-58	63	61-66	-	-
1996	62	61-63	72	70-73	50	49-51
2000	73	69-77	74	70-78	57	53-59
2005	92	87-93	75	73-76	66	65-68

4.7 REPRÉSENTATIVITÉ DES SOURCES**4.7.1 Sérogroupes**

Dans l'analyse stratifiée, les sérogroupes rares étaient sous-représentés pour la DO. Cependant, la répartition des cas estimés selon le séro groupe (N_e), quand celui-ci est connu, ne diffère pas statistiquement (test du chi2) de celle des cas observés dans la DO, le CNR et Epibac ($p=0,91$; $0,95$ et $0,91$ respectivement) (tableau 13).

4.7.2 Âge

L'analyse stratifiée a montré que, pour la DO, les cas âgés de 25 ans et plus étaient sous-représentés. Cependant, la répartition des cas estimés selon les groupes d'âge (N_e) ne diffère pas statistiquement (test du chi2) de celle des cas observés dans la DO, le CNR et Epibac ($p=0,89$; $0,92$ et $0,11$ respectivement) (tableau 14).

4.7.3 Région

La répartition des cas estimés selon les régions de résidence (N_e) ne diffère pas statistiquement (test du chi2) de celle des cas observés dans la DO, le CNR et Epibac ($p=0,99$; $0,99$ et $0,60$ respectivement) (tableau 15).

4.8 ÉVALUATION DES TAUX D'INCIDENCE À PARTIR DES EFFECTIFS ESTIMÉS (TABLEAU 16)

Le nombre total estimé en 2005 d'IIM confirmés par culture sur le LCR et/ou le sang était de 639 [IC 95 % : 630-656] ce qui, rapporté à l'effectif de la population (source : Institut national de la statistique et

des études économiques, correspondait à un taux d'incidence annuel de 1,0 (1,0-1,1) pour 100 000 habitants sur le territoire métropolitain.

Les calculs présentés dans le tableau 16 montrent que l'incidence estimée était de 0,6/100 000 pour les IIM B, 0,2/100 000 pour les IIM C et de 0,1/100 000 pour les IIM de sérogroupes autres.

L'incidence des IIM est la plus élevée chez les 0-3 ans (6,6/100 000) et la plus faible chez les plus de 25 ans (0,4/100 000).

L'incidence des IIM est la plus faible en Île-de-France (0,7/100 000) et dans le quart sud-ouest (0,8/100 000).

4.9 ESTIMATION DE L'EXHAUSTIVITÉ DE LA DO PAR UNE ANALYSE CAPTURE-RECAPTURE RESTREINTE AUX LABORATOIRES PARTICIPANTS À ÉPIBAC

Les résultats de cette analyse sont présentés en annexes 3 et 3bis.

Le nombre total de cas d'IIM, estimé par l'analyse globale selon les modèles log-linéaires, était de 519 [IC 95 % = 512-531]. L'exhaustivité de la DO était de 92 % [IC 95 % = 90-93], celle du CNR de 76 % [IC 95 % = 75-77] et celle d'Epibac de 82 % [IC 95 % = 80-83].

Les niveaux d'exhaustivité de la DO pour les sérogroupes, les groupes d'âges et les régions obtenus par les analyses stratifiées étaient comparables à ceux estimés dans l'analyse globale même si pour certaines strates ("sérogroupes rares", "plus de 25 ans", "régions 1 et 4"), ils étaient légèrement supérieurs.

Distribution des cas observés selon le séro groupe dans les trois sources et des cas estimés avec les modèles retenus dans l'analyse stratifiée

Séro groupe	DO observés	%	CNR observés	%	Epibac observés	%	N_e	%
B	368	65	310	64	271	65	390	65
C	138	25	118	25	106	25	145	24
Rares	55	10	53	11	42	10	63	11
Total	561	100	481	100	419	100	598	100

| TABLEAU 14 |

Distribution des cas observés selon l'âge dans les trois sources et des cas estimés avec les modèles retenus dans l'analyse stratifiée

Âge	DO observés	%	CNR observés	%	Epibac observés	%	N _E	%
0-3 ans	192	32	158	33	142	33	199	31
4-15 ans	110	19	87	18	72	17	120	19
16-25 ans	128	22	104	22	92	22	136	21
>25 ans	158	27	132	27	119	28	183	29
Total	588	100	481	100	425	100	638	100

| TABLEAU 15 |

Distribution des cas observés selon la région dans les trois sources et des cas estimés avec les modèles retenus dans l'analyse stratifiée

Région	DO observés	%	CNR observés	%	Epibac observés	%	N est	%
Région 1	72	12	61	13	52	12	78	12
Région 2	140	24	113	24	82	19	148	23
Région 3	156	27	131	27	123	29	172	27
Région 4	154	26	122	25	113	27	163	26
Région 5	66	11	54	11	55	13	73	12
Total	588	100	481	100	425	100	634	100

| TABLEAU 16 |

Taux d'incidence estimés [IC 95 %] par séro groupe, tranche d'âge et par région

	Effectif population	N _E	IC 95 %	Taux d'incidence estimé (/100 000)	IC 95 %
Sérogroupe					
B	61 771 000	390	388-395	0,6	0,6-0,6
C	61 771 000	145	145-149	0,2	0,2-0,2
Rares	61 771 000	63	63-68	0,1	0,1-0,1
Groupes d'âges					
0-3 ans	3 037 446	199	199-203	6,6	6,6-6,7
4-15 ans	9 040 245	120	118-125	1,3	1,3-1,4
16-25 ans	7 965 717	136	136-140	1,7	1,7-1,8
>25 ans	41 727 592	183	175-202	0,4	0,4-0,5
Régions					
Île-de-France	11 491 000	78	77-85	0,7	0,7-0,7
Quart nord-ouest	12 272 000	148	147-153	1,2	1,2-1,2
Quart nord-est	14 194 000	172	168-186	1,2	1,2-1,3
Quart sud-est	14 385 000	163	162-167	1,1	1,1-1,2
Quart sud-ouest	8 825 000	73	73-78	0,8	0,8-0,9

5. Discussion

5.1 RÉSULTATS

L'analyse effectuée sur l'ensemble des 627 observations a estimé le nombre total d'IIM confirmées par culture dans le sang ou le LCR en France en 2005 à 639 cas [IC 95 % = 630-656].

L'exhaustivité de la déclaration obligatoire (DO) des IIM confirmées par isolement d'un méningocoque dans le sang ou le LCR a été estimée à 92,0 % [IC 95 % = 89,6-93,3], celle par le CNR à 75,3 % [IC 95 % = 73,3-76,3] et celle par le réseau de laboratoires Epibac à 66,5 % [IC 95 % = 64,8-67,5].

Dans l'analyse restreinte aux laboratoires Epibac, l'exhaustivité de la DO a été estimée à 92,1 % [IC 95 % = 90,0-93,4] et celle de la surveillance CNR à 76,3 % [IC 95 % = 74,6-77,3], ces chiffres étant très proches de ceux correspondant à l'estimation issue de l'analyse portant sur toutes les observations. L'exhaustivité du réseau Epibac a été estimée à 81,9 % (80,0-83,0) et comme attendu, ces chiffres sont supérieurs à ceux issus de la première analyse.

Comparées aux estimations faites par deux précédentes analyses capture-recapture à trois sources, nos résultats sont en faveur d'une amélioration notable de l'exhaustivité qui était pour la DO de 62 % en 1996 et 73 % en 2000 [2]. Cependant, l'évolution de l'exhaustivité du CNR est plus modérée (72 % en 1996 et 74 % en 2000). L'analyse restreinte aux laboratoires Epibac montre une amélioration de l'exhaustivité du réseau qui était également de 74 % en 2000 [2].

La progression de l'exhaustivité de la DO entre 2000 et 2005 pourrait être liée à la médiatisation suite à des campagnes de vaccinations contre le méningocoque C conduites en 2000 dans le Puy-de-Dôme et en 2002, dans trois départements du Sud-Ouest [15,16]. Elle est sans doute également liée à la campagne d'information de l'InVS en 2003 sur la déclaration obligatoire, à une meilleure adhésion des cliniciens et biologistes aux actions de santé publique grâce aux actions de communication des Ddass et des Cire (bulletins d'informations sur les MDO édités par certaines cellules de l'InVS en région (Cire) chaque année) et à une plus grande sensibilisation aux recommandations de prophylaxie des contacts mises en œuvre par les Ddass, notamment quand le cas fréquente une collectivité.

Cependant, concernant une maladie pour laquelle des mesures de prévention pour l'entourage doivent être prises d'urgence par les Ddass quand notamment le cas fréquente une collectivité, l'objectif reste d'atteindre une exhaustivité proche de 100 %.

Le nombre estimé de 639 IIM confirmées par culture dans le sang ou le LCR en 2005 dans cette analyse est très proche de celui estimé à 638 par la même méthode en 2000. Ce résultat est donc en faveur d'une comparabilité de l'incidence réelle des IIM, à définition constante, même si entre les deux années a été observé un pic de cas déclarés (augmentation en 2002 et 2003 suivi d'une baisse en 2004) [17].

En 2002, il y a eu un changement des critères de déclaration des IIM conduisant à un changement de définition des cas. Le principal changement était la prise en compte de la notion de *Purpura fulminans*

sans confirmation biologique mais également le diagnostic par PCR fait devant un LCR évocateur de méningite bactérienne.

Le nombre de cas déclarés en 2005 qui répondaient à l'ancienne définition était de 588 (soit 79 % du total des IIM déclarées). Ceci explique qu'en 2005, 139 cas déclarés par la DO et sans isolement de méningocoque ont été exclus de l'analyse, en plus des cas pour lesquels l'isolement s'était fait en dehors du LCR ou du sang.

En 2006, les critères de déclaration des IIM ont à nouveau évolué afin de prendre en compte des résultats de PCR positifs, la PCR étant dorénavant considérée comme une technique de confirmation biologique d'une infection à méningocoque et de prédiction du sérotype à elle seule.

Les estimations successives par analyses capture-recapture ont permis de tenir compte de la sous-déclaration dans l'estimation de l'incidence réelle des IIM en France et ainsi, de mieux suivre les tendances. Le taux d'exhaustivité est utilisé pour corriger l'incidence nationale à partir du nombre total d'IIM répondant aux critères de DO (et donc depuis 2002, aux cas dont le diagnostic ne repose pas sur la culture). Cette correction repose sur différentes hypothèses : 1) la progression de l'exhaustivité est linéaire entre 2 estimations et 2) le taux d'exhaustivité de la DO est identique pour les cas confirmés par culture et les autres cas, notamment ceux confirmés par PCR.

À ce jour, l'exhaustivité de la DO pour les cas non diagnostiqués par culture n'a pu être estimée car les résultats de PCR n'ont été intégrés dans les transmissions Epibac qu'en 2009. Pour sa part, le CNR transmet à l'InVS les résultats PCR issus de son laboratoire depuis 2008. Même si la recommandation pour les laboratoires de bactériologie qui pratiquent la PCR méningocoque est d'envoyer l'ADN au CNR, cette pratique n'est pas aussi courante qu'elle ne l'est pour les souches. Compte tenu de la dépendance observée entre la DO et le CNR, une analyse de capture-recapture à partir de ces deux sources n'est pas fondée.

Entre 2000 et 2005, le nombre total d'IIM déclarés en France métropolitaine est passé de 480 à 741, soit une augmentation de 54 %. En 2005, l'incidence corrigée pour la sous-notification des IIM en France est donc de 1,3 pour 100 000 habitants en France métropolitaine.

Toujours sous l'hypothèse que l'exhaustivité de la DO calculée pour les IIM confirmées par culture dans le sang ou le LCR peut également s'appliquer au reste des IIM déclarées (définition de cas : paragraphe 3.2), l'augmentation à définition de cas constante de 22 % du nombre de cas déclarés ((588-480)/480), est expliquée essentiellement par l'amélioration de l'exhaustivité sachant que les autres cas (153=741-588) sont liés à la nouvelle définition contribuant à une augmentation de 32 % des cas déclarés (153/480).

Entre 2006 et 2009, le nombre déclaré par la DO a progressivement baissé, passant de 716 à 629 (baisse de 12 %). Cette baisse concerne essentiellement les cas déclarés par la DO avec isolement de méningocoque, leur nombre étant passé de 603 à 456 alors que le

nombre de cas diagnostiqués par PCR a presque doublé sur cette période (66 à 112).

Le niveau élevé de l'exhaustivité des IIM confirmées par culture en 2005 et ce changement observé dans les pratiques diagnostiques depuis cette date doit conduire à un suivi régulier de l'exhaustivité afin d'écartier une possible baisse de l'exhaustivité DO pour cette catégorie de cas depuis 2005.

Les analyses stratifiées sur le sérotype ont montré une hétérogénéité de capture. Les sérotypes autres que B et C (c'est-à-dire Y, W135 ou autres) conduisent au taux d'exhaustivité le plus bas pour la DO et Epibac. Les techniques d'identification de ces sérotypes plus rares peuvent ne pas être disponibles dans certains laboratoires ce qui peut induire un délai plus long pour identifier ces sérotypes et conduire à une non-déclaration (DO). Par contre, les laboratoires souhaitant confirmer ces sérotypes compte tenu de leur plus faible fréquence sont sans doute plus vigilants pour envoyer les souches isolées chez ces cas au CNR.

Les analyses stratifiées sur l'âge ont montré qu'il existait une hétérogénéité de capture pour les trois sources, l'exhaustivité la plus élevée étant estimée chez les moins de 4 ans, tranche d'âge où l'incidence des IIM est la plus élevée. Il semble que la moins bonne exhaustivité de la DO chez les 25 ans et plus soit liée à une moins bonne déclaration des cas de sérotypes rares (W135, Y ou autre) qui sont plus fréquents dans cette tranche d'âge [17].

L'interaction entre DO et Epibac chez les 0-3 ans pourrait s'expliquer par une meilleure implication des biologistes au signalement de ces cas, sachant que des mesures urgentes de prévention doivent être mises en œuvre dans l'entourage et notamment, auprès des fratries des jeunes enfants.

Les analyses stratifiées sur la région ont estimé des taux d'exhaustivité entre 90 et 95 % pour la DO sans vraiment montrer que la région était un facteur d'hétérogénéité de capture. L'exhaustivité de la DO est plus faible dans les régions Nord-Est (région 3) et Sud-Ouest (région 5) alors que c'était dans la région Nord-Est que l'exhaustivité était la plus élevée en 2000. La somme des nombres estimés dans chacune des strates était très proche du nombre estimé par le modèle global.

Les analyses stratifiées suggèrent des hétérogénéités de capture pour l'âge et le sérotype qui pourraient remettre en cause les résultats de l'analyse globale. Cependant, la somme des estimations par strates est très proche de l'estimation obtenue par le modèle global ce qui minimise le risque que ces hétérogénéités soient source de biais dans l'estimation globale.

Les différences dans les probabilités de capture en fonction de l'âge et du sérotype pourraient entraîner, dans certaines sources, un défaut de représentativité des données observées. Les comparaisons des distributions entre les effectifs observés et estimés (paragraphe 4.7) ne sont pas en faveur d'un tel phénomène.

5.2 MÉTHODE

Comme dans les précédentes analyses de capture-recapture à trois sources pour les IIM, l'approche retenue dans ce travail a été celle des modèles log-linéaires [2]. Contrairement aux analyses précédentes qui intégraient des variables d'hétérogénéité dans les modèles log-linéaires,

notre approche a consisté à réaliser des analyses séparées dans des strates prédéfinies à partir de variables considérées pertinentes.

Par ailleurs, les intervalles de confiance des nombres estimés ont été obtenus par la méthode proposée par Regal et Hook [14] qui repose sur une statistique de qualité d'ajustement. Cette méthode présente, par rapport à une méthode basée sur un calcul de variance asymptotique, l'avantage de reposer sur moins d'hypothèses et de garantir des bornes cohérentes avec les observations (en particulier, une borne inférieure toujours supérieure ou égale au nombre total de cas observés). Toutefois, cette méthode ne prend pas en compte l'incertitude liée au choix du modèle et conduit à des intervalles de confiance probablement trop étroits.

À partir de l'ensemble des données transmises à l'InVS par les Ddass pour la DO, par le CNR et les laboratoires du réseau Epibac, il a été possible de croiser les cas d'IIM correspondant à la même définition de cas et issus de trois sources indépendantes de par leur fonctionnement, l'une étant basée principalement sur les cliniciens, les deux autres sur les laboratoires avec pour l'un, un envoi extemporané des souches et pour l'autre, un recueil papier annuel des données. Cependant, lors du transfert de patients d'un hôpital à l'autre, des erreurs ont pu survenir si les deux hôpitaux ne figuraient pas sur la fiche de DO du cas. Dans la mesure du possible, nous en avons tenu compte avant de considérer, comme distincts, des cas présentant des variables communes (date de naissance ou âge, sexe, date de début de l'évènement...).

Une des conditions implicites de l'analyse capture-recapture est que tous les cas identifiés correspondent à des "vrais" cas, selon la définition de cas retenue. Le nombre de cas observés par la DO était systématiquement plus élevé par rapport aux autres sources et au niveau de presque toutes les analyses stratifiées. Une explication pourrait être que pour certaines fiches de DO, l'information sur l'isolement d'un méningocoque dans le LCR ou le sang ait été cochée par excès (confusion avec PCR ou antigènes solubles, par exemple). L'intégration de ces cas, impossibles à identifier (sauf à réinterroger les déclarants) et dont le nombre est inconnu, a pu entraîner une surestimation des cas déclarés par la DO uniquement et donc, une sous-estimation des taux d'exhaustivité. L'inverse peut également être avancé par le fait d'avoir inclus dans l'analyse des cas (3 au total) pour lesquels la notion d'isolement n'apparaissait pas dans la DO mais qu'on avait retrouvé dans les bases CNR et Epibac.

La dépendance observée entre certaines sources n'est pas inattendue dans la mesure où pour le CNR et la DO, il a été demandé par circulaire à l'ensemble des médecins déclarants et aux médecins inspecteurs des Ddass de vérifier que la déclaration d'un cas d'IIM s'accompagne de l'envoi de la souche au CNR. Pour le CNR et le réseau Epibac, il s'agit de la participation volontaire des laboratoires hospitaliers et même si les envois de la souche au CNR et des données à Epibac sont décalés dans le temps, la participation volontaire à la surveillance indique un intérêt pour la santé publique qui pourrait expliquer le lien entre la participation aux 2 modalités de surveillance.

Le niveau élevé de l'exhaustivité des IIM confirmées par culture en 2005 et la place de plus en plus importante de la PCR pour la confirmation biologique des IIM doit conduire à un suivi régulier de l'exhaustivité afin d'identifier une possible baisse de l'exhaustivité DO. Comme lors des analyses précédentes, celle réalisée sur les données 2005 montre qu'il n'existe pas de dépendance entre la DO et le réseau Epibac. Un suivi de l'évolution de l'exhaustivité en croisant ces deux sources est donc adapté et plus facilement reproductible entre la réalisation

de deux analyses de capture-recapture à trois sources. En 2006, ce travail avait conduit à estimer l'exhaustivité de la DO pour les cas confirmés par culture sur sang et LCR à 92,0% [IC 95%=91-94]. En 2009, la surveillance des cas diagnostiqués par PCR a été prise en compte dans le réseau Epibac et permettra une analyse portant à la

fois sur les cas confirmés par culture et par PCR. D'autres sources comme celle du Programme de médicalisation des systèmes d'information (PMSI) ont montré des limites (problèmes de codage), quant à leur utilisation dans des analyses de capture-recapture des IIM [18].

6. Conclusion

Les données de la DO des infections invasives à méningocoque (IIM) ainsi que les données bactériologiques du CNR et du réseau Epibac sont régulièrement analysées par l'InVS pour détecter les situations d'alerte, suivre les tendances et contribuer à l'évaluation des politiques de prévention.

Une estimation de l'exhaustivité de ces dispositifs est très utile afin de pouvoir affiner les analyses et estimer l'incidence réelle des infections. Dans ce but, la méthode capture-recapture permet une analyse fiable et reproductible de l'exhaustivité des systèmes de surveillance pour les IIM confirmées par culture dans le LCR et/ou le sang.

Les résultats de cette analyse conduite sur les données 2005 montrent une nette amélioration de l'exhaustivité de la DO et du réseau Epibac. On peut considérer qu'à cette date, 92 % des IIM confirmées par culture sont rapportées à travers la déclaration obligatoire à l'InVS, 75 % des souches isolés dans le LCR ou le sang sont envoyées au CNR et que

82 % des informations bactériologiques des laboratoires du réseau Epibac sont transmises à l'InVS.

Cependant, la survenue d'une IIM devant conduire sans délai à des mesures de prévention pour protéger l'entourage, l'objectif reste d'atteindre une exhaustivité proche de 100 % du signalement et de la déclaration des cas d'IIM aux Ddass.

Une des limites actuelles est l'utilisation des taux d'exhaustivité calculés à partir des cas confirmés par culture pour corriger les taux d'incidence des IIM répondant à l'ensemble des critères de DO et notamment, la confirmation par PCR. L'adaptation des différents systèmes de surveillance microbiologique (CNR et Epibac) à l'évolution des pratiques, en matière de confirmation biologique par le recueil des données sur les cas confirmés par PCR et non par culture, devrait permettre de lever cette limite dans les analyses futures.

Références bibliographiques

- [1] Parent du Châtelet I, Taha MK. Les infections invasives à méningocoque en France en 2005. *Bull Epidemiol Hebd* 2007;49:383-7.
- [2] Perrocheau A. La surveillance des infections invasives à méningocoques en France en 2000. Évaluation quantitative par la méthode capture-recapture à trois sources. Rapport InVS. Saint-Maurice (Fra): Institut de veille sanitaire, février 2006, 40 p. Disponible sur : www.invs.sante.fr.
- [3] Cormack RM. The statistics of capture-recapture methods. *Oceanogr Mar Biol Ann Rev* 1968;6:455-506.
- [4] Laporte RE, Stull E, McCarty D. Monitoring the incidence of myocardial infarctions: applications of capture-marks-recapture technology. *Intern J Epid* 1992;21:258-63.
- [5] Hook EB, Albright SG, Cross PK. Use of Bernoulli census and log-linear methods for estimating the prevalence of *Spina bifida* of livebirths and the completeness of vital records reports in New York state. *Am J Epidemiol* 1980;112(6):750-8.
- [6] Bernillon P, Lievre L, Pillonel J, Laporte A, Costagliola D, Groupe d'épidémiologie des CISH. Estimation de la sous-déclaration des cas de Sida en France par la méthode capture-recapture. *Bull Epidemiol Hebd* 1997;5:19-21.
- [7] Reintjes R *et al.* Assessing the sensitivity of STD surveillance in the Netherlands: an application of the capture-recapture method. *Epidemiol Infect* 1999;122(1):97-102.
- [8] Hubert B, Desenclos JC. Évaluation de l'exhaustivité et de la représentativité d'un système de surveillance par la méthode capture-recapture. Application à la surveillance des infections à méningocoque en France en 1989 et 1990. *Rev Epidém et Santé Publ* 1993;41:241-9.
- [9] Galloway A, Vaillant V, Bouvet P, Grimont P, Desenclos JC. How many foodborne outbreaks of Salmonella infection occurred in France in 1995? Application of the capture-recapture method to three surveillance systems. *Am J Epidemiol*. 2000;152(2):171-7.
- [10] Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;17:243-64.
- [11] International working group for disease monitoring and forecasting. Capture-recapture and multiple-record systems estimation II: applications in human diseases. *Am J Epidemiol* 1995;142:1059-68.
- [12] International working group for disease monitoring and forecasting. Capture-recapture and multiple-record systems estimation I: history and Theoretical development. *Am J Epidemiol* 1995;142:1047-88.
- [13] Galloway A, Nardone A, Vaillant V, Desenclos JC. La méthode capture-recapture appliquée à l'épidémiologie : principes, limites et applications. *Revue d'épidémiologie et de santé publique*, 2002;(50):219-32.
- [14] Regal RR, Hook EB. Goodness-of-fit based confidence intervals for estimates of the size of a closed population. *Stat Med*. 1984;3:287-91.
- [15] Lévy-Bruhl D, Perrocheau A, Mora M, *et al.* Vaccination campaign following an increase in incidence of serogroup C meningococcal diseases in the department of Puy-de-Dôme (France). *Euro Surveill*. 2002 May;7(5):74-6.
- [16] Charron M et Hemery C. Étude de la couverture vaccinale suite à la campagne de vaccination contre le méningocoque C dans les Landes, les Pyrénées-Atlantiques et les Hautes-Pyrénées en 2002. Rapport InVS. Saint-Maurice (Fra) : Institut de veille sanitaire, août 2004, 52 p. Disponible sur : www.invs.sante.fr.
- [17] Parent du Châtelet I, Taha MK, Lepoutre A, Lévy-Bruhl D. Les infections invasives à méningocoque en France en 2007. *Bull Epidemiol Hebd* 2008;51/52:498-503.
- [18] Dubos F, Maréchal I, Tilmont B, Courouble C, Leclerc F, Martinot A. Incidence of invasive meningococcal diseases in children in northern France: usefulness and limits of the discharge code database for correcting compulsory notification data *Arch Pediatr*. 2009;16(7):984-90.

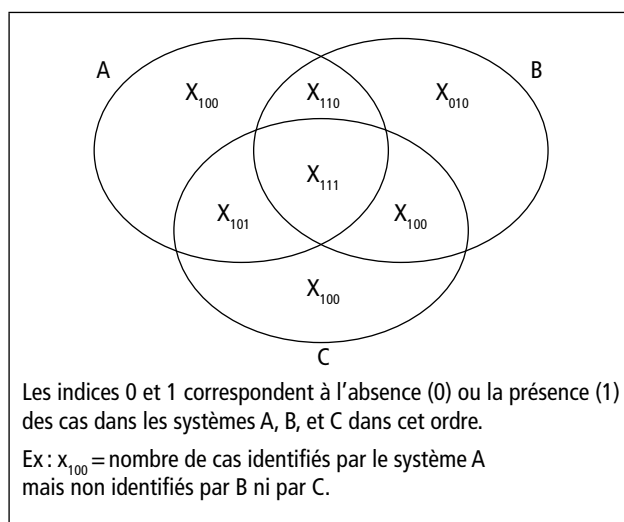
Annexe 1 – Principe et conditions de la méthode capture-recapture

1. PRINCIPE

Cette méthode permet, en croisant les cas d'une maladie recensés par plusieurs systèmes de surveillance dans une population et pendant une période donnée, d'estimer le nombre x_{000} de cas non identifiés par les systèmes de surveillance. Cette méthode nécessite une identification des cas communs aux différents systèmes. L'estimation de x_{000} obtenue permet d'estimer le nombre total de cas de la maladie dans la population d'étude et l'exhaustivité de chaque système (figure 1).

FIGURE 1

Répartition des cas d'une maladie recensés par trois systèmes A, B, C



2. LES CONDITIONS D'APPLICATION DE LA MÉTHODE

2.1 Conditions "implicites"

- **Tous les cas identifiés sont des vrais cas.** Une définition différente des cas selon les sources peut remettre en cause la validité des estimations. La présence de cas définis selon des critères non admis dans les autres sources entraînerait une surestimation du nombre total de cas non déclarés et une sous-estimation de l'exhaustivité des autres sources.
- **Les systèmes de surveillance couvrent les mêmes zones géographiques durant la même période.** Si les zones géographiques couvertes sont disjointes, la probabilité qu'un cas soit commun aux deux systèmes de surveillance est proche de zéro. Le nombre total de cas sera surestimé et l'exhaustivité sera alors sous-estimée.
- **Tous les vrais cas communs et seulement les vrais cas communs sont identifiés.** Une surestimation du nombre de cas communs induit une sous-estimation du nombre total de cas et donc, une surestimation de l'exhaustivité des sources.

2.2 Conditions "statistiques"

- **Les sources sont indépendantes**, c'est-à-dire que la probabilité qu'un individu soit recensé dans une source ne dépend pas de la probabilité qu'il soit recensé par une autre source. Il y a dépendance positive entre deux systèmes lorsque l'identification des cas par un système augmente la probabilité d'être identifiés par l'autre système pour ces cas. La dépendance positive induit une sous-estimation de N. Inversement, il y a dépendance négative lorsque le fait d'être identifié dans un système diminue la probabilité d'être identifié dans l'autre système. Dans ce cas-là, on observera une surestimation de N. Avec plus de deux sources, la dépendance entre les sources peut être évaluée et prise en compte dans l'estimation de N grâce aux modèles log-linéaires.
- **L'homogénéité de capture des cas** : pour une source donnée, tous les individus de la population étudiée ont la même probabilité de capture. La notification des cas dans une source ne doit pas être liée à des variables caractérisant les cas (âge, sexe, lieu de résidence, gravité de la maladie...). Cependant, cette probabilité peut être différente suivant les sources. La présence de variables d'hétérogénéité de capture peut induire une dépendance positive ou négative entre les sources [a]. Les interactions entre les variables d'hétérogénéité et les sources peuvent être prises en compte en stratifiant sur ces variables pour créer des strates de probabilité de capture homogène et réduire les biais pour l'estimation de N [b,c]. Le nombre total de cas est ainsi estimé par la somme des N des sous-groupes. Il n'est pas recommandé de stratifier sur les variables pour lesquelles les probabilités d'identification sont homogènes car la stratification entraîne une augmentation de la variance de N d'où une diminution de la précision de l'estimation.
- **La population étudiée est close**, c'est-à-dire qu'il n'y a pas de mouvement de population pendant la période d'étude. Le non-respect de cette condition peut induire une sous-estimation du nombre de cas communs et donc, une surestimation de N. Cette condition est très liée à celle de l'homogénéité de capture.

3. ESTIMATION DU NOMBRE TOTAL DE CAS AVEC DEUX SOURCES

Le tableau de contingence 2x2 (tableau 1) permet de répartir les cas selon leur présence (indice 1) ou absence (indice 2) dans l'une ou l'autre source.

TABLEAU 1

Tableau de contingence 2x2

		A		
		Oui	Non	
B	Oui	X_{11}	X_{01}	N_2
	Non	X_{10}	$X_{00}=?$	
		N_1		N

N_1 est l'effectif total de la source A ; N_2 est l'effectif total de la source B ; N est le nombre total de cas à estimer ; X_{00} est le nombre de cas identifiés par aucune des sources, à estimer.

Sous l'hypothèse d'indépendance des sources ($P[\text{identifié par A}/\text{identifié par B}] = P[\text{identifié par A}]$), les estimateurs de Sekar et Deming [d] permettent d'estimer le nombre de cas identifiés par aucune des sources X_{22} , le nombre total de cas N , sa variance $\text{Var}(N)$ et son intervalle de confiance à 95 % [$\text{IC}_{95\%}(N)$]:

$$\hat{X}_{00} = \frac{X_{10} X_{01}}{X_{11}}$$

$$\hat{N} = X_{11} + X_{01} + X_{10} + \frac{X_{10} X_{01}}{X_{11}} = \frac{N_1 N_2}{X_{11}}$$

$$\text{Var}(N) = \frac{N_1 N_2 X_{10} X_{01}}{X_{11}^3} \quad \text{IC}_{95\%}(N) = \hat{N} \pm 1,96 \times \sqrt{\text{Var}(N)}$$

Chapman [e] et Seber [f] ont montré que ces estimateurs pouvaient être biaisés lorsque les effectifs sont faibles et donc que X_{11} a une probabilité non nulle d'être égal à 0. Ils ont proposé d'autres estimateurs:

$$\hat{N} = \frac{(N_1+1)(N_2+1)}{X_{11}+1} - 1 \quad \text{Var}(N) = \frac{(N_1+1)(N_2+1)X_{10}X_{01}}{(X_{11}+1)^2(X_{11}+2)}$$

Les taux d'exhaustivité des deux sources A et B sont estimés respectivement par:

$$\text{Exh}_A = \hat{p}_A = \frac{N_1}{N} = \frac{X_{11}}{N_2} \quad \text{Exh}_B = \hat{p}_B = \frac{N_2}{N} = \frac{X_{11}}{N_1}$$

L'intervalle de confiance à 95 % de chaque taux d'exhaustivité est alors calculé en rapportant le nombre de cas déclarés aux bornes de l'intervalle de confiance à 95 % de l'estimation du nombre total de cas.

4. MÉTHODE CAPTURE-RECAPTURE AVEC PLUS DE DEUX SOURCES

4.1 Dépendance entre les sources

Il est possible de suspecter l'existence d'une dépendance entre deux sources. Par exemple, la présence d'interlocuteurs communs au sein de deux sources (un médecin biologiste participant à deux réseaux de surveillance par exemple) est susceptible d'entraîner une dépendance entre ces deux sources.

4.1.1 Évaluation de la dépendance en croisant les sources deux à deux

Pour plus de deux sources, Wittes [g,h] propose d'estimer N pour chaque couple de sources. La dépendance entre les sources est évaluée en comparant les estimations de chacun des couples de sources. Si l'un d'eux donne une estimation très différente des autres couples, on peut suspecter une dépendance entre ces sources.

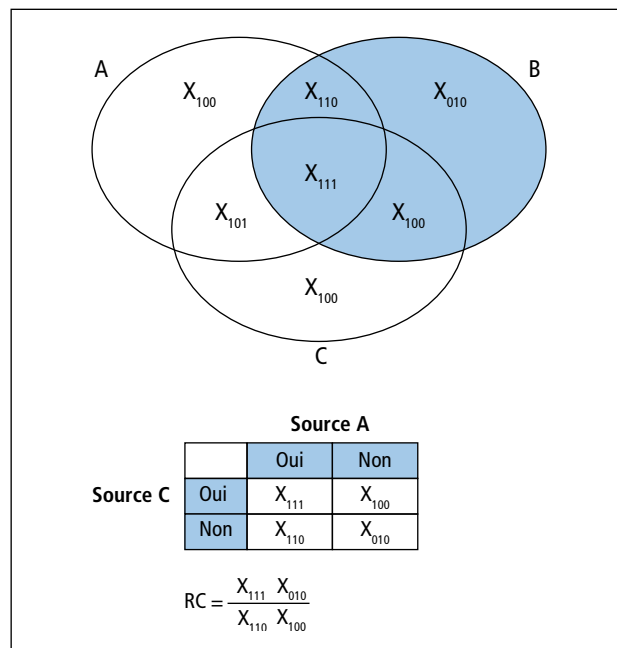
4.1.2 Tests d'indépendances des sources avec plus de deux sources

Pour plus de deux sources, l'indépendance entre deux sources peut également être testée avec un test du χ^2 appliqué sur un tableau de contingence 2×2 à l'intérieur d'une 3^e source (figure 2). En outre, le calcul des rapports de cote (RC) permet d'indiquer le sens et la force des dépendances éventuelles. Si le RC est inférieur à 1, il existe une dépendance négative entre les sources, d'où une surestimation de N (lorsque N est estimé à l'aide d'un modèle qui suppose l'indépendance). Si le RC est supérieur à 1, il existe une dépendance positive entre les sources, d'où une sous-estimation de N .

La figure 2 représente la distribution des cas identifiés par les sources A, B et C. La dépendance entre les sources A et C peut être évaluée en calculant le RC à partir des cas identifiés par la source B. Les cas X_{111} ont été identifiés à la fois par la source A et C, les cas X_{110} et X_{100} ont été respectivement et uniquement identifiés par la source A et par la source C, les cas X_{010} n'ont été identifiés ni par A ni par C.

FIGURE 2

Évaluation de la dépendance par estimation du rapport des cotes (RC)



On peut estimer le nombre total de cas en regroupant les sources dépendantes en une source unique. L'analyse est alors réalisée en considérant deux sources indépendantes, les deux sources fusionnées et la troisième source [i].

4.2 Estimation du nombre total de cas à l'aide des modèles log-linéaires

4.2.1 Principe

Les modèles log-linéaires permettent de calculer des estimations prenant en compte les dépendances entre les sources [b,i].

Le modèle log-linéaire permet d'étudier les relations entre k variables qualitatives croisées dans un tableau de contingence. Il représente le logarithme népérien de la fréquence attendue d'une cellule du tableau comme une combinaison linéaire d'effets principaux et d'interactions [i].

4.2.2 Analyse globale

En absence de variable d'hétérogénéité, le tableau de contingence a une cellule structurellement vide correspondant à l'absence d'identification de cas par les sources. Pour estimer les effectifs attendus, le modèle utilise toutes les cellules du tableau, sauf celle définie comme étant structurellement vide et pour laquelle on attend une estimation. La présence d'une cellule structurellement vide rend impossible l'ajustement d'un modèle prenant en compte l'interaction d'ordre maximum entre toutes les sources [c].

Pour trois sources, les cas se répartissent en fonction de leur présence ou absence dans chacune des sources A, B et C, dans le tableau de contingence suivant :

TABLEAU 2 |

Tableau de contingence à trois sources

	A			
	Oui		Non	
	B	B	B	B
C	Oui	Non	Oui	Non
Oui	x_{111}	x_{101}	x_{011}	x_{001}
Non	x_{110}	x_{100}	x_{010}	$x_{000}?$

1=présence, 0=absence des cas dans les sources A, B, C dans cet ordre.

L'analyse globale consiste à ajuster un modèle log-linéaire aux 7 données observées (en considérant la cellule vide comme une donnée manquante) ; huit modèles peuvent être ajustés aux données :

le modèle supposant l'indépendance entre les trois sources :

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k$$

trois modèles intégrant une interaction entre deux sources :

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{AB}^{ij}$$

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{AC}^{ik}$$

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{BC}^{jk}$$

trois modèles intégrant deux interactions entre deux sources :

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{AB}^{ij} + u_{AC}^{ik}$$

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{BC}^{jk} + u_{AC}^{ik}$$

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{AC}^{ik} + u_{BC}^{jk}$$

le modèle complet (ou saturé) intégrant toutes les interactions d'ordre 2 :

$$\ln m_{ijk} = u + u_A^i + u_B^j + u_C^k + u_{AB}^{ij} + u_{AC}^{ik} + u_{BC}^{jk}$$

où

- m_{ijk} est le nombre de cas attendu pour la cellule ijk ($0 \leq i \leq 1$, $0 \leq j \leq 1$, $0 \leq k \leq 1$) sous le modèle considéré
- u_A , u_B et u_C sont les effets principaux correspondant à la présence dans chaque source
- u_{AB} , u_{AC} et u_{BC} sont les termes d'interaction d'ordre 2 correspondant à la présence dans les deux sources indexées.

Le modèle complet est dit saturé car il comporte le même nombre de paramètres qu'il y a de données à ajuster. Ce modèle prédit parfaitement les observations. Tous les autres modèles sont emboîtés dans celui-ci (un modèle est emboîté dans un autre plus général lorsqu'il est un cas particulier de ce modèle plus général).

4.2.3 Adéquation du modèle aux données

L'adéquation (ou ajustement) d'un modèle M aux données peut être évaluée à l'aide de la statistique de la déviance G_M^2 qui fournit une mesure de l'écart entre les données observées x_{ijk} et les données prédites m_{ijk} par le modèle ajusté :

$$G_M^2 = 2 \sum_{\substack{0 \leq i \leq 1, \\ 0 \leq j \leq 1, \\ 0 \leq k \leq 1}} x_{ijk} \ln(x_{ijk}/m_{ijk})$$

Remarque : la déviance du modèle saturé est toujours nulle puisque les prédictions obtenues, après ajustement de ce modèle, sont égales aux observations.

La déviance suit asymptotiquement une loi du Chi 2 à n-p degrés de liberté (où n est le nombre de données observées dans le tableau de contingence et p le nombre de paramètres du modèle). L'hypothèse nulle d'un test d'adéquation est que le modèle présente une bonne adéquation. En pratique, on ne conservera pas un modèle pour lequel le test d'adéquation est significatif.

4.2.4 Sélection de modèle

En règle générale, un plus grand nombre de paramètres conduit à un meilleur ajustement aux données mais à une moins bonne précision des prédictions. Ainsi, les procédures de sélection de modèle visent à trouver un compromis entre qualité d'ajustement et principe de parcimonie.

► *Comparaison de modèles emboîtés à l'aide de la déviance*
Deux modèles emboîtés peuvent être comparés au moyen de la différence de leur déviance : dans le cas de 2 modèles emboîtés M_1 et M_2 ayant respectivement p_1 et p_2 paramètres ($p_1 > p_2$), la différence $G_{M_1}^2 - G_{M_2}^2$ suit approximativement une loi du Chi2 à $p_1 - p_2$ degrés de liberté.

Lorsque la différence des déviances n'est pas significative, on aura tendance à conserver le modèle le plus parcimonieux.

Remarque : ce test est un test de rapport de vraisemblances.

► *Comparaison de modèles à l'aide de critères d'information*

D'autres critères ont été proposés pour le choix de modèle [j,k]. Ce sont des fonctions de la déviance G^2 et du nombre de degrés de liberté du modèle (ddl = n - p, où n est le nombre de cellules du tableau de contingence pour lesquelles on dispose d'observations et p le nombre de paramètres inclus dans le modèle) : le critère d'information d'Akaike (AIC) et le critère d'information Bayésien (BIC). Draper [1] a proposé une petite altération au BIC, noté ici DIC.

$$AIC = G^2 - 2(\text{ddl})$$

$$BIC = G^2 - (\ln N_{\text{obs}})(\text{ddl})$$

$$DIC = G^2 - (\ln(N_{\text{obs}}/2\pi))(\text{ddl})$$

Quel que soit le critère, on privilégiera un modèle pour lequel la valeur associée est la plus basse. Le modèle saturé ayant un G^2 et un nombre de degré de liberté nul, a également un AIC, un BIC et DIC nuls. Seuls les modèles ayant un AIC et un BIC négatifs seront favorisés par rapport au modèle saturé.

► *Calcul des intervalles de confiance de la taille de la population totale*

Le calcul des intervalles de confiance peut être basé sur un calcul de variance asymptotique et une hypothèse de normalité pour la distribution de la taille de la population [c].

Il est également possible de les calculer en considérant un critère de qualité d'ajustement (la déviance) [m,n] ; cette méthode garantit une borne inférieure de l'intervalle de confiance de la taille de la population totale toujours supérieure ou égale au nombre total de cas observés.

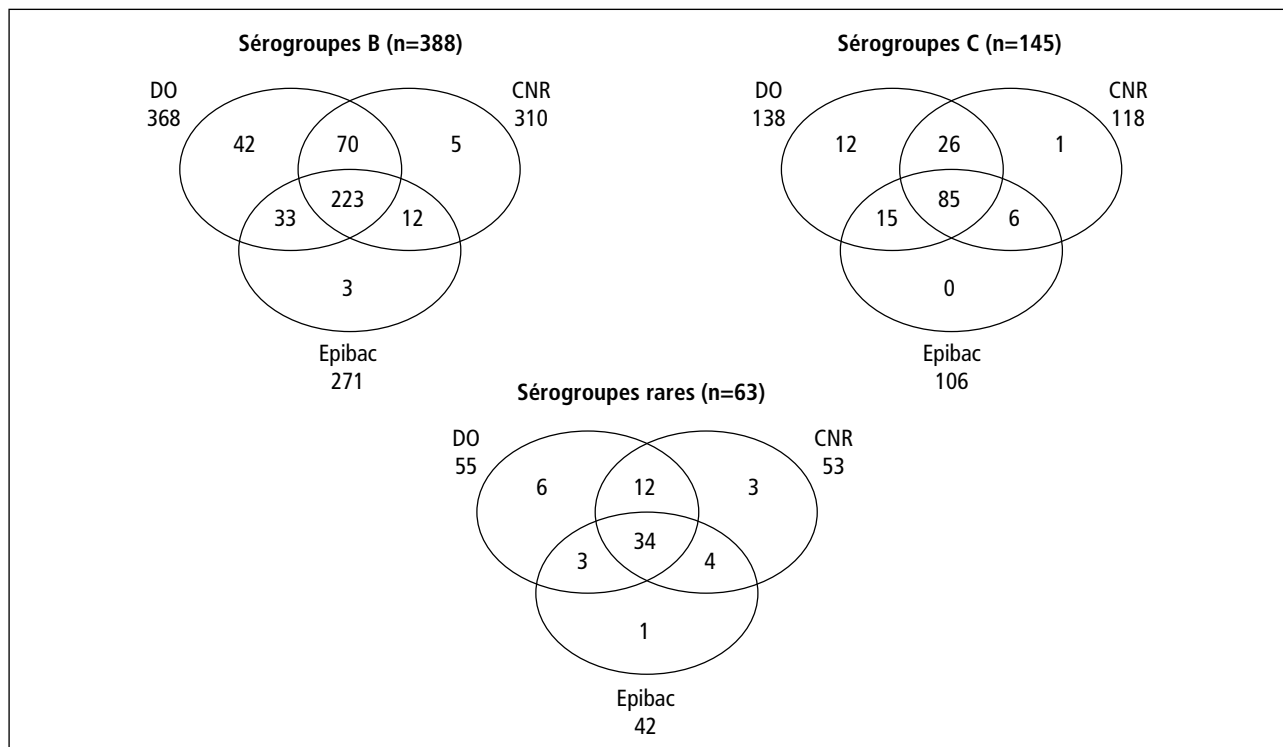
Références bibliographiques

- [a] Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;17:243-64.
- [b] Cormack RM, log-linear models for capture-recapture. *Biometrics* 1989;45:395-413.
- [c] Bishop YMM, Fienberg SE, Holland PW. *Discrete multivariate analysis: theory and practice*. Cambridge: MIT press. 1975. pp 220-56.
- [d] Sekar CC, Deming WE. On a method of estimating birth and death rates and the extent of registration. *American statistical association Journal* 1949;44:101-15.
- [e] Chapman CJ. Some properties of the hypergeometric distribution with its application to zoological censuses. *U California Public Stat* 1951;1:131-60.
- [f] Seber GAF. The effects of trap response on tag recapture estimates. *Biometrics* 1970;26:13-22.
- [g] Wittes J, Sidel VW. A generalization of the simple capture-recapture model with the application to epidemiological research. *J Chron Dis* 1968;21:287-301.
- [h] Wittes JT, Colton T, Sidel VW. Capture-recapture method for assessing the completeness of case ascertainment when using multiple information source. *J Chron Dis* 1974;27:25-36.
- [i] Rice J. Loglinear analysis: analysis of categorical variables in the logit setting. *Advances in social science methodology*. Volume 2. pp. 1-52.
- [j] Hook EB, Regal RR. Validity of methods for model selection. Weighting for model uncertainty and small sample adjustment in capture-recapture estimation. *Am J Epidemiol* 1997;145(12):1138-44.
- [k] Hook EB, Regal RR. Effect in variation of probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. *Am J Epidemiol* 1993;137(10):1148-66.
- [l] Draper D. Assessment and propagation of model uncertainty. *J R Stat Soc [B]* 1995;57:45-70.
- [m] Regal RR, Hook EB. Goodness-of-fit based confidence intervals for estimates of the size of a closed population. *Stat Med*. 1984;3:287-91.
- [n] Cormack RM. Interval estimation for mark-recapture studies of closed populations. *Biometrics* 1992;42:567-76.

Annexe 2 – Illustrations de l'analyse capture-recapture

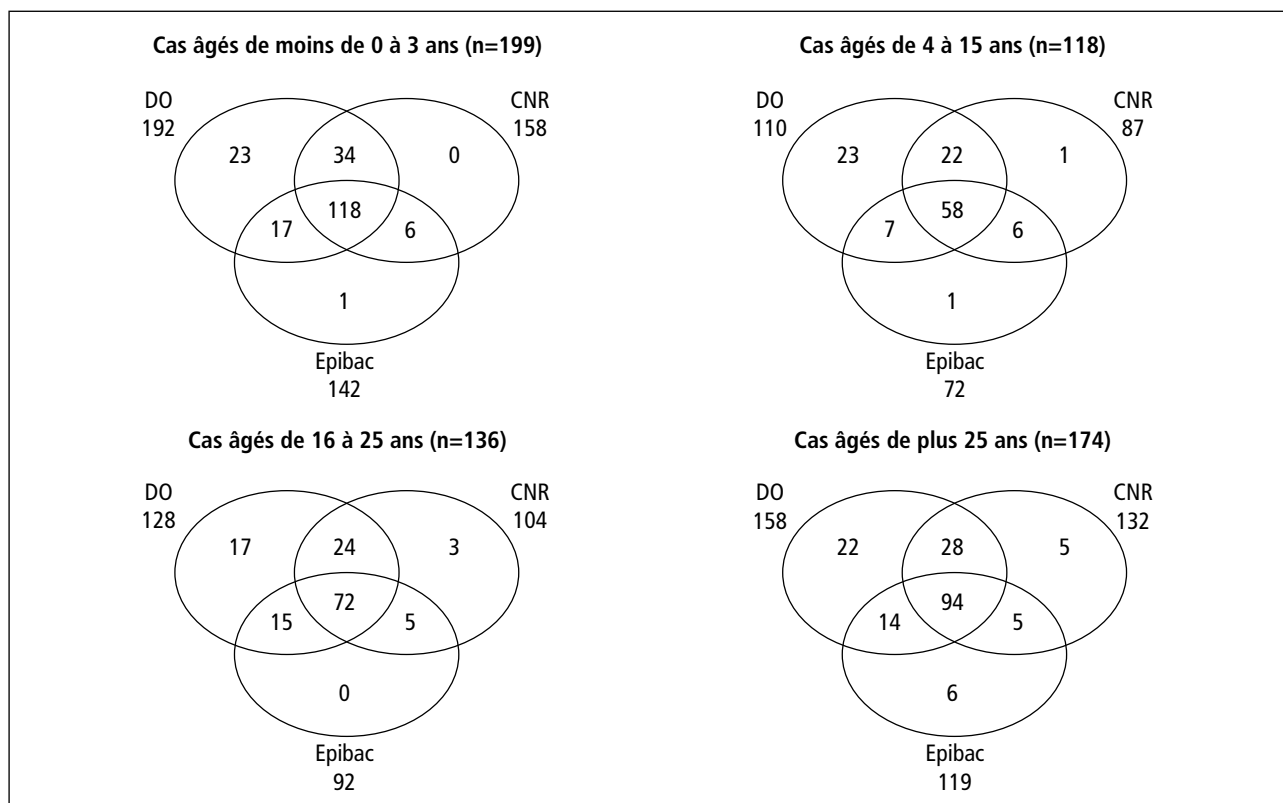
| FIGURE 1 |

Distribution des cas selon le sérotype et les sources de déclaration (n=596)

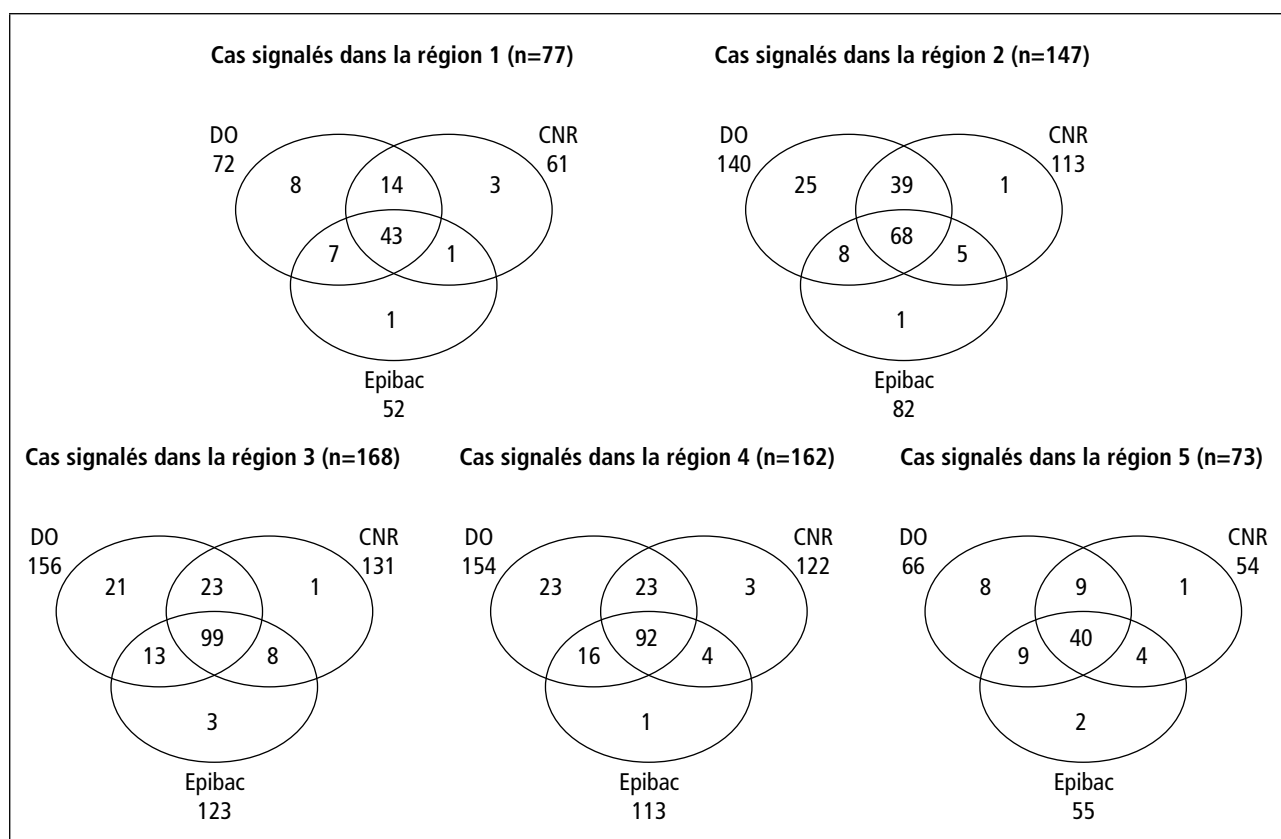


| FIGURE 2 |

Distribution des cas selon l'âge et les sources de déclaration (n=627)



Distribution des cas selon la région et la (les) source(s) de déclaration (n=627)



| TABLEAU 1 |

Résultats de l'analyse globale par modèles log-linéaires : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E, critères utilisés pour la sélection du modèle : déviance (G²), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G ²	p	AIC	BIC	Xest	N _E	IC 95 %
Modèle saturé	0	0,0	1,0	0,0	0,0	16	643	630-678
DE et CE	1	3,4	0,07	1,4	1,4	7	634	628-644
DC et CE	1	0,4	0,5	-1,6	-1,6	12	639	630-656
DC et DE	1	64,5	<10 ⁻⁴	62,5	62,6	3	630	627-638
CE	2	3,4	0,2	-0,6	-0,5	6	633	628-641
DE	2	64,7	<10 ⁻⁴	60,7	60,8	2	629	627-635
DC	2	64,7	<10 ⁻⁴	60,7	60,8	3	630	627-637
Pas d'interaction	3	65,0	<10 ⁻⁴	59,0	59,2	3	630	627-635

D: DO; C: CNR; E: Epibac; AIC=G² - 2×ddl; BIC=G² - (log N_{obs}/2π)×(ddl); Xest: nombre estimé de cas notifiés par aucune des sources; N: Nombre total de cas estimé.

Le modèle retenu est présenté en gras.

| TABLEAU 2 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe B: estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	5	393	388-413
DE et CE	1	0,6	0,46	-1,44	-1,39	3	391	388-398
DC et CE	1	0,3	0,61	-1,74	-1,69	3	391	388-402
DC et DE	1	27,2	$<10^{-4}$	25,19	25,25	1	389	388-395
CE	2	0,7	0,71	-3,31	-3,2	2	390	388-395
DE	2	27,2	$<10^{-4}$	23,19	23,3	1	389	388-393
DC	2	27,2	$<10^{-4}$	23,2	23,31	1	389	388-394
Pas d'interaction	3	27,2	$<10^{-4}$	21,2	21,36	1	389	388-393

Le modèle retenu est présenté en gras.

| TABLEAU 3 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe C: estimations des nombres de cas recensés par aucune des sources Xest et du total, critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	0	145	145-149
DE et CE	1	1,3	0,26	-0,74	-0,69	0	145	145-149
DC et CE	1	0,3	0,56	-1,66	-1,6	0	145	145-149
DC et DE	1	4,5	0,03	2,49	2,55	0	145	145-147
CE	2	1,4	0,48	-2,55	-2,44	0	145	145-149
DE	2	6,4	0,04	2,36	2,46	0	145	145-148
DC	2	5,2	0,08	1,16	1,27	0	145	145-147
Pas d'interaction	3	6,7	0,08	0,74	0,9	0	145	145-148

Le modèle retenu est présenté en gras.

| TABLEAU 4 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe à séro groupe rare (W_{135} , Y, autre): estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	4	67	63-130
DE et CE	1	0,6	0,44	-1,41	-1,36	1	64	63-82
DC et CE	1	0,8	0,37	-1,21	-1,16	1	64	63-82
DC et DE	1	5,3	0,02	3,28	3,34	0	63	63-71
CE	2	1,1	0,56	-2,85	-2,74	0	63	63-68
DE	2	5,3	0,07	1,33	1,43	0	63	63-67
DC	2	5,6	0,06	1,56	1,67	0	63	63-67
Pas d'interaction	3	5,6	0,13	-0,43	-0,27	0	63	63-66

Le modèle retenu est présenté en gras.

| TABLEAU 5 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de 0 à 3 ans : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	0	199	199-204
DE et CE	1	0,02	0,9	-1,98	-1,93	0	199	199-202
DC et CE	1	2,26	0,13	0,26	0,31	1	200	199-208
DC et DE	1	17,41	<10 ⁻⁴	15,41	15,47	0	199	199-201
CE	2	2,37	0,31	-1,63	-1,52	0	199	199-203
DE	2	17,60	<10 ⁻⁴	13,60	13,71	0	199	199-201
DC	2	20,79	<10 ⁻⁴	16,79	16,90	0	199	199-203
Pas d'interaction	3	20,81	<10 ⁻⁴	14,81	14,97	0	199	199-202

Le modèle retenu est présenté en gras.

| TABLEAU 6 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de 4 à 15 ans : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	1	119	118-140
DE et CE	1	0,07	0,79	-1,93	-1,87	1	119	118-125
DC et CE	1	0,65	0,42	-1,35	-1,30	3	121	118-139
DC et DE	1	22,13	<10 ⁻⁴	20,13	20,19	0	118	118-121
CE	2	0,82	0,66	-3,18	-3,07	2	120	118-125
DE	2	22,77	<10 ⁻⁴	18,77	18,88	0	118	118-121
DC	2	24,37	<10 ⁻⁴	20,37	20,48	0	118	118-123
Pas d'interaction	3	24,57	<10 ⁻⁴	18,57	18,73	0	118	118-122

Le modèle retenu est présenté en gras.

| TABLEAU 7 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de 16 à 25 ans : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	0	136	136-146
DE et CE	1	1,21	0,27	-0,79	-0,73	2	138	136-145
DC et CE	1	0,56	0,45	-1,44	-1,39	0	136	136-140
DC et DE	1	8,33	<10 ⁻⁴	6,33	6,39	0	136	136-139
CE	2	2,13	0,34	-1,87	-1,76	1	137	136-141
DE	2	10,45	0,01	6,45	6,56	0	136	136-141
DC	2	8,43	0,01	4,43	4,54	0	136	136-139
Pas d'interaction	3	10,87	0,01	4,87	5,03	0	136	136-140

Le modèle retenu est présenté en gras.

| TABLEAU 8 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de plus de 25 ans : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	31	205	178-344
DE et CE	1	9,3	<10 ⁻⁴	7,30	7,36	3	177	174-187
DC et CE	1	3,16	0,08	1,16	1,21	9	183	175-202
DC et DE	1	17,69	<10 ⁻⁴	15,69	15,74	6	180	174-202
CE	2	10,19	0,01	6,19	6,30	2	176	174-182
DE	2	21,46	<10 ⁻⁴	18,46	18,57	1	175	174-180
DC	2	19,04	<10 ⁻⁴	15,04	15,15	2	176	174-183
Pas d'interaction	3	22,68	<10 ⁻⁴	16,68	16,84	1	175	174-179

Le modèle retenu est présenté en gras.

| TABLEAU 8BIS |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe B ou C âgés de plus de 25 ans : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	4	130	126-175
DE et CE	1	1,78	0,18	-0,22	-0,17	1	127	126-133
DC et CE	1	0,98	0,32	-1,02	-0,96	1	127	126-135
DC et DE	1	6,2	0,01	4,2	4,26	1	127	126-139
CE	2	2,25	0,32	-1,75	-1,64	0	126	126-130
DE	2	7,18	0,03	3,18	3,29	0	126	126-130
DC	2	6,7	0,04	2,7	2,81	0	126	126-130
Pas d'interaction	3	7,39	0,06	1,39	1,55	0	126	126-129

Le modèle retenu est présenté en gras.

| TABLEAU 8TER |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe rare âgés de plus de 25 ans : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	9	46	37-458
DE et CE	1	1,75	0,19	-0,25	-0,2	0	37	37-44
DC et CE	1	1,24	0,27	-0,76	-0,71	2	39	37-125
DC et DE	1	2,51	0,11	0,51	0,56	1	38	37-58
CE	2	2,42	0,3	-1,58	-1,48	0	37	37-41
DE	2	3,24	0,2	-0,76	-0,65	0	37	37-41
DC	2	3,3	0,19	-0,7	-0,59	0	37	37-43
Pas d'interaction	3	3,71	0,29	-2,29	-2,12	0	37	37-40

Le modèle retenu est présenté en gras.

| TABLEAU 9 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant en Île-de-France : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	10	87	77-438
DE et CE	1	1,38	0,24	-0,62	-0,57	1	78	77-85
DC et CE	1	4,14	0,04	2,14	2,19	1	78	77-86
DC et DE	1	4,35	0,04	2,35	2,41	3	80	77-165
CE	2	4,46	0,11	0,46	0,57	0	77	77-81
DE	2	5,16	0,08	1,16	1,27	0	77	77-82
DC	2	7,5	0,02	3,5	3,61	0	77	77-81
Pas d'interaction	3	7,59	0,06	1,59	1,75	0	77	77,80

Le modèle retenu est présenté en gras.

| TABLEAU 10 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans l'interrégion 2 (quart nord-ouest de la France) : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	1	148	147-164
DE et CE	1	0,19	0,66	-1,81	-1,76	0	147	147-152
DC et CE	1	1,09	0,3	-0,91	-0,85	3	150	147-167
DC et DE	1	16,12	$<10^{-4}$	14,12	14,18	0	147	147-151
CE	2	1,51	0,47	-2,49	-2,38	1	148	147-153
DE	2	16,29	$<10^{-4}$	12,29	12,4	0	147	147-150
DC	2	18,31	$<10^{-4}$	14,31	14,41	0	147	147-153
Pas d'interaction	3	18,31	$<10^{-4}$	12,31	12,47	0	147	147-151

Le modèle retenu est présenté en gras.

| TABLEAU 11 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans l'interrégion 3 (quart nord-Est de la France) : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	2	170	168-193
DE et CE	1	1,78	0,18	-0,22	-0,17	0	168	168-174
DC et CE	1	0,38	0,54	-1,62	-1,57	4	172	168-186
DC et DE	1	22,25	$<10^{-4}$	20,25	20,31	0	168	168-172
CE	2	2,49	0,29	-1,51	-1,4	1	169	168-174
DE	2	22,42	$<10^{-4}$	18,42	18,53	0	168	168-171
DC	2	23,74	$<10^{-4}$	19,74	19,85	1	169	168-173
Pas d'interaction	3	24,21	$<10^{-4}$	18,21	18,37	0	168	168-172

Le modèle retenu est présenté en gras.

| TABLEAU 12 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans l'interrégion 4 (quart sud-est de la France) : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	4	166	162-208
DE et CE	1	0,09	0,76	-1,91	-1,85	2	164	162-174
DC et CE	1	1,75	0,19	-0,25	-0,19	1	163	162-171
DC et DE	1	19,91	$<10^{-4}$	17,91	17,97	0	162	162-171
CE	2	1,75	0,42	-2,25	-2,14	1	163	162-167
DE	2	19,99	$<10^{-4}$	15,99	16,1	1	163	162-167
DC	2	20,41	$<10^{-4}$	16,41	16,52	0	162	162-166
Pas d'interaction	3	20,65	$<10^{-4}$	14,65	14,82	0	162	162-166

Le modèle retenu est présenté en gras.

| TABLEAU 13 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans l'interrégion 5 (quart sud-ouest de la France) : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	1	74	73-101
DE et CE	1	0,67	0,41	-1,33	-1,28	0	73	73-80
DC et CE	1	0,01	0,93	-1,99	-1,94	1	74	73-83
DC et DE	1	5,06	0,02	3,06	3,11	0	73	73-79
CE	2	0,68	0,71	-3,32	-3,22	0	73	73-78
DE	2	5,22	0,07	1,22	1,32	0	73	73-76
DC	2	5,15	0,08	1,15	1,25	0	73	73-77
Pas d'interaction	3	5,37	0,15	-0,63	-0,46	0	73	73-76

Le modèle retenu est présenté en gras.

Annexe 3 – Exhaustivité de la surveillance des IIM par le réseau Epibac

La première partie de l'analyse portait sur l'ensemble des cas identifiés par les trois sources, sachant que cette analyse sous-estimait l'exhaustivité de la surveillance par Epibac (paragraphe 3.3). Elle a permis d'estimer l'exhaustivité de la DO et du CNR.

La seconde analyse, présentée ci-après, a été restreinte aux hôpitaux dont les laboratoires participent au réseau Epibac et visait à estimer l'exhaustivité de ce réseau.

1. CAS RETENUS POUR L'ANALYSE

Les cas retenus sont les cas répondant à la définition de cas (chapitre 3.3) et identifiés à partir des hôpitaux dont la bactériologie était traitée par un laboratoire participant à Epibac (www.invs.sante.fr/surveillance/index.htm) en 2005 dans les sources DO et CNR.

1.1 DO

Parmi les 588 cas retenus dans la 1^{re} analyse, **478** cas ont été retenus pour l'analyse spécifique du réseau Epibac.

1.2 CNR

Parmi les 481 cas notifiés au CNR, **396** cas ont été inclus dans l'analyse.

1.3 Epibac

L'ensemble des **425** cas notifiés à Epibac ont été inclus dans l'analyse.

Au total, 511 cas ont été inclus dans l'analyse.

2. CARACTÉRISTIQUES DES CAS DÉCLARÉS PAR SOURCE

2.1 Âge

La distribution des cas par groupe d'âge (tableau 1) était similaire dans les trois sources ($p=1,0$).

2.2 Sérotype

La distribution des sérotypes (tableau 2) différait significativement selon la source ($p<0,01$). Toutes les souches analysées par le CNR ont été caractérisées pour le sérotype contrairement aux deux autres sources pour lesquelles des souches sont restées de sérotype inconnu. Si l'on exclut les sujets dont le sérotype est inconnu, la distribution des cas ne diffère pas significativement ($p=0,97$).

2.3 Trimestre de diagnostic

La distribution des cas par trimestre d'hospitalisation, prélèvement ou isolement (tableau 3) était comparable dans les trois sources ($p=0,96$).

2.4 Région d'origine

La distribution des cas par région (tableau 4) était comparable dans les trois sources ($p=1,0$).

| TABLEAU 1 |

Distribution des cas déclarés dans les trois sources selon le groupe d'âge, France, 2005

Groupe d'âge	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
0-3 ans	163 (32)	156	33	133	33	142	33
4-15 ans	97 (19)	89	19	70	18	72	17
16-25 ans	104 (20)	98	20	82	21	88	21
>25 ans	147 (29)	135	28	111	28	123	29
Total	511 (100)	478	100	396	100	425	100

| TABLEAU 2 |

Distribution des cas déclarés dans les trois sources selon le sérotype, France, 2005

Sérotype	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
B	315 (62)	298	62	253	64	271	64
C	122 (24)	116	24	98	25	106	25
Rares	52 (10)	46	10	45	11	42	10
Inconnu	22 (4)	18	4	0	0,0	6	1
Total	511 (100)	478	100	396	100	425	100

| TABLEAU 3 |

Distribution des cas déclarés dans les trois sources selon le trimestre de diagnostic, France, 2005

Trimestre	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
1 ^{er} trimestre	193 (38)	180	38	159	40	164	39
2 ^e trimestre	126 (25)	119	25	86	22	103	24
3 ^e trimestre	77 (15)	72	15	59	15	61	14
4 ^e trimestre	115 (22)	107	22	92	23	97	23
Total	511 (100)	478	100	396	100	425	100

| TABLEAU 4 |

Distribution des cas déclarés dans les trois sources selon la région de résidence ou de prise en charge, France, 2005

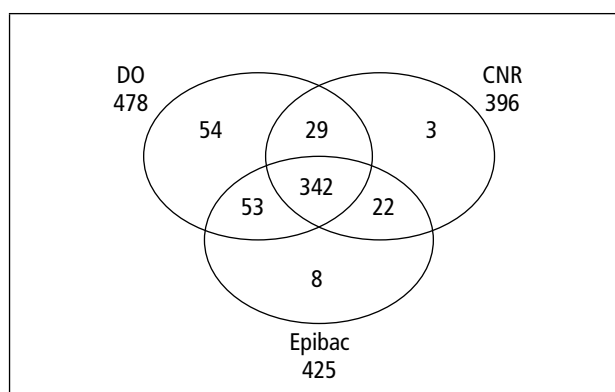
Région téléphonique	Effectifs (%)	DO		CNR		Epibac	
		n	%	n	%	n	%
Région 1	61 (12)	57	12	49	12	52	12
Région 2	105 (21)	98	21	80	20	82	19
Région 3	146 (29)	135	28	116	29	123	29
Région 4	136 (26)	131	27	104	27	113	27
Région 5	63 (12)	57	12	47	12	55	13
Total	511 (100)	478	100	396	100	425	100

2.5 Distribution des cas selon leur(s) source(s) d'appartenance

En 2005, 511 cas ont été identifiés par les trois sources dont 425 (83,2 %) par Epibac.

La distribution des cas selon leur(s) source(s) d'appartenance est présentée à la figure 1 : 87,3 % (446/511) des cas étaient communs à au moins deux sources, 66,9 % (342/511) étaient communs aux trois sources et 12,7 % (65/511) des cas n'ont été déclarés que par une seule source. Parmi ces 65 cas, 54 cas, soit 83,1 % ont été déclarés par le système de DO.

| FIGURE 1 |

Distribution des 511 cas selon la (les) source(s) de déclaration

3. ÉTUDE DE LA DÉPENDANCE ENTRE LES SOURCES

3.1 Étude des sources deux à deux (méthode de Wittes)

En croisant les sources 2 à 2, on obtenait une estimation du nombre total de cas variant de 462 à 514 (tableau 5). Le nombre de cas estimés était inférieur en croisant les sources CNR et Epibac, suggérant une dépendance positive entre ces deux sources (sous-estimation du nombre de cas par rapport aux deux autres estimations).

3.2 Estimation de la dépendance par le rapport de cote (méthode de Wittes)

La recherche de la dépendance entre deux sources a été effectuée à partir de la 3^e source en calculant le rapport de cote. Il existait une dépendance positive entre CNR et Epibac, la dépendance entre DO et CNR était à la limite de la significativité (tableau 6).

| TABLEAU 5 |

Estimation du nombre total de cas en croisant les sources 2 à 2, exhaustivité et intervalle de confiance, France, 2005

Source 1, source 2	n1	n2	n12	n	N _E	IC 95 %	Exhaustivité1 (%)	IC 95 %	Exhaustivité2 (%)	IC 95 %
DO, CNR	478	396	371	503	510	491-530	94	91-96	78	74-81
DO, Epibac	478	425	395	508	514	498-530	93	91-95	83	79-86
CNR, Epibac	396	425	364	457	462	449-476	86	82-89	92	89-94

n1 : cas dans la source 1 ; n2 : cas source 2 ; n12 : cas communs aux deux sources ; n : cas observés dans les deux sources ; N_E : cas estimés en croisant les deux sources.

| TABLEAU 6 |

Rapport de cote de la répartition entre deux sources des cas présents dans la troisième source – Test d'indépendance des sources deux à deux, méthode de Wittes

Sources étudiées	Système dans lequel sont faites les observations	RC	IC 95 %	Indépendance
DO, CNR	Epibac	2,3	0,9-5,8	Pas de dépendance
DO, Epibac	CNR	1,6	0,3-5,8	Pas de dépendance
CNR, Epibac	DO	12,0	6,8-21,3	Dépendance positive

3.3 Analyse globale selon les modèles log-linéaires

Les résultats de l'analyse globale (ajustement des 8 modèles log-linéaires aux données globales) sont présentés en annexe 2 – tableau 1.

Seuls trois modèles présentaient une bonne adéquation aux données observées : le modèle saturé, le modèle prenant en compte les deux dépendances DO*CNR et CNR*Epibac, et le modèle prenant en compte uniquement l'interaction entre le CNR et Epibac.

Le modèle retenu (valeurs AIC et BIC les plus basses) est celui considérant deux dépendances entre les sources, DO*CNR et CNR*Epibac. Le nombre total de cas d'IIM estimé était de 519 [IC 95 %=512-531] (tableau 7).

L'exhaustivité de la DO était de 92,1 % [IC 95 %=90,0-93,4], celle du CNR de 76,3 % [IC 95 %=74,6-77,3] et celle du réseau Epibac de 81,9 % [IC 95 %=80,0-83,0] (tableau 7).

| TABLEAU 7 |

Exhaustivité des trois sources à partir de l'estimation du nombre total de cas estimé

N _E = 519	Exhaustivité (%)	IC 95 %
DO	92	90-93
CNR	76	75-77
Epibac	82	80-83

4. ANALYSE STRATIFIÉE

4.1 Sérogroupes

Comme pour l'analyse globale, aucun cas de séro groupe inconnu n'ayant été recensé par le CNR, l'estimation du nombre d'IIM n'a pas été réalisée dans cette catégorie (annexe 3bis – figure 1).

Les modèles ayant conduit aux estimations apparaissent en gras dans les tableaux 2, 3 et 4 en annexe 3bis. Les trois modèles retenus présentaient une interaction entre les sources CNR et Epibac.

L'exhaustivité de la DO était de 94 % [IC 95 %=93-95] pour le séro groupe B, de 95 % [IC 95 %=93-95] pour le séro groupe C et de 87 % [IC 95 %=81-87] pour les autres sérogroupes (tableau 8). L'exhaustivité

du CNR était de 80 % [IC 95 %=79-80] pour le séro groupe B, de 80 % [IC 95 %=78-80] pour le séro groupe C et de 86 % [IC 95 %=82-86] pour les autres sérogroupes. Ces chiffres sont comparables à ceux de l'analyse globale.

L'exhaustivité d'Epibac était de 85 % [IC 95 %=84-86] pour le séro groupe B, de 87 % [IC 95 %=85-87] pour le séro groupe C et de 81 % [IC 95 %=76-81] pour les autres sérogroupes.

Les résultats sont en faveur d'une hétérogénéité de capture selon le séro groupe pour la source Epibac, l'exhaustivité du réseau était moins élevée pour les sérogroupes différents de B et C (tableau 8). On retrouve, comme dans l'analyse présentée dans le chapitre 4, la moins bonne exhaustivité de la DO pour les sérogroupes rares.

| TABLEAU 8 |

Exhaustivité des trois sources selon le séro groupe

Trois sources	N _E	DO		CNR		Epibac	
		Exhaustivité (%)	IC 95 %	Exhaustivité (%)	IC 95 %	Exhaustivité (%)	IC 95 %
B	316	94	93-95	80	79-80	86	84-86
C	122	95	93-95	80	78-80	87	85-87
Rares	63	89	84-88	86	82-87	81	76-81

4.2 Âge

Les cas âgés de 4 à 15 ans représentaient 19 % (97/511) des cas alors que ceux âgés de moins de 3 ans représentaient 32 % (annexe 3bis – figure 2).

Les modèles ayant conduit aux estimations apparaissent en gras dans les tableaux 5 à 8 en annexe 3bis. Les modèles retenus présentaient une interaction entre les sources CNR et Epibac avec en plus pour une interaction entre DO et CNR chez les plus de 25 ans.

Pour les cas d'IIM survenant chez des adultes de plus de 25 ans, le tableau 8 en annexe 1 montre qu'aucun modèle ne présente un AIC et BIC de valeurs inférieures à celles observées avec le modèle saturé. L'estimation avec le modèle saturé paraît surestimée. Parmi les autres modèles, le modèle avec une interaction entre les sources DO et CNR et CNR et Epibac avait les valeurs d'AIC et BIC les plus basses (mais positives).

Les trois sources ont identifié 163 cas de 0-3 ans, 97 cas de 4-15 ans, 104 cas de 16-25 ans, 147 cas de plus de 25 ans et les nombres estimés de cas par les modèles étaient respectivement de 163 [IC 95 %=163-167], 99 [IC 95 %=97-104], de 104 [IC 95 %=104-108] et de 152 [IC 95 %=147-164].

L'exhaustivité de la DO variait entre 96 % [IC 95 %=93-96] chez les 0-3 ans et 89 % [IC 95 %=82-92] chez les plus de 25 ans.

Pour Epibac, l'analyse a également mis en évidence une hétérogénéité de capture selon les groupes d'âges. L'exhaustivité du réseau Epibac était plus élevée pour les 0-3 ans [87 % ; IC 95 %=85-87] et plus basse pour les 4-15 ans [73 % ; IC 95 %=69-74] (tableau 9).

Quand on somme les nombres estimés d'IIM dans chaque strate, on obtient un total de 518, nombre similaire à l'estimation par le modèle global (519). L'exhaustivité globale de la DO serait de 92,3 % (478/518), celle du CNR de 76,4 % (396/518), et celle du réseau Epibac de 82,0 % (425/518).

4.3 Région

Plus de la moitié des cas (55,2 %) ont été signalés dans les régions 3 et 4 (annexe 3bis – figure 3).

Les modèles ayant conduit aux estimations apparaissent en gras dans les tableaux 9 à 13 en annexe 3 bis. Les modèles retenus présentaient tous une interaction entre les sources CNR et Epibac avec en plus pour certains une interaction entre DO et Epibac (région 1) ou DO et CNR (région 3) comme dans l'analyse globale.

Les trois sources ont identifié 61 cas dans la région 1 (Île-de-France), 105 cas dans la région 2 (quart nord-ouest), 146 dans la région 3 (quart nord-est), 136 dans la région 4 (quart sud-est) et 63 dans la région 5 (quart sud-ouest).

Comme pour les deux autres variables de stratification, les nombres estimés de cas par les modèles sont égaux ou très proches des nombres observés et étaient respectivement de 63 [IC 95 % = 61-82],

106 [IC 95 % = 105-111], 149 [IC 95 % = 146-159], 136 [IC 95 % = 136-140], 63 [IC 95 % = 63-67] (tableau 10).

Pour la DO, les niveaux d'exhaustivité varient entre 90 et 95 % selon les régions, avec une exhaustivité plus élevée dans la région 4 ce qui n'était pas le cas dans l'analyse globale. Pour le CNR, les exhaustivités sont comparables d'une région à l'autre.

Il apparaît que pour Epibac, l'exhaustivité de la région 2 (quart nord-ouest) est inférieure aux régions 4 et 5.

Compte tenu des interactions prises en considération dans les modèles choisis selon les critères définis dans la méthode, il apparaît que les dépendances entre les sources varient selon les régions.

Quand on additionne les nombres estimés d'IIM dans chaque strate, on obtient un total de 517, nombre très proche de l'estimation par le modèle global (519). L'exhaustivité globale de la DO serait de 92,5 % (478/517), celle du CNR de 76,6 % (396/517), et celle du réseau Epibac de 82,2 % (425/517).

| TABLEAU 9 |

Exhaustivité des trois sources selon l'âge

3 sources	N _E	DO		CNR		Epibac	
		Exhaustivité (%)	IC 95 %	Exhaustivité (%)	IC 95 %	Exhaustivité (%)	IC 95 %
0-3 ans	163	96	93-96	82	80-82	87	85-87
4-15 ans	99	90	86-92	71	67-72	73	69-74
16-25 ans	104	94	91-94	79	76-79	85	82-85
>25 ans	152	89	82-92	73	68-75	81	75-84
Total	518	92	-	76	-	82	-

| TABLEAU 10 |

Exhaustivité des trois sources selon la région de résidence

3 sources	N _E	DO		CNR		Epibac	
		Exhaustivité (%)	IC 95 %	Exhaustivité (%)	IC 95 %	Exhaustivité (%)	IC 95 %
Région 1	63	90	69-93	78	60-80	82	63-85
Région 2	106	92	88-93	75	72-76	77	74-78
Région 3	149	91	85-92	78	73-80	83	77-84
Région 4	136	96	94-96	76	74-77	83	81-83
Région 5	63	90	85-90	75	70-75	87	82-87
Total	517	92	-	77	-	82	-

5. ÉVOLUTION DE L'EXHAUSTIVITÉ SELON LES SOURCES [2,8]

L'exhaustivité du réseau Epibac a augmenté entre 2000 et 2005.

| TABLEAU 11 |

Évolution de l'exhaustivité d'Epibac – 2000-2005

Années	Epibac	
	Exhaustivité (%)	IC 95 %
2000	74	70-78
2005	82	80-83

6. REPRÉSENTATIVITÉ DES SOURCES

6.1 Sérogroupes

L'analyse stratifiée a montré que pour la source Epibac, les sérogroupes autres que B et C étaient sous-représentés. Cependant, la répartition

des cas estimés selon le séro groupe quand celui-ci est connu (N_E) ne diffère pas statistiquement (test du χ_2) de celle des cas observés dans la DO, le CNR et Epibac ($p=0,45$; 0,86 et 0,48 respectivement) (tableau 12).

6.2 Âge

L'analyse stratifiée a montré que pour la source Epibac, les enfants de 4-15 ans étaient sous-représentés. Cependant, la répartition des cas estimés selon les groupes d'âges (N_E) ne diffère pas statistiquement (test du χ_2) de celle des cas observés dans la DO, le CNR et Epibac ($p=0,91$; 0,86 et 0,82 respectivement) (tableau 13).

6.3 Région

L'analyse stratifiée a montré que pour la source Epibac, les cas résidents en région 2 (Nord-Ouest) étaient sous-représentés. Cependant, la répartition des cas estimés selon la région (N_E) ne diffère pas statistiquement (test du χ_2) de celle des cas observés dans la DO, le CNR et Epibac ($p=0,99$; 0,99 et 0,99 respectivement) (tableau 14).

| TABLEAU 12 |

Distribution des cas observés selon le séro groupe dans les trois sources et des cas estimés avec les modèles retenus dans l'analyse stratifiée

	DO observés	%	CNR observés	%	Epibac observés	%	N_E	%
B	298	65	253	64	271	65	316	63
C	116	25	98	25	106	25	122	24
Rares	46	10	45	11	42	10	63	13
Total	460	100	396	100	419	100	501	100

| TABLEAU 13 |

Distribution des cas observés selon l'âge dans les trois sources et des cas estimés avec les modèles retenus dans l'analyse stratifiée

	DO observés	%	CNR observés	%	Epibac observés	%	N_E	%
0-3 ans	156	33	133	33	142	33	163	30
4-15 ans	89	19	70	18	72	17	99	18
16-25 ans	98	20	82	21	88	21	104	20
>25 ans	135	28	111	28	123	29	152	32
Total	478	100	396	100	425	100	518	100

| TABLEAU 14 |

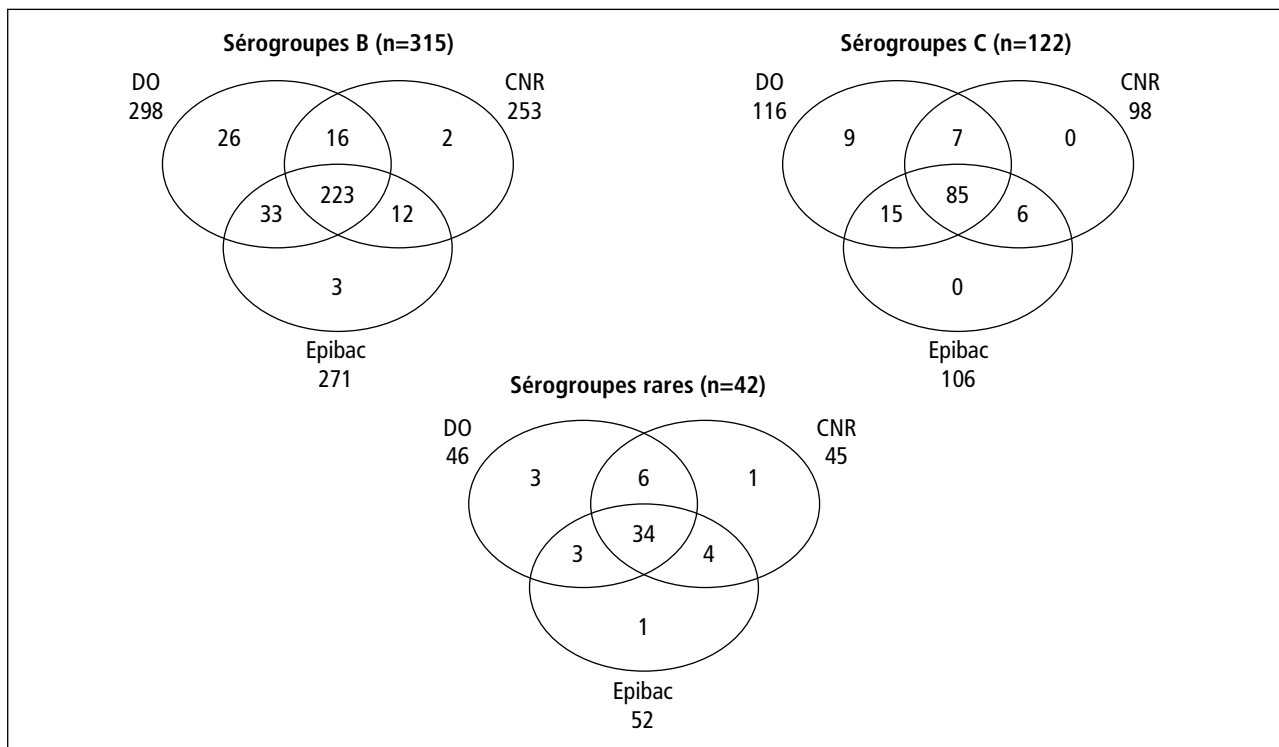
Distribution des cas observés selon la région dans les trois sources et des cas estimés avec les modèles retenus dans l'analyse stratifiée

	DO observés	%	CNR observés	%	Epibac observés	%	N_E	%
Région 1	57	12	49	12	52	12	63	12
Région 2	98	21	80	20	82	19	106	21
Région 3	135	28	116	29	123	29	149	29
Région 4	131	27	104	27	113	27	136	26
Région 5	57	12	47	12	55	13	63	12
Total	478	100	396	100	425	100	517	100

Annexe 3bis – Illustrations de l'analyse capture-recapture restreinte aux laboratoires Epibac

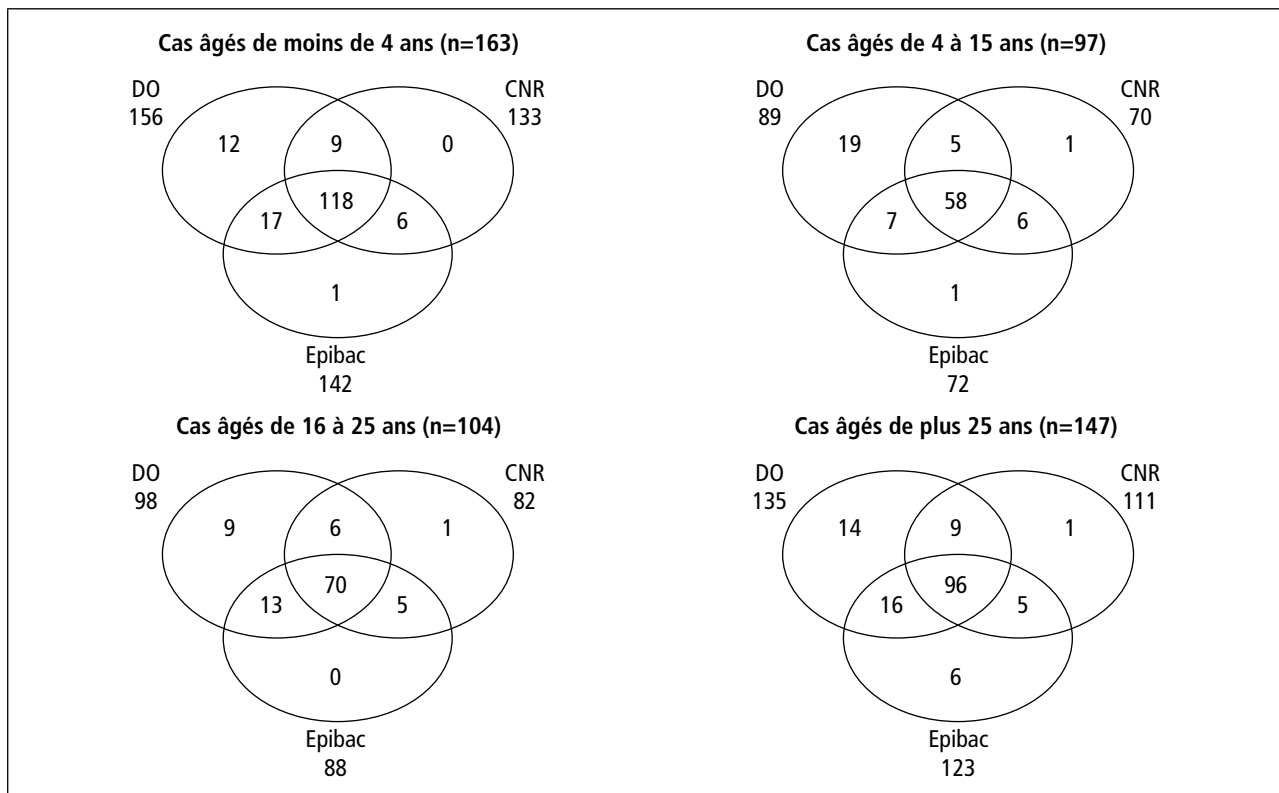
| FIGURE 1 |

Distribution des cas selon le sérotype et les sources de déclaration

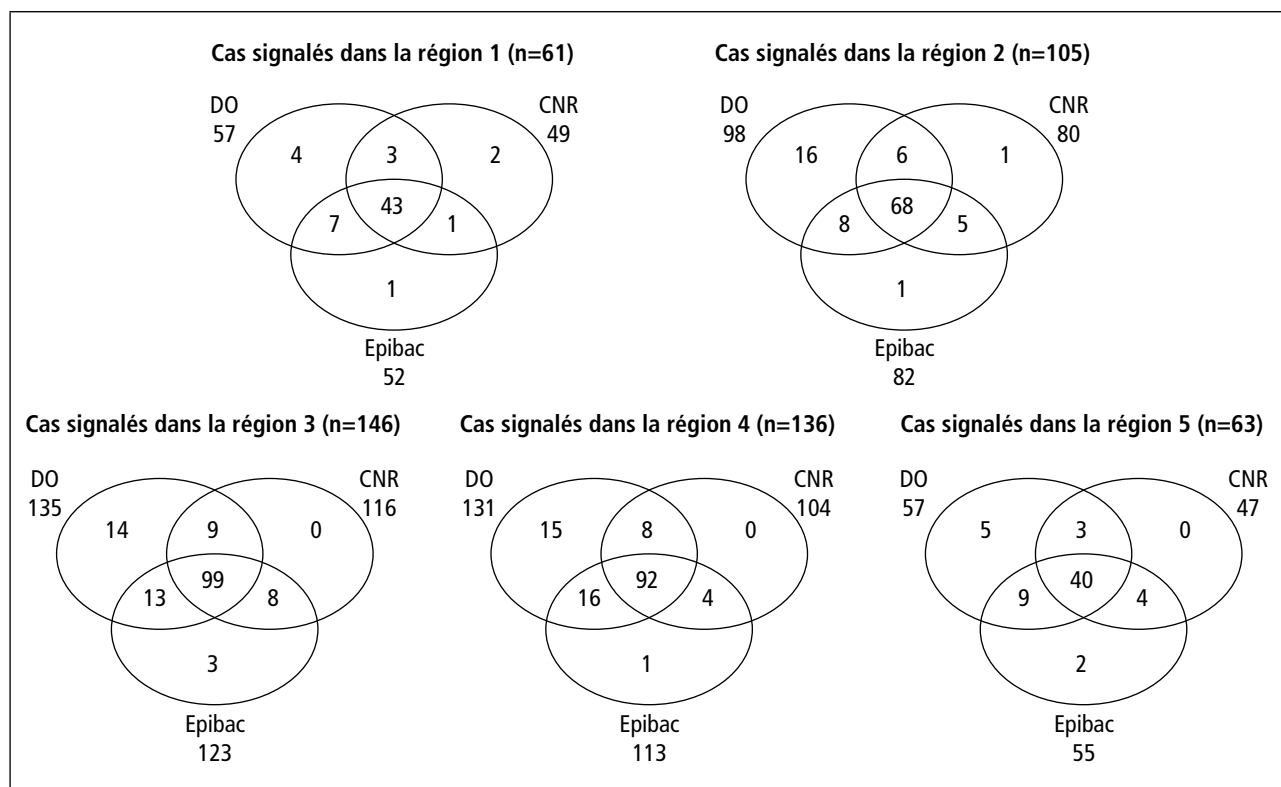


| FIGURE 2 |

Distribution des cas selon l'âge et les sources de déclaration



Distribution des cas selon la région et la (les) source(s) de déclaration



Résultats de l'analyse globale par modélisations log-linéaires : estimations des nombres de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95%
Modèle saturé	0	0	1	0	0	13	524	512-565
DE et CE	1	3,37	0,07	1,37	1,42	5	516	511-530
DC et CE	1	0,49	0,48	-1,51	-1,46	8	519	512-531
DC et DE	1	89,47	<10 ⁻⁴	87,47	87,52	1	512	511-517
CE	2	3,59	0,17	-0,41	-0,3	4	515	511-521
DE	2	89,75	<10 ⁻⁴	85,75	85,86	0	511	511-516
DC	2	89,99	<10 ⁻⁴	85,99	86,1	1	512	511-517
Pas d'interaction	3	90,45	<10 ⁻⁴	84,45	84,61	1	512	511-516

Le modèle retenu est présenté en gras.

D: DO; C: CNR; E: Epibac; $AIC = G^2 - 2 \times ddl$; $BIC = G^2 - (\log N_{obs} / 2\pi) \times (ddl)$; Xest: nombre estimé de cas notifiés par aucune des sources; N_E : nombre total de cas estimé.

| TABLEAU 2 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe B : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	5	320	315-351
DE et CE	1	0,56	0,46	-1,44	-1,39	3	318	315-329
DC et CE	1	0,93	0,34	-1,07	-1,02	2	317	315-324
DC et DE	1	43,99	<10 ⁻⁴	41,99	42,04	0	315	315-319
CE	2	1,33	0,51	-2,67	-2,56	1	316	315-321
DE	2	43,99	<10 ⁻⁴	39,99	40,09	0	315	315-319
DC	2	43,99	<10 ⁻⁴	39,99	40,09	0	315	315-318
Pas d'interaction	3	43,99	<10 ⁻⁴	37,99	38,15	0	315	315-318

Le modèle retenu est présenté en gras.

| TABLEAU 3 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupe C : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	0	122	122-126
DE et CE	1	1,26	0,26	-0,74	-0,69	0	122	122-127
DC et CE	1	0,42	0,51	-1,58	-1,52	0	122	122-125
DC et DE	1	11,81	<10 ⁻⁴	9,81	9,86	0	122	122-123
CE	2	1,59	0,45	-2,41	-2,3	0	122	122-125
DE	2	13,81	<10 ⁻⁴	9,81	9,92	0	122	122-124
DC	2	12,92	<10 ⁻⁴	8,92	9,03	0	122	122-124
Pas d'interaction	3	14,67	<10 ⁻⁴	8,67	8,83	0	122	122-124

Le modèle retenu est présenté en gras.

| TABLEAU 4 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de séro groupes rares (W_{135} , Y, autres) : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	1	53	52-88
DE et CE	1	0,59	0,44	-1,41	-1,36	0	52	52-57
DC et CE	1	0,08	0,78	-1,92	-1,87	0	52	52-63
DC et DE	1	3,34	0,07	1,34	1,4	0	52	52-56
CE	2	0,62	0,73	-3,38	-3,27	0	52	52-55
DE	2	3,51	0,17	-0,49	-0,38	0	52	52-54
DC	2	3,34	0,19	-0,66	-0,55	0	52	52-55
Pas d'interaction	3	3,51	0,32	-2,49	-2,33	0	52	52-54

Le modèle retenu est présenté en gras.

| TABLEAU 5 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de 0-3 ans : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95%
Modèle saturé	0	0	1	0	0	0	163	163-173
DE et CE	1	0,02	0,9	-1,98	-1,93	0	163	163-168
DC et CE	1	0,38	0,54	-1,62	-1,57	0	163	163-169
DC et DE	1	18,93	$<10^{-4}$	16,93	16,99	0	163	163-165
CE	2	0,41	0,81	-3,59	-3,48	0	163	163-167
DE	2	19,02	$<10^{-4}$	15,02	15,13	0	163	163-165
DC	2	20,02	$<10^{-4}$	16,02	16,13	0	163	163-165
Pas d'interaction	3	20,05	$<10^{-4}$	14,05	14,21	0	163	163-165

Le modèle retenu est présenté en gras.

| TABLEAU 6 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de 4-15 ans : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	5	102	97-185
DE et CE	1	0,07	0,79	-1,93	-1,87	3	100	97-124
DC et CE	1	0,28	0,6	-1,72	-1,67	2	99	97-115
DC et DE	1	38,54	$<10^{-4}$	36,54	36,6	0	97	97-100
CE	2	0,33	0,85	-3,67	-3,57	2	99	97-104
DE	2	39,35	$<10^{-4}$	35,35	35,46	0	97	97-101
DC	2	39,15	$<10^{-4}$	35,15	35,26	0	97	97-100
Pas d'interaction	3	39,72	$<10^{-4}$	33,72	33,88	0	97	97-101

Le modèle retenu est présenté en gras.

| TABLEAU 7 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de 16-25 ans : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	0	104	104-116
DE et CE	1	1,05	0,31	-0,95	-0,9	1	105	104-116
DC et CE	1	0,45	0,5	-1,55	-1,5	0	104	104-107
DC et DE	1	12,14	$<10^{-4}$	10,14	10,19	0	104	104-106
CE	2	1,65	0,44	-2,35	-2,24	0	104	104-108
DE	2	13,93	$<10^{-4}$	9,93	10,04	0	104	104-107
DC	2	12,14	$<10^{-4}$	8,14	8,25	0	104	104-106
Pas d'interaction	3	14,01	$<10^{-4}$	8,01	8,17	0	104	104-106

Le modèle retenu est présenté en gras.

| TABLEAU 8 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM de plus de 25 ans : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	11	158	147-254
DE et CE	1	8,51	<10 ⁻⁴	6,51	6,57	1	148	147-158
DC et CE	1	0,38	0,54	-1,62	-1,57	5	152	147-164
DC et DE	1	20,37	<10 ⁻⁴	18,37	18,42	1	148	147-157
CE	2	8,53	0,01	4,53	4,63	1	148	147-152
DE	2	25,28	<10 ⁻⁴	21,28	21,39	0	147	147-150
DC	2	20,37	<10 ⁻⁴	16,37	16,47	1	148	147-152
Pas d'interaction	3	25,56	<10 ⁻⁴	19,56	19,72	0	147	147-150

Le modèle retenu est présenté en gras.

| TABLEAU 9 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans la région 1 : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	16	77	61-777
DE et CE	1	1,38	0,24	-0,62	-0,57	2	63	61-82
DC et CE	1	6,3	0,01	4,3	4,35	0	61	61-67
DC et DE	1	5,86	0,02	3,86	3,92	1	62	61-124
CE	2	6,66	0,04	2,66	2,77	0	61	61-64
DE	2	6,77	0,03	2,77	2,88	0	61	61-65
DC	2	10,31	0,01	6,31	6,41	0	61	61-64
Pas d'interaction	3	10,46	0,02	4,46	4,62	0	61	61-63

Le modèle retenu est présenté en gras.

| TABLEAU 10 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans la région 2 : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95 %
Modèle saturé	0	0	1	0	0	4	109	105-181
DE et CE	1	0,19	0,66	-1,81	-1,76	2	107	105-123
DC et CE	1	0,42	0,52	-1,58	-1,53	2	107	105-118
DC et DE	1	32,18	<10 ⁻⁴	30,18	30,23	0	105	105-109
CE	2	0,55	0,76	-3,45	-3,34	1	106	105-111
DE	2	32,38	<10 ⁻⁴	28,38	28,49	0	105	105-108
DC	2	32,29	<10 ⁻⁴	28,29	28,4	0	105	105-108
Pas d'interaction	3	32,45	<10 ⁻⁴	26,45	26,61	0	105	105-108

Le modèle retenu est présenté en gras.

| TABLEAU 11 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans la région 3 : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95%
Modèle saturé	0	0	1	0	0	0	146	146-166
DE et CE	1	1,78	0,18	-0,22	-0,17	0	146	146-152
DC et CE	1	0,77	0,38	-1,23	-1,17	3	149	146-159
DC et DE	1	23,9	<10 ⁻⁴	21,9	21,95	0	146	146-149
CE	2	2,8	0,25	-1,2	-1,1	1	147	146-151
DE	2	24,21	<10 ⁻⁴	20,21	20,32	0	146	146-148
DC	2	26,08	<10 ⁻⁴	22,08	22,19	0	146	146-150
Pas d'interaction	3	26,67	<10 ⁻⁴	20,67	20,84	0	146	146-149

Le modèle retenu est présenté en gras.

| TABLEAU 12 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans la région 4 : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95%
Modèle saturé	0	0	1	0	0	0	136	136-154
DE et CE	1	0,09	0,76	-1,91	-1,85	0	136	136-143
DC et CE	1	0,23	0,64	-1,77	-1,72	0	136	136-143
DC et DE	1	23,03	<10 ⁻⁴	21,03	21,09	0	136	136-138
CE	2	0,35	0,84	-3,65	-3,54	0	136	136-140
DE	2	23,07	<10 ⁻⁴	19,07	19,18	0	136	136-138
DC	2	23,96	<10 ⁻⁴	19,96	20,07	0	136	136-139
Pas d'interaction	3	23,97	<10 ⁻⁴	17,97	18,13	0	136	136-138

Le modèle retenu est présenté en gras.

| TABLEAU 13 |

Résultats des modélisations log-linéaires dans la strate des cas d'IIM résidant dans la région 5 : estimation du nombre de cas recensés par aucune des sources Xest et du total N_E , critères utilisés pour la sélection du modèle : déviance (G^2), p-value du test d'adéquation, AIC et BIC

Modèles	ddl	G^2	p	AIC	BIC	Xest	N_E	IC 95%
Modèle saturé	0	0	1	0	0	0	63	63-89
DE et CE	1	0,67	0,41	-1,33	-1,28	0	63	63-71
DC et CE	1	0,16	0,69	-1,84	-1,79	1	64	63-70
DC et DE	1	6,23	0,01	4,23	4,28	0	63	63-66
CE	2	0,91	0,63	-3,09	-2,98	0	63	63-67
DE	2	6,44	0,04	2,44	2,55	0	63	63-65
DC	2	6,86	0,03	2,86	2,97	0	63	63-66
Pas d'interaction	3	7,2	0,07	1,2	1,36	0	63	63-65

Le modèle retenu est présenté en gras.

Surveillance des infections invasives à méningocoque en France métropolitaine en 2005

Évaluation quantitative par la méthode de capture-recapture à trois sources

En France, l'incidence des infections invasives à méningocoque (IIM) est évaluée à travers le dispositif de déclaration obligatoire (DO). Régulièrement, l'exhaustivité du dispositif est évaluée par analyse capture-recapture pour corriger l'incidence et suivre les tendances. Le Centre national de référence (CNR) des méningocoques reçoit les souches invasives de méningocoque de l'ensemble des laboratoires de bactériologie. Le réseau de surveillance des infections invasives bactériennes (EpiBac) collecte des données sur des cas avec isolement de méningocoque dans le sang ou le liquide céphalo-rachidien (LCR).

Une analyse capture-recapture à trois sources (DO, CNR, EpiBac) a permis d'estimer, pour l'année 2005, l'exhaustivité de chacune des trois sources.

L'analyse a porté sur les cas pour lesquels une souche de méningocoque a été isolée dans le sang ou le LCR, critère commun aux trois systèmes. En raison d'une couverture nationale restreinte, l'estimation de l'exhaustivité du réseau EpiBac a été effectuée à partir des cas admis dans des hôpitaux participant effectivement au réseau EpiBac en 2005. Les dépendances entre les sources ont été recherchées en croisant les sources deux à deux (méthode de Wittes). Des modèles log-linéaires ont été utilisés pour estimer le nombre de cas recensés dans aucune des sources. Une approche globale puis stratifiée selon l'âge, le sérotype et la région a été faite.

En 2005, 627 cas d'IIM confirmés par culture dans le sang et/ou le LCR ont été déclarés dans au moins une des trois sources (DO=588, CNR=481, EpiBac=425). Des dépendances entre le CNR et la DO et entre le CNR et EpiBac ont été identifiées. Le nombre total d'IIM estimé en 2005 était de 639 [IC 95%=630-656]. Le taux d'exhaustivité de la DO a été estimé à 92% [IC 95%=90-93], celui du CNR à 76% [IC 95%=75-77]. Le taux d'exhaustivité au sein du réseau EpiBac a été estimé à 82% [IC 95%=80-83]. Les analyses stratifiées ont mis en évidence une moindre exhaustivité pour les sérotypes différents de B et C ainsi qu'une exhaustivité plus élevée pour les enfants de moins de 4 ans.

Les résultats de cette analyse ont montré une nette amélioration de l'exhaustivité de la DO et du réseau EpiBac en 2005 par rapport à une précédente évaluation en 2000 (73% pour la DO et 57% pour EpiBac), celle du CNR étant restée stable. Le niveau élevé de l'exhaustivité de la DO des IIM confirmées par culture en 2005 doit conduire à un suivi régulier de cette exhaustivité afin d'écartier une possible baisse dans les années qui suivent. Il doit également tenir compte de l'évolution des pratiques en matière de confirmation biologique comme la PCR.

Mots clés : méningocoque, surveillance épidémiologique, déclaration maladie, réseau surveillance, modèle capture-recapture, France

Surveillance of the invasive meningococcal infections in France in 2005: Quantitative evaluation by the three sources capture-recapture method.

In France, incidence of Invasive Meningococcal Disease (IMD) is estimated through Mandatory Notification (MN). Regularly, the completeness of this national surveillance system is evaluated by capture-recapture analysis and results are used to adjust incidence trends. The National Reference Centre for Meningococci (NRCM) collects the invasive strains from all the French laboratories. Data on cases with strains isolated in blood or cerebrospinal fluid are also collected through a sentinel network of hospital laboratories for bacterial invasive diseases (EPIBAC). A three-source capture-recapture analysis was performed to assess the completeness of MN, NRC and EPIBAC in 2005.

The analysis was restricted to laboratory culture confirmed IMD cases (isolation of Neisseria meningitidis strains from blood and/or CSF). Because EPIBAC has an incomplete geographical coverage, an analysis restricted to cases hospitalized in EPIBAC participating hospitals was performed. Dependence between paired sources was studied using Wittes method. Log-linear models contributed to estimate the number of cases not registered in any of the sources. A global approach followed by stratified analyses on age, serogroup, and geographical region were done.

A total of 627 laboratory confirmed cases were identified from the three sources: MN=588, NRC=488, EPIBAC=425. Dependences between MN and NRC and between NRC and EPIBAC were identified. The overall estimated number of cases was 639 [95% CI=630-656], yielding a completeness of reporting of 92% [CI 95%=90-93] for MN and of 76% [CI 95%=75-77] for NRC. The completeness of EPIBAC was 82% [CI 95%=80-83]. Stratified analyses suggested a lower completeness of reporting for non B and non C serogroups, and a higher completeness of reporting for children under 4 years of age.

These analysis results indicate that the completeness of MN and EPIBAC improved substantially in 2005 compared to 2000 (73% for MN and 57% for EPIBAC) while it was stable for the NRC source. These results allow a more accurate monitoring of IMD incidence. The high level of MN completeness for culture confirmed MD cases has to lead to a regular follow-up in order to detect a decrease of the completeness and to take into account the part of PCR among confirmed cases.

Citation suggérée :

Berger F, Parent du Châtelet I, Bernillon P et Gallay A. Surveillance des infections invasives à méningocoque en France métropolitaine en 2005 – Évaluation quantitative par la méthode de capture-recapture à trois sources. Saint-Maurice (Fra) : Institut de veille sanitaire, août 2010, 43 p. Disponible sur : www.invs.sante.fr

INSTITUT DE VEILLE SANITAIRE

12 rue du Val d'Osne

94 415 Saint-Maurice Cedex France

Tél. : 33 (0)1 41 79 67 00

Fax : 33 (0)1 41 79 67 67

www.invs.sante.fr

ISSN : 1956-6956

ISBN-NET : 978-2-11-099278-9

Réalisé par Diadeis-Paris

Dépôt légal : août 2010