

Anne Perrocheau

Evaluation de la surveillance des infections
à **méningocoques**
en **France** en 1996
par la méthode capture-recapture



Personnes ayant participé à l'étude

**Centre National de Référence des
Méningocoques :**

Jean Michel Alonso
Martine Guibourdenche
Jean-Yves Riou

Institut de Veille Sanitaire :

Anne Claire de Benoist
Jean-Claude Desenclos
Anne Gallay
Véronique Goulet
Daniel Levy-Bruhl
Véronique Vaillant

CENTRE NATIONAL DE REFERENCE
DES MENINGOCOQUES



INSTITUT PASTEUR



INSTITUT DE
VEILLE SANITAIRE

Sommaire

	pages
1. Introduction	5
1.1 Les infections à méningocoque en France	5
1.2 Les objectifs de l'étude	5
2. Matériels et méthodes	7
2.1 La population d'étude	7
2.2 Le recueil des données	7
2.3 La définition de cas	8
2.4 La méthode capture-recapture	8
2.4.1 Principe	8
2.4.2 Les conditions d'application de la méthode	9
2.4.2.1 Conditions « implicites »	9
2.4.2.2 Conditions « statistiques »	9
2.4.3 La recherche des cas communs	9
2.4.3.1 Les variables disponibles selon les fichiers	9
2.4.4 L'identification des cas communs	10
2.4.4.1 Définition spécifique des doublons	10
2.4.4.2 Définition sensible des doublons	10
2.5 Estimation du nombre total de cas avec 2 sources	11
2.6 Dépendance entre les sources	11
2.6.1 Critères qualitatifs	11
2.6.2 Evaluation de la dépendance en croisant les sources deux à deux	11
2.6.3 Tests d'indépendances des sources avec plus de deux sources	11
2.6.4 Prise en compte de la dépendance par regroupement des sources dépendantes	11
2.7 Estimation du nombre total de cas à l'aide des modèles log-linéaires	11
2.7.1. Principe	11
2.7.2 Choix d'un modèle avec la statistique du rapport de vraisemblance G^2	12
2.7.3 Autres critères pour le choix du meilleur modèle	12
2.7.4 Estimation du nombre de cas notifiés par aucune des sources et du nombre total de cas	12
2.7.5 Identification et prise en compte des variables d'hétérogénéité de capture	13
2.8 Etude de la représentativité des sources	13
3. Résultats	15
3.1 Description	15
3.1.1 Nombre de cas déclarés	15
3.1.2 Caractéristiques des cas	15
3.2 Recherche des cas communs	16

3.2.1 Les données disponibles	16
3.2.2 L'identification des doublons	17
3.2.2.1 Selon les critères spécifiques	17
3.2.2.2 Selon les critères sensibles	17
3.2.2.3 Sensibilité et choix des critères de définition des doublons	18
3.4 L'indépendance des sources	19
3.4.1 Etude des sources deux à deux	19
3.4.2 Test d'indépendance des sources	19
3.4.3 Regroupement des sources selon les dépendances	19
3.5 Estimation du nombre total de cas avec les modèles log linéaires	19
3.6 Etude de l'hétérogénéité de capture selon les caractéristiques des cas	20
3.6.1 L'âge	20
3.6.2 Le trimestre de diagnostic	20
3.6.3 La zone géographique	20
3.6.4 Le séro groupe	20
3.7 Estimation du nombre total de cas avec les variables d'hétérogénéité de capture	20
3.8 Exhaustivité selon les sources	21
3.8.1 Résultats de l'analyse des modèles log-linéaires	21
3.8.2 Résultats des analyses stratifiées	22
3.8.3 Evolution de l'exhaustivité selon les sources	22
3.9 Etude de la représentativité des sources	23
3.9.1 La zone géographique	23
3.9.2 Le séro groupe	23
4. Discussion	25
4.1 Les principaux résultats	25
4.2 Comparaison avec les études antérieures	25
4.3 Définition de cas et validité externe	25
4.4 sensibilité de la recherche de doublons	26
4.5 Dépendance entre les sources et variables d'hétérogénéité de capture	26
4.6 Représentativité des différentes sources	27
5. Conclusion	29
6. Références	31
Annexe 1 : Fiche de déclaration, déclaration obligatoire	33
Annexe 2 : Fiche de déclaration, CNR	35
Annexe 3 : Fiche de déclaration, EPIBAC	37
Annexe 4 : Distribution géographique d'EPIBAC et taux d'exhaustivité	39
Annexe 5 : Répartition des cas de sérogroupes rares et inconnus selon les sources	41

Introduction

1.1 LES INFECTIONS À MÉNINGOCOQUE EN FRANCE

La France fait partie des pays européens présentant un taux d'incidence faible, inférieur à 1 cas pour 100 000 habitants. Le taux de létalité varie entre 8 et 10 % selon les années. Aucune épidémie, détectable par notre système de surveillance, n'a été observée depuis une vingtaine d'années.

Le nombre d'infections à méningocoque (IM) déclaré chaque année a diminué à partir des années 80 et cette tendance a persisté jusqu'en 1995, année où le taux d'incidence le plus bas jamais atteint en France a été observé, 0,5 cas pour 100 000 habitants. Depuis 1996, on observe une augmentation progressive du nombre de cas déclarés.

Les IM sont des maladies à déclaration obligatoire (DO) qui justifient la mise en œuvre rapide d'une prophylaxie dans l'entourage immédiat du cas. Il existe un Centre National de Référence des Méningocoques (CNR) qui reçoit les souches isolées chez les malades. Il existe également un réseau de surveillance des méningites bactériennes basé sur la participation volontaire d'un réseau de laboratoires hospitaliers couvrant l'ensemble du territoire français (EPIBAC).

Deux études de l'exhaustivité de la DO des infections à méningocoque ont été réalisées en comparant les souches reçues au CNR et les fiches de DO reçues à la DGS. Pour la DO l'exhaustivité était de 51 % en 1989-90 et de 56 % en 1991-92. Pour le CNR l'exhaustivité était de 53 % en 1989-90 et de 63 % en 1991-92 [1, 2]. L'évolution de la participation des cliniciens et laboratoires aux différents systèmes de surveillance est une préoccupation majeure des responsables des sys-

tèmes de surveillance. Des variations de cette participation modifient le nombre total de cas portés à la connaissance des autorités sanitaires et doivent être prises en compte lors de l'analyse des tendances de l'évolution de l'incidence de la maladie. Nous avons réalisé une nouvelle étude capture-recapture en utilisant trois sources de données différentes.

Les résultats de cette étude permettront d'évaluer l'exhaustivité de la DO, du CNR et d'EPIBAC en 1996 et la représentativité de chacune des sources. Ces informations permettront de corriger les données de surveillance pour estimer le nombre total de cas d'IM survenus sur le territoire français en 1996 et les années suivantes.

1.2 LES OBJECTIFS DE L'ÉTUDE

Général :

- Déterminer le nombre total d'infections à méningocoque prises en charge en milieu hospitalier et confirmées par isolement de *Neisseria meningitidis* dans le sang ou le LCR en France en 1996

Spécifiques :

- Évaluer l'exhaustivité des différents systèmes de surveillance : DO, CNR et EPIBAC.
- Connaître l'évolution de l'exhaustivité de la DO et du CNR depuis 1992.
- Évaluer l'indépendance des sources deux à deux.
- Identifier les facteurs influençant la représentativité de chacune des sources.
- Calculer le taux de redressement annuel pour chaque système de surveillance à prendre en compte lors de l'analyse de l'évolution de la maladie.

Matériels et méthodes

Il s'agit d'une étude rétrospective sur l'ensemble des cas d'infections à méningocoque déclarés en France métropolitaine entre le 1er janvier 1996 et le 31 décembre 1996 par l'un au moins des trois systèmes de surveillance suivant : la déclaration obligatoire, le Centre National de Référence des Méningocoques, le réseau EPIBAC.

2.1 LA POPULATION D'ÉTUDE

La population résidant en France métropolitaine en 1996.

2.2 LE RECUEIL DES DONNÉES

Le système de la déclaration obligatoire : DO

En France, les IM sont des maladies à déclaration obligatoire. La déclaration est réalisée sur un formulaire officiel par le médecin hospitalier et est adressée à la Direction Départementale des Affaires Sanitaires et Sociales (DDASS) du département de résidence du patient. La déclaration d'une IM entraîne la mise en place en urgence de mesures de prophylaxie dans l'entourage immédiat du cas selon les recommandations de la circulaire de la Direction Générale de la Santé DGS / PGE / 1, numéro 79 du 5 Février 1990 [3]. Les DDASS transmettent sur une base hebdomadaire par minitel le nombre de nouveaux cas déclarés à l'Institut de Veille Sanitaire (InVS). Elles envoient, après validation, les fiches de DO correspondantes à l'InVS. Les fiches reçues à l'InVS représentent 90 % des déclarations faites par minitel. Les 10 % restants correspondent soit à une déclaration télématique non suivie de la transmission d'une fiche ou, le plus souvent, à l'absence de confirmation biologique du diagnostic.

Les critères de déclaration reposent sur la confirmation biologique de l'infection : « l'isolement dans le sang ou le liquide céphalo-rachidien (LCR) de *Neisseria meningitidis* ou la présence d'antigènes solubles dans le sang, les urines ou le LCR ». Les informations recueillies sur la fiche de déclaration sont l'initiale du nom et le prénom du malade, sa date de naissance, son département de résidence, la date d'hospitalisation, le sérotype du méningocoque et le lieu de prélèvement, la présence d'un purpura fulminans et l'évolution de la maladie, le nombre de personnes dans l'entourage proche ayant bénéficié des mesures de prophylaxie ainsi que la nature de l'antibioprophylaxie administrée (annexe 1). Ces données sont informatisées et analysées annuellement à l'InVS. Tous les départements sont tenus de déclarer leur cas, y compris les départements d'Outre Mer.

Le CNR reçoit les souches de méningocoques adressées volontairement pour groupage et typage par les laboratoires hospitaliers. Les souches sont accompagnées d'une fiche de renseignements avec le nom, le prénom, l'âge du malade, la date et le liquide d'isolement du méningocoque, et les coordonnées du laboratoire expéditeur (annexe 2). Ces données et les résultats des analyses biologiques sont informatisées et analysées annuellement pour l'élaboration d'un rapport annuel d'activité. Il existe une politique active d'information de l'expertise disponible au CNR auprès des laboratoires hospitaliers qui a contribué à l'augmentation constante du nombre de souches envoyées depuis 1987. Les raisons déterminant l'envoi des souches sont multiples : envoi systématique, demande d'expertise pour les groupes rares, ou partenariat sur des projets de recherche.

Le Centre National de Référence des Méningocoques : CNR

Le CNR reçoit des souches de France métropolitaine, des départements et territoires d'Outre Mer et de pays d'Afrique et d'Asie avec lesquels il existe une collaboration scientifique.

Le CNR reçoit des souches de France métropolitaine, des départements et territoires d'Outre Mer et de pays d'Afrique et d'Asie avec lesquels il existe une collaboration scientifique.

Le réseau EPIBAC

C'est un réseau de laboratoires hospitaliers de microbiologie dont l'objectif est la surveillance de certaines

infections invasives communautaires. Il a pour but d'estimer l'incidence de ces infections, de suivre leur évolution dans le temps et de décrire quelques caractéristiques épidémiologiques des patients atteints par ces infections. Un cas est défini par l'isolement dans le sang ou le LCR de *Haemophilus influenza*, *Neisseria meningitidis*, *Listeria monocytogenes*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, et *Streptococcus agalactiae*. La participation des laboratoires de microbiologie à ce réseau repose sur le volontariat. Un recueil mensuel est réalisé par le laboratoire sur une fiche standard transmise à l'InVS. Pour chaque isolement sont recueillis le site et la date du prélèvement ainsi que l'âge et le sexe du patient (annexe 3). En 1996, 249 laboratoires hospitaliers ont participé sur l'ensemble de l'année à EPIBAC. Le réseau regroupait 60 % des laboratoires hospitaliers français. Les admissions dans les services hospitaliers dont la bactériologie était traitée par un laboratoire appartenant à EPIBAC représentaient 60 % des admissions en Médecine des établissements hospitaliers susceptibles de prendre en charge les pathologies étudiées. Le réseau EPIBAC couvre l'ensemble de la France mais dans 6 régions les établissements hospitaliers dépendant d'un laboratoire EPIBAC représentaient moins de 50 % des admissions en médecine de la région (annexe 4).

2.3 LA DÉFINITION DE CAS

Pour l'étude, nous avons retenu la définition de cas suivante : tout patient, hospitalisé en France métropolitaine entre le premier janvier et le 31 décembre 1996, chez lequel une souche de *Neisseria meningitidis* NM a été isolée du sang ou du LCR.

Les patients dont seule la présence d'antigènes solubles dans le sang, les urines ou le LCR avait permis de confirmer le diagnostic d'IM, et les patients hospitalisés dans les départements d'Outre Mer ont été exclus de l'étude car ils avaient une probabilité nulle d'être présents dans le système de surveillance du CNR pour les premiers et dans le système EPIBAC pour les seconds.

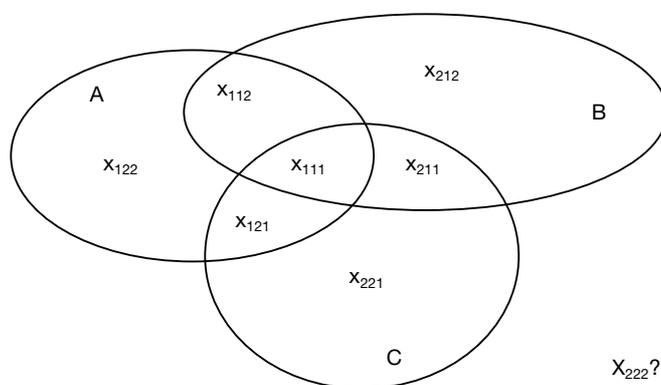
La période d'étude était comprise entre le premier janvier 1996 et le 31 décembre 1996. Tous les cas dont la date de début de maladie (pour la DO) ou la date de prélèvement (pour le CNR et EPIBAC) était comprise entre le 1er janvier et le 31 décembre 1996 et ayant fait l'objet d'une notification à l'une au moins des trois sources ont été inclus. Le prélèvement ou l'isolement du méningocoque pouvant survenir un à deux jours après la date d'hospitalisation, nous avons comparé les cas DO déclarés pendant les 8 derniers jours de 1995 et de 1996 avec les cas CNR et EPIBAC déclarés pendant les 8 premiers jours de 1996 et 1997 afin de rechercher les doublons qui auraient pu être déclarés sur des années différentes.

2.4 LA MÉTHODE CAPTURE-RECAPTURE

2.4.1 Principe

Cette méthode permet, en croisant les cas d'une maladie recensés par plusieurs systèmes de surveillance dans une population et une période définie, et après avoir identifié les cas communs entre les différents systèmes et sous certaines hypothèses, d'estimer le nombre de cas identifiés par aucun des systèmes ; l'estimation ainsi obtenue permet d'estimer le nombre de cas total de la maladie et l'exhaustivité de chaque système (figure 1).

FIGURE 1 : Répartition des cas d'une maladie recensés par trois systèmes A, B, C.



Les indices 1 et 2 correspondent à la présence (1) ou absence (2) des cas dans les sources A, B, et C dans cet ordre.

2.4.2 Les conditions d'application de la méthode [4]

2.4.2.1 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 cas identifiés appartiennent à la zone géographique et à la période étudiées. Si les zones géographiques d'où sont issus les cas sont disjointes, zone A et zone B, la probabilité de retrouver un doublon d'un cas d'une zone A dans la source B sera nulle. Si les cas sont issus de population différentes, on obtiendra un nombre total de cas qui permettra d'approcher le nombre total de cas survenus dans les deux zones A et B mais l'absence de cas communs ne permet pas d'appliquer la méthode capture-recapture.
- Si la population de la source A est un sous-ensemble de la population B, les cas communs ont une probabilité non nulle d'être identifiés ; tous les cas de la population B ont une probabilité non nulle d'être identifié dans une des sources et la méthode est valable.
- Tous les vrais cas communs et seulement les vrais cas communs sont identifiés. Une surestimation de cas communs induit une sous-estimation du nombre total de cas et inversement. En l'absence d'identifiant commun unique entre les sources, l'identification des cas repose sur une combinaison de critères. Plusieurs définitions des cas communs peuvent ainsi être proposées selon les sources considérées. Une analyse de leur sensibilité permet d'orienter le choix sur une des définitions des cas communs.

2.4.2.2 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 pour ces cas d'être identifié par l'autre système. 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 sur-estimation de N. Avec plus de deux sources, la dépendance entre les sources (le nombre maximum de sources dépendantes entre elles = nombres de sources - 1) peut être évaluée et prise en compte dans l'estimation de N grâce à l'application des modèles

log-linéaires à la méthode capture-recapture.

- L'homogénéité de capture des cas : pour une source donnée, tous les individus de la population étudiée ont la même probabilité d'identification. 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, ...). Cette probabilité peut cependant être différente suivant les sources. L'estimation de N dépend du sens de variation des probabilités de capture suivant les variables d'hétérogénéité, dans les sources. Par ailleurs la présence de variables d'hétérogénéité de capture peut induire une dépendance positive ou négative entre les sources [5].
- Les interactions entre les variables d'hétérogénéité et les systèmes 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 [5, 6, 7]. 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, i.e. il n'y a pas d'arrivée ni de départ pendant la période d'étude. Le non respect de cette condition peut induire une sous-estimation du nombre de cas communs et donc une sur-estimation de N. Cette condition est très liée à celle de l'homogénéité de capture.

2.4.3 La recherche des cas communs

2.4.3.1 Les variables disponibles selon les fichiers.

En l'absence d'un identifiant idéal commun à tous les fichiers, nous avons tenu compte des variables disponibles dans les 3 sources et de la proportion de données renseignées pour chaque variable. Nous avons retenu quatre caractéristiques permettant l'identification des cas communs : le prénom (DO et CNR), la date de naissance (DO et EPIBAC) ou l'âge (CNR), la date de début de maladie (DO) ou d'isolement (CNR) ou de prélèvement (EPIBAC), le département de la DDASS déclarant (DO) ou le département du laboratoire (CNR et EPIBAC).

L'âge :

Dans le fichier DO, l'âge était calculé par soustraction de la date de maladie - la date de naissance. Dans le fichier EPIBAC, l'âge était calculé par une soustraction année de maladie - année de naissance. Dans le fichier CNR, les âges avaient été arrondis. Pour l'analyse finale, l'âge retenu pour un doublon était celui du fichier comportant la date de naissance du cas.

La date de maladie :

Différentes dates étaient disponibles selon les sources : date de début des signes où à défaut date de la déclaration pour la DO, date d'isolement ou de prélèvement

ou à défaut date de réception de la demande d'analyse pour le CNR et EPIBAC. Pour l'analyse finale, pour les doublons avec la DO nous avons retenu la date de début de maladie, pour les doublons CNR-EPIBAC nous avons repris la date de prélèvement et à défaut d'isolement. Les cas DO de décembre 1995 communs avec des cas CNR et/ou EPIBAC de janvier 96 ont été recherchés, ainsi que les cas CNR et EPIBAC de janvier 1997 communs avec des cas DO de décembre 1996.

Le code département :

Pour la DO nous avons utilisé le code département de la DDASS déclarante sachant qu'un cas doit être déclaré dans son département de résidence. Pour EPIBAC et CNR, le département mentionné était celui de l'hôpital déclarant. Nous avons accepté un code département différent s'il correspondait à un département limitrophe : un patient peut être hospitalisé dans l'hôpital le plus proche de son domicile qui n'est pas forcément situé sur son département de résidence.

2.4.4 L'identification des cas communs

Elle a été réalisée manuellement en comparant les fichiers deux à deux.

Une première sélection a été réalisée avec des critères d'appariement très spécifiques (définition a). Puis une seconde sélection a été réalisée avec des critères plus sensibles qui permettaient de tenir compte d'erreurs ou d'imprécisions survenues lors du remplissage des fiches de déclaration ou lors de la saisie des données (définition b).

2.4.4.1 Définition spécifique des doublons (définition a)

Entre DO et CNR

La recherche de doublons entre les fichiers DO et CNR a été effectuée en utilisant une combinaison de quatre variables :

1. le prénom ; lorsque seule l'initiale du prénom était rapportée dans la DO et correspondait à l'initial du prénom du CNR, ou lorsque l'orthographe était différente mais le prénom phonétiquement identique, les prénoms ont été considérés identiques.
2. le département de la DDASS déclarant (DO) ou le département du laboratoire de l'hôpital (CNR) ; les départements devaient être les mêmes ou correspondre à deux départements limitrophes.
3. la date de début de maladie (DO) ou la date d'isolement (CNR) ; lorsque la date de maladie de la DO précédait de 1 à 3 jours celle rapportée dans le CNR les dates ont été considérées identiques.
4. l'âge du malade (calculé à partir de la date de naissance pour la DO et EPIBAC, et rapporté directement pour le CNR) en acceptant un écart maximum de 1 an.

Quand les données sur une ou plusieurs des quatre variables citées ci-dessus étaient incomplètes nous avons utilisé le sérotype (dont la valeur discriminante augmentait avec la rareté de la souche) et le sexe.

Entre DO et EPIBAC

La recherche de doublons a été réalisée en utilisant :

1. la date de naissance
2. la date de début de maladie (DO) et la date de prélèvement (EPIBAC) ; une différence de 3 jours était acceptée si la date DO précédait la date EPIBAC
3. le département de la DDASS déclarant (DO) et de l'hôpital déclarant (EPIBAC) qui devait être identique ou correspondre à des départements limitrophes.

Lorsque une ou plusieurs de ces variables étaient inconnues, nous avons utilisé l'âge, le sérotype et le sexe du malade.

Entre CNR et EPIBAC

Nous avons utilisé :

1. la date de prélèvement d'EPIBAC et la date d'isolement du CNR - qui devait être identique ou différente de moins de 3 jours.
2. l'âge du patient considéré identique si l'écart ne dépassait pas 1 an.
3. le code postal du département du laboratoire déclarant qui devait être identique.
4. le sérotype quand il était connu.
5. le sexe du malade.

Les codes laboratoire de EPIBAC ou du CNR ont été utilisés pour valider des doublons pour lesquels une partie des autres informations manquaient.

Recherche des cas communs aux trois sources

Les informations différentes contenues dans les trois sources permettaient d'identifier facilement les triplets. A partir d'un prénom commun dans DO et CNR, la date de naissance connue dans EPIBAC et DO permettait de retrouver les cas communs. Le code postal, sexe et sérotype permettait ensuite de confirmer qu'il s'agissait bien du même cas.

2.4.4.2 Définition sensible des doublons (définition b)

Un second tri a été réalisé avec les cas non sélectionnés lors du premier tri. Lorsqu'une seule variable sur les 4 variables d'identification retenues ne correspondait pas aux critères de sélection, nous avons revu les cas individuellement afin de tenir compte d'erreurs de saisies, d'imprécisions dans les dates ou d'arrondis sur l'âge.

Par exemple, lorsque la différence de date excédait trois jours, nous avons tenu compte de la similitude des trois autres variables, plus du sexe et du sérotype pour considérer le cas comme un doublon. Si la différence d'âge excédait un an, et toutes les autres variables identiques, y compris le sexe et le sérotype, nous avons considéré qu'il s'agissait d'un doublon.

De plus, lors de la mise en communs des trois fichiers, les informations complémentaires sur la date de naissance et le prénom apportées par le troisième fichier permettaient de vérifier qu'ils s'agissaient bien de cas communs.

2.5 ESTIMATION DU NOMBRE TOTAL DE CAS AVEC 2 SOURCES

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

		A		
		oui	Non	
B	oui	x_{11}	x_{21}	N_2
	non	x_{12}	x_{22}	N_1
		N_1	N_2	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_{22} est le nombre de cas présents dans aucune des sources, à estimer.

Sous l'hypothèse d'indépendance des sources (la probabilité d'être déclaré dans une source est indépendante de la probabilité d'être déclaré dans la deuxième source), les estimateurs de Sekar et Deming [8] permettent d'estimer le nombre de cas identifiés par aucune des sources (x_{22}), le nombre total de cas (N), sa variance ($Var(N)$) et son intervalle de confiance à 95 % ($IC(N)$) :

$$x_{22} = x_{12}x_{21} / x_{11} \quad N = N_1N_2 / x_{11}$$

$$Var(N) = N_1N_2x_{12}x_{21}/(x_{11})^3$$

$$IC(N)_{95\%} = N \pm 1.96 (\text{racine carrée } (Var(N)))$$

Chapman et Seber 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 [9, 10] :

$$N = (N_1 + 1)(N_2 + 1) / (x_{11} + 1) - 1$$

$$VarN = (N_1+1)(N_2+1)x_{12}x_{21} / (x_{11})^2(x_{11}+2)$$

Les taux d'exhaustivité des deux sources sont respectivement :

$$pA = N_1/N = x_{11}/N_2 \quad pB = N_2/N = x_{11}/N_1$$

L'intervalle de confiance à 95 % de chaque taux d'exhaustivité est 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.

2.6 DÉPENDANCE ENTRE LES SOURCES

2.6.1 Critères qualitatifs

Selon la nature des sources utilisées on peut anticiper l'existence de dépendance entre des sources. Ainsi des sources basées sur les mêmes interlocuteurs, médecins des laboratoires, par exemple, auront plus de

chance d'être dépendantes par rapport à des systèmes s'appuyant sur des réseaux différents.

2.6.2 Evaluation de la dépendance en croisant les sources deux à deux

Wittes propose d'estimer N pour chaque combinaison des sources deux à deux et d'évaluer la dépendance entre les sources en comparant les estimations. Si un des couples de sources donne une estimation très différente des autres couples, on peut suspecter une dépendance entre ces sources [11, 12].

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

Soient K sources S_1, S_2, \dots, S_k dans lesquelles ont été notifiés n_1, n_2, \dots, n_k cas. Soit n_{rs} le nombre de cas présents dans l'une ou les deux sources S_r et S_s , mais pas dans les autres sources. L'indépendance entre ces deux sources peut être testée avec un test du χ^2 appliqué sur un tableau de contingence 2x2 qui répartit tous les n -nrs cas notifiés au moins une fois par les sources autres que S_r et S_s dans les classes définies par la présence ou l'absence dans les sources S_r et S_s [11, 12].

Le calcul des odds ratio (OR) permet en outre d'indiquer le sens et la force des dépendances éventuelles. Si l'OR est inférieur à 1, il existe une dépendance négative entre les sources, d'où une sur-estimation de N . Si l'OR est supérieur à 1, il existe une dépendance positive entre les sources, d'où une sous-estimation de N .

2.6.4 Prise en compte de la dépendance par regroupement des sources dépendantes

Wittes a proposé d'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 2 sources fusionnées et la troisième [12].

2.7 ESTIMATION DU NOMBRE TOTAL DE CAS À L'AIDE DES MODÈLES LOG-LINÉAIRES

2.7.1. Principe

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 [13].

Dans la situation particulière de la capture-recapture, le tableau de contingence a une cellule structurellement vide correspondant à l'absence de notifications de cas dans l'ensemble des 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 [7].

Le modèle complet, avec les interactions d'ordre 2, pour trois sources A avec i niveaux, B avec j niveaux et C avec k niveaux, formant un tableau de contingence i x j x k cellules, s'écrit :

$$\ln F_{ijk} = + A_i + B_j + C_k + AB_{ij} + AC_{ik} + BC_{jk}$$

F_{ijk} est la fréquence attendue pour la cellule ijk sous le modèle considéré. A_i, B_j et C_k sont les effets principaux correspondant à la présence dans chaque source. AB_{ij}, AC_{ik} et BC_{jk} sont les termes d'interactions d'ordre 2 correspondant à la présence dans les deux sources indexées.

Les modèles log-linéaires permettent de calculer des estimations prenant en compte les dépendances entre les sources ainsi que les variables d'hétérogénéité de capture [5, 6, 7].

2.7.2 Choix d'un modèle avec la statistique du rapport de vraisemblance G²

L'ajustement des modèles aux données observées est réalisé par une procédure pas à pas descendante sur modèles emboîtés. Les critères de choix du meilleur modèle retenu ont été basés sur la statistique du rapport de vraisemblance G² qui permet :

- de comparer les données attendues F_{ijk} sous un modèle M aux données observées f_{ijk} :

$$G^2M = 2 \sum_{ijk} f_{ijk} \ln(f_{ijk}/F_{ijk})$$

- de comparer l'adéquation des deux modèles emboîtés par la différence entre les G² des deux modèles M1 et M2 :

$$G^2M1 - M2 = G^2M1 - G^2M2.$$

Le modèle retenu est celui qui contient le moins de termes d'interaction (principe de parcimonie) tout en ayant une bonne adéquation avec les données observées du tableau de contingence (G² non significatif), de façon à ce que la variance de l'estimation du nombre total de cas N soit la plus petite possible [7].

Nous avons réalisé une stratégie pas à pas descendante et choisi le meilleur modèle : chacun des termes ne peut être retiré sans entraîner une différence significative entre les deux modèles avec et sans le terme lors du test d'amélioration dont l'hypothèse nulle est : le modèle le plus simple décrit les données aussi bien que le modèle avec le terme en plus.

2.7.3 Autres critères pour le choix du meilleur modèle

D'autres critères ont été proposés pour le choix du meilleur modèle [14]. Ce sont des fonctions de la statistique de vraisemblance G² : L'Akaike Information Criterion (AIC) et le Bayesian Information Criterion (BIC). Draper [15] a proposé une petite altération au BIC, présenté ici comme le DIC. Pour chaque modèle log-linéaire, la valeur de ces critères est calculée en appliquant les formules :

$$AIC = G^2 - 2(ddl)$$

$$BIC = G^2 - (\ln Nobs)(ddl)$$

$$DIC = G^2 - (\ln(Nobs/2))(ddl)$$

Quel que soit le critère choisi, le meilleur modèle est celui pour lequel la valeur associée est la plus basse [15]. Le modèle saturé ayant un G² et un degré de liberté nul, a également un AIC et un BIC et DIC nuls. Seuls les modèles ayant un AIC et un BIC négatifs seront meilleurs que le modèle saturé.

Draper propose aussi une estimation pondérée sur le DIC du nombre total de cas à partir d'une moyenne pondérée de toutes les estimations obtenues avec chaque modèle possible :

$$NwDIC = \text{Somme} (N_i \times e^{-(DIC_i/2)}) / \text{Somme} e^{-(DIC_i/2)}$$

2.7.4 Estimation du nombre de cas notifiés par aucune des sources et du nombre total de cas

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

		A			
		oui		non	
		B		B	
C		oui	non	oui	Non
oui		X ₁₁₁	X ₁₂₁	X ₂₁₁	X ₂₂₁
non		X ₁₁₂	X ₁₂₂	X ₂₁₂	m ₂₂₂ ?

1 = présence, 2=absence des cas dans les sources A,B,C dans cet ordre.

Les valeurs estimées sous le modèle retenu, pour chaque cellule non vide ont permis d'estimer le nombre de cas (m₂₂₂) non identifiés par aucune des sources, puis celle de N et de sa variance avec les formules proposées par Bishop (tableau 1) [7].

Lorsque les valeurs de certaines cellules sont faibles ou nulles, l'estimation de N est calculée en ajoutant 1 au nombre des cellules qui apparaissent au dénominateur de l'estimateur selon la méthode proposée par Hook et regal [16].

TABLEAU 1 : Capture-recapture à trois sources, estimateurs du nombre m222 de cas identifiés par aucune source et de la variance de N, selon le nombre d'interaction entre les sources [7].

Modèle	ddl	Estimateur de m222	Variance de N
Indépendant	3	$m_{222} = N - N_{obs}$ N est la solution de l'équation $(N - N_1)(N - N_2)(N - N_3) = N^2(N - N)$	$N m_{222} / (m_{112} + m_{121} + m_{211} + m_{111})$
interaction A, B	2	$(x_{112} + x_{122} + x_{212})(x_{221}) / (x_{111} + x_{121} + x_{211})$	$(m_{222})^2 (1 / (x_{112} + x_{122} + x_{212}) + 1 / (x_{111} + x_{121} + x_{211}) + 1 / x_{221} + 1 / m_{222})$
interaction A, C	2	$(x_{121} + x_{122} + x_{221})(x_{212}) / (x_{111} + x_{112} + x_{211})$	$(m_{222})^2 (1 / (x_{121} + x_{122} + x_{221}) + 1 / (x_{111} + x_{112} + x_{211}) + 1 / x_{212} + 1 / m_{222})$
interaction B, C	2	$(x_{211} + x_{221} + x_{212})(x_{122}) / (x_{111} + x_{121} + x_{112})$	$(m_{222})^2 (1 / (x_{211} + x_{221} + x_{212}) + 1 / (x_{111} + x_{121} + x_{112}) + 1 / x_{122} + 1 / m_{222})$
interaction (A, B) et (A, C)	1	$(x_{212})(x_{221}) / x_{211}$	$(m_{222})^2 (1 / x_{212} + 1 / x_{221} + 1 / x_{211} + 1 / m_{222})$
interaction (A, B) et (B, C)	1	$(x_{122})(x_{221}) / x_{121}$	$(m_{222})^2 (1 / x_{221} + 1 / x_{122} + 1 / x_{121} + 1 / m_{222})$
interaction (A, C) et (B, C)	1	$(x_{212})(x_{122}) / x_{112}$	$(m_{222})^2 (1 / x_{212} + 1 / x_{122} + 1 / x_{112} + 1 / m_{222})$
interaction (A, B) ; (A, C) et (B, C)	0	$(x_{111})(x_{122})(x_{221})(x_{212}) / (x_{121})(x_{112})(x_{211})$	$(m_{222})^2 (1 / x_{111} + 1 / x_{121} + 1 / x_{112} + 1 / x_{122} + 1 / x_{211} + 1 / x_{221} + 1 / x_{212} + 1 / m_{222})$

ddl : degrés de liberté, N : estimation du nombre total de cas, Nobs : nombre de cas observés, m222 : nombre de cas estimés identifiés par aucune des sources, m : estimation de chaque cellule selon la présence (1) ou l'absence (2) dans chaque source.

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

A l'intérieur d'une même source, tous les cas doivent avoir la même probabilité de capture. Pour vérifier cette hypothèse, nous avons défini des sous-groupes pour les variables étudiées, par rapport à la probabilité de capture des cas.

En présence de facteurs d'hétérogénéité de capture il convient d'introduire ces facteurs dans le modèle final pour obtenir l'estimation du nombre total de cas prenant en compte l'hétérogénéité de capture pour ces variables [17].

Nous avons défini trois classes d'âge correspondant à la probabilité de développer une IM et à la probabilité d'être déclaré : les enfants de moins de 5 ans, groupe caractérisé par un taux d'incidence et de létalité élevés ; Les enfants et adolescents âgés de 5 à 19 ans représentent un groupe homogène par rapport à la déclaration à la DDASS car il s'agit d'enfants scolarisés avec mise en œuvre des mesures de prophylaxie dans la collectivité ; et les sujets âgés de 20 ans et plus où le risque de survenue des IM est faible [18].

En raison de la saisonnalité des IM, la date de survenue de la maladie a été considérée selon le trimestre de diagnostic. Le premier et le dernier trimestre de l'année correspondent à une période hivernale caractérisée par une incidence élevée des IM ; le deuxième et troisième trimestres présentent des taux d'incidence plus faibles. Le grand nombre de département et le faible nombre

de cas dans certains départements ne permettaient pas une stratification sur le département pour l'analyse avec les modèles log-linéaires. Nous avons défini 5 zones géographiques selon les zones correspondant aux indices téléphoniques : le nord-est (03), le nord-ouest (02), le sud-est (04), le sud-ouest (05), et la région île de France (01).

La zone Nord-Est comportait les régions Alsace, Bourgogne, Champagnes Ardennes, Franche Comté, Lorraine, Nord Pas de Calais et Picardie. La zone Nord-Ouest : Basse et Haute Normandie, Bretagne, Centre, Pays de la Loire. La zone sud-est : Auvergne, Languedoc Roussillon, la région PACA, et Rhône Alpes. La zone sud-ouest : Aquitaine, Limousin, Midi Pyrénées, Poitou Charente. La zone Ile de France était définie par la région Ile de France.

Pour l'analyse selon le sérotype nous avons regroupés les sérotypes rares en une catégorie incluant les sérotypes A, W135, Y, Z et E19.

2.8 ETUDE DE LA REPRÉSENTATIVITÉ DES SOURCES

La proportion de cas observée dans chaque sous-groupe a été comparée par le test du chi2 à la proportion attendue sous l'hypothèse d'un taux d'exhaustivité homogène dans chaque sous-groupe.

Résultats

3.1 DESCRIPTION

3.1.1 Nombre de cas déclarés

En 1996, 294 infections à méningocoque ont été déclarées par DO. Après exclusion de 9 cas confirmés uniquement par la présence d'antigènes solubles (6) ou déclarés dans les départements d'Outre Mer (3), l'analyse a porté sur 285 patients.

Trois cent trente neuf patients pour lesquels une ou plusieurs souches avaient été isolées du sang ou du LCR ont été enregistrés au CNR. Après exclusion de 10 cas hospitalisés dans les départements d'Outre Mer (7) ou dont le site de prélèvement et le groupe étaient inconnus (3), l'analyse a porté sur 329 patients.

Pour EPIBAC, 231 cas ont été déclarés et ont été inclus dans l'étude.

Parmi les cas DO de décembre 1995 nous avons retrouvé 2 cas communs avec 2 cas CNR et un cas commun avec CNR et EPIBAC. Nous n'avons retrouvé aucun cas CNR ou EPIBAC de janvier 1997 commun avec un cas DO de décembre 1996.

3.1.2 Caractéristiques des cas

Le sexe ratio était de 1,1 dans la DO, 1,2 dans le CNR et 1,3 dans EPIBAC (tableau 2).

La distribution par âge variait très peu d'une source à l'autre : les enfants de 0-4 ans représentaient 39 à 40 % des cas, les 5-9 ans et les 10-14 ans 8 à 12 %, les 15-19 ans, 14 à 17 %, et les adultes de 20 ans et plus 22 à 25 % des cas.

La distribution par trimestre de diagnostic était identique dans les trois sources : 27 à 32 % des cas pendant le premier et dernier trimestre ; 22 à 26 % pendant le second trimestre et 14 à 17 % pendant le troisième trimestre (tableau 3).

Dans les trois sources, les 2 zones géographiques de la moitié nord de la France présentaient le plus grand nombre de cas. Pour la DO, le nombre de cas était inférieur dans la région Ile de France, pour le CNR dans la zone géographique sud-ouest, et pour EPIBAC dans la zone géographique sud-est (tableau 4).

La proportion de cas de sérogroupes A, B, C ou rares était inférieure dans les sources DO et EPIBAC par rap-

TABLEAU 2 : Répartition des cas d'IM par sexe et groupes d'âge selon la source, France 1996.

	DO	CNR	EPIBAC
	cas (%)	cas (%)	cas (%)
Sexe			
Homme	150 (53)	181 (55)	131 (57)
Femme	135 (47)	148 (45)	99 (43)
Groupe d'âge			
0 - 4	110 (39)	128 (39)	93 (40)
5 - 9	27 (9)	37 (11)	25 (11)
10 - 14	35 (12)	31 (9)	19 (8)
15-19	49 (17)	52 (16)	32 (14)
20 et +	64 (22)	81 (25)	53 (23)
Inconnu	0 (0)	0 (0)	9 (4)
Total	285 (100)	329 (100)	231 (100)

port au CNR (tableau 5). Le diagnostic de sérotype rare (A, E19, W135, X, Y, Z) nécessitant l'expertise du CNR, les souches de sérotypes rares dans les sources DO et EPIBAC étaient de fait communes avec le CNR. Les

souches non typées (sérotype inconnu) correspondaient soit à des sérotypes rares pour lesquels la souche n'avait pas été adressée au CNR, soit à des cas pour lesquels l'information n'avait pas été reportée.

TABLEAU 3 : Répartition des cas par trimestre de diagnostic selon la source, France 1996.

	DO	CNR	EPIBAC
	Cas (%)	cas (%)	cas (%)
Premier trimestre	85 (30)	104 (32)	69 (30)
Second trimestre	62 (22)	80 (24)	60 (26)
Troisième trimestre	47 (16)	56 (17)	32 (14)
Quatrième trimestre	91 (32)	89 (27)	70 (30)
Total	285 (100)	329 (100)	231 (100)

TABLEAU 4 : Distribution géographique des cas par zone géographique selon la source, France 1996.

	DO	CNR	EPIBAC
	Cas (%)	cas (%)	cas (%)
Zone nord-est	73 (26)	89 (27)	66 (29)
Zone nord-ouest	78 (27)	86 (26)	45 (19)
Zone sud-ouest	42 (15)	43 (13)	42 (18)
Zone sud-est	49 (17)	57 (17)	33 (14)
Zone Ile de France	40 (14)	54 (17)	45 (20)
Inconnu	3 (1)	0 (0)	0 (0)
Total	285 (100)	329 (100)	231 (100)

TABLEAU 5 : Distribution des cas par sérotype selon la source, France 1996.

	DO	CNR	EPIBAC
	Cas (%)	cas (%)	cas (%)
B	175 (61)	237 (72)	132 (57)
C	51 (18)	66 (20)	33 (14)
rares	8 (3)	26 (8)	12 (5)
inconnu	51 (18)	0 (0)	54 (23)
Total	285 (100)	329 (100)	231 (100)

3.2 RECHERCHE DES CAS COMMUNS

3.2.1 Les données disponibles

Le prénom ou son initial, étaient connus pour 92 % des cas DO et 100 % des cas CNR (tableau 6). L'âge était connu pour tous les cas DO et CNR, et pour 96 % des cas EPIBAC. La date de naissance était connue pour 92 et 93 % des cas DO et EPIBAC respectivement. Le sexe était connu pour tous les cas dans les trois sources. Le code postal de la DDASS déclarante était connu dans 99 % des cas et le code

postal du malade dans 97 % pour la DO. Le code postal de l'hôpital était connu pour tous les cas CNR et EPIBAC. La date d'hospitalisation était complétée dans 100 % des cas DO, la date d'isolement de la bactérie était complétée pour 98 % des cas CNR, et la date de prélèvement était complétée dans 100 % des cas EPIBAC.

La présence de purpura fulminans a été retrouvée pour 55 % des cas DO et 100 % des cas CNR. L'évolution était complétée pour 90 % des cas DO et 33 % des cas CNR. Le sérotype de la souche responsable de la maladie était connu pour 82 % des cas DO, 100 % des cas CNR et 77 % des cas EPIBAC.

TABLEAU 6 : Variables disponibles selon les sources et pourcentage de cas renseignés.

	DO	CNR	EPIBAC
	Pourcentage de cas renseignés	Pourcentage de cas renseignés	Pourcentage de cas renseignés
Initial du nom	100	ND*	ND
Prénom	92	100	ND
Age	100	100	96
Date de naissance	92	ND	93
Sexe	100	100	100
Code postal résidence	98	ND	ND
Code postal hôpital	ND	100	100
Code postal DDASS	97	ND	ND
Ville de l'hôpital	ND	100	100
Date d'hospitalisation	100	ND	ND
Date de prélèvement	ND	ND	100
Date d'isolement	ND	98	ND
Date de déclaration	100	ND	ND
Purpura fulminans	55	100	ND
Evolution	90	33	ND
Sérogroupe	82	100	77

* ND : information non demandée

3.2.2 L'identification des doublons

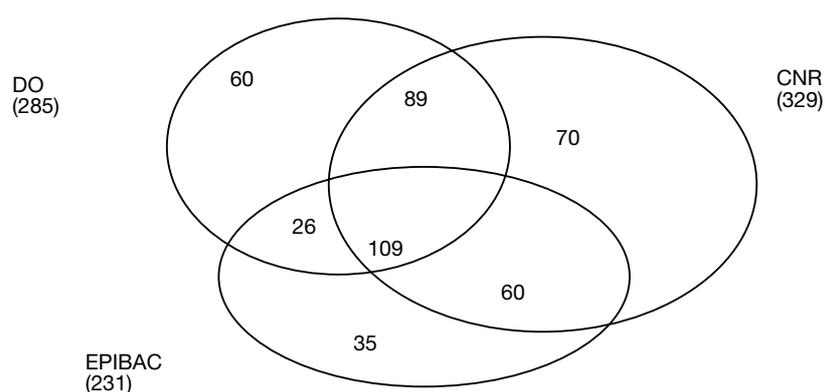
3.2.2.1 Selon les critères spécifiques : définition a

Entre les sources DO et CNR 198 doublons ont été identifiés, entre les sources DO et EPIBAC 135

doublons et entre les sources CNR et EPIBAC 169 doublons ont été identifiés (figure 2).

Au total, 109 cas étaient déclarés dans les trois sources.

FIGURE 2 : Nombre et répartition des doublons identifiés selon la définition a



3.2.2.2 Selon les critères sensibles : définition b

Entre DO et CNR

Pour 11 cas, la différence de date dépassait trois jours : 2 fois il s'agissait de différence de 4 jours et de 9 jours, 3 fois de 6 jours, 1 fois de 7, 10, 11 et 12 jours. Nous

avons tenu compte de la similitude des prénoms, pour la plupart peu communs, ainsi que de l'âge, du code postal, du sexe et du sérogroupe identiques pour considérer ces cas comme des doublons. Pour 10 cas, l'âge était différent de plus de 1 an : 4 fois la différence

était de 2 ans, 2 fois de 5 ans, et une fois de 3, 7, 9 et 10 ans. Quand l'écart dépassait 2 ans, nous avons tenu compte des prénoms identiques, de la date de maladie, du code postal, du sexe et du sérotype identiques pour considérer ces cas comme des doublons.

Un cas DO non apparié avec un cas CNR, mais apparié avec un cas EPIBAC s'est révélé par la suite être apparié avec un doublon CNR-EPIBAC.

Au total, 21 nouveaux doublons ont été identifiés au second tri. Il y avait 220 cas communs aux fichiers DO et CNR.

Entre DO et EPIBAC

Pour 6 cas, les dates de naissance étaient différentes d'un chiffre alors que les autres variables d'identification des doublons étaient identiques. Pour 4 cas, la date d'hospitalisation précédait de plus de 3 jours la date de prélèvement, ces cas avaient la même date de naissance, le même code postal, le même sexe et le même

sérotipe. Nous avons considéré ces 10 cas comme des doublons.

Au total il y avait 145 cas communs aux fichiers DO et EPIBAC.

Entre EPIBAC et CNR

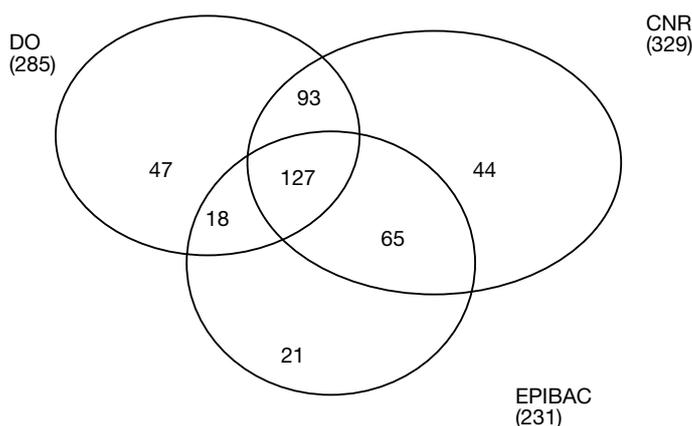
Pour 14 cas la date de début était différente de plus de 3 jours, alors que les prénoms, le code postal, l'âge, le sexe et le sérotipe étaient identiques.

Pour 4 cas l'âge était différent de plus de 1 an et l'ensemble des autres variables étaient identiques. Nous avons considéré ces 18 cas comme des doublons.

La mise en commun des trois fichiers a permis d'identifier 5 autres doublons pour lesquels les informations manquantes dans EPIBAC ou CNR étaient complétées par les informations de la DO ce qui permettait de reconnaître qu'il s'agissait du même cas.

Au total il y avait 192 cas communs aux fichiers CNR et EPIBAC (figure 3).

FIGURE 3 : Nombre et répartition des doublons selon la définition b.



3.2.2.3 Sensibilité et choix des critères de définition des doublons

Le nombre total de cas déclarés par au moins une des trois sources en 1996 variait de 415 à 452 selon la définition des doublons utilisée (tableau 7).

La définition b était plus sensible et moins spécifique que la définition a, mais aussi plus complexe et moins reproductible. Cependant, la définition a reposant sur des critères très spécifiques de recherche des dou-

blons, le risque de faux positif (cas différents dans deux sources identifiés comme doublon) était faible et le risque de faux négatif (cas commun à deux sources non reconnus comme doublon) élevé. Avec la définition b, l'élargissement des critères de définition permettait de réduire le risque de faux négatif mais risquait d'augmenter le nombre de faux positif. L'étude individuelle des cas suspectés identiques entre deux sources et la spécificité des variables utilisées pour l'appariement,

TABLEAU 7 : Nombre de cas observés communs à deux et trois sources selon les définitions.

Sélection	Doublons			Cas présents dans les trois sources	Total cas observés
	DO-CNR	CNR-EPIBAC	DO-EPIBAC	DO-CNR-EPIBAC	
Définition a	198	169	135	109	452
Définition b	220	192	145	127	415

telles que la date de naissance, le prénom s'il s'agissait d'un prénom rare, ou la date de début des signes associée à l'hôpital ou au département, permettait de réduire le risque de faux positif. Compte tenu de la qualité des données disponibles dans les sources, nous avons considéré que la définition b permettait d'identifier plus précisément les vrais doublons sans trop augmenter le risque de faux positif et nous avons conduit l'analyse à partir des doublons identifiés selon cette définition.

3.4 L'INDÉPENDANCE DES SOURCES

3.4.1 Etude des sources deux à deux.

L'estimation du nombre total de cas déclarés ou non déclarés était de 396 en croisant les fichiers CNR et EPIBAC, 426 en croisant les fichiers DO et CNR, et 453 en croisant les fichiers DO et EPIBAC (tableau 8).

TABLEAU 8 : Estimation du nombre total de cas obtenu en croisant les sources 2 à 2, exhaustivité des sources et intervalle de confiance à 95 %, France, 1996.

Source 1, source 2	n1	n2	n12	n	N	x	Exh 1	IC 95 %	Exh 2	IC 95 %
DO, CNR	285	329	220	394	426	32	67 %	65-69	77 %	75-80
DO, EPIBAC	285	231	145	371	453	82	63 %	59-68	51 %	48-55
CNR, EPIBAC	329	231	192	368	396	28	83 %	80-86	58 %	56-61

L'intervalle de confiance à 95 % de l'estimation obtenue avec les sources CNR et EPIBAC (381-411) ne se superposait pas, sauf pour la borne supérieure, avec les intervalles de confiance des deux autres estimations (411-441 et 422-484). Ce résultat, ajouté au plus faible nombre de cas obtenu en croisant les fichiers CNR et EPIBAC, indiquait une dépendance positive entre les sources CNR et EPIBAC.

3.4.2 Test d'indépendance des sources

Une dépendance positive entre DO et CNR était confirmée par un odds ratio de 2,28, intervalle de confiance excluant 1, calculé pour les cas présents dans EPIBAC répartis dans les sources DO et CNR (tableau 9).

Une dépendance positive entre CNR et EPIBAC était confirmée par un odds ratio de 3,57, intervalle de confiance excluant 1, calculé pour les cas présents dans DO et répartis dans les sources CNR et EPIBAC.

L'absence de dépendance entre DO et EPIBAC était confirmée par l'odds ratio de 0,92, intervalle de confiance incluant 1, calculé pour les cas présents dans CNR et répartis dans les sources DO et EPIBAC.

3.4.3 Regroupement des sources selon les dépendances

Les dépendances existant entre le CNR et les deux autres sources, DO et EPIBAC ne permettaient pas le regroupement des sources selon les dépendances. En effet, en regroupant DO avec CNR la dépendance EPIBAC-CNR persistait, et en regroupant CNR avec EPIBAC, la dépendance DO-CNR persistait. L'estimation du nombre total de cas par la méthode de Sekar et Demings étaient impossibles à réaliser [8].

3.5 ESTIMATION DU NOMBRE TOTAL DE CAS AVEC LES MODÈLES LOG LINÉAIRES

L'analyse pas à pas descendante à partir du modèle saturé prenant en compte toutes les interactions d'ordre 2 entre les 3 sources, sans les variables de stratification, montrait qu'en dehors du modèle saturé, un seul modèle

TABLEAU 9 : Test d'indépendance des sources deux à deux, méthode de Wittes, infection à méningocoque, France, 1996.

Source 1, source 2 / source 3	OR	IC 95 %
DO, CNR / EPIBAC	2,28	1,08 – 4,84
DO, EPIBAC / CNR	0,92	0,56 – 1,51
EPIBAC, CNR / DO	3,57	1,87 – 6,84

le présentait une bonne adéquation avec les données. Ce modèle incluait les deux dépendances observées précédemment : entre les sources DO et CNR et entre CNR et EPIBAC (modèle numéro 4) (tableau 10).

L'estimation du nombre total de cas d'IM confirmés par isolement de NM dans le sang ou le LCR en France, déclarés par au moins une des trois sources ou non déclarés était de 470 cas avec ce modèle.

TABLEAU 10 : Analyse log-linéaire et modèles, estimation du nombre de cas d'infection à méningocoque et IC 95 %, AIC et BIC, France, 1996.

Numéro	Modèles	N	IC 95 %	G ²	ddl	p	AIC	BIC
1	DC, CE, DE	466	421-511	0	0	1	0	0
2	DE, CE	437	421-453	5,38	1	0,02	3,38	1,19
3	DC, DE	429	418-440	18,6	1	0	16,59	14,4
4	DC, CE	470	430-510	0,11	1	0,74	-1,89	-4,08
5	DC, E	433	421-445	21	2	0	17,04	12,7
5	DE, C	428	419-436	18,7	2	0	14,69	10,3
7	D, CE	441	427-455	6,37	2	0,04	2,37	2,01
8	D, C, E	431	424-438	21,5	3	0,00	15,5	8,93

Estimation de N pondéré selon le BIC = 461

N : nombre de cas total estimés, G² : test du rapport de vraisemblance, ddl : degré de liberté, p : adéquation du modèle, D : DO, C : CNR, E : EPIBAC, AIC : Akaike information criteria, BIC : Bayesian information criteria

3.6 ETUDE DE L'HÉTÉROGÉNÉITÉ DE CAPTURE SELON LES CARACTÉRISTIQUES DES CAS

3.6.1 L'âge

L'introduction de l'âge en 3 groupes (< 5 ans, 5-19 ans, 20 ans et +) dans le modèle saturé montrait par une stratégie pas à pas descendante que le meilleur modèle retenu (plus petit BIC et AIC) incluait deux termes d'interaction pour les dépendances entre DO et CNR et entre CNR et EPIBAC et l'âge seul. Il n'existait pas d'interaction entre l'âge et les trois sources. La probabilité d'être capturé dans chacune des sources ne variait pas selon le groupe d'âge.

3.6.2 Le trimestre de diagnostic

L'analyse des modèles log-linéaires incluant le trimestre de diagnostic et les trois sources montrait que le meilleur modèle retenu (plus petit BIC et AIC) incluait deux termes d'interaction pour les dépendances entre DO et CNR et entre CNR et EPIBAC et le trimestre seul et ne mettait pas en évidence d'interaction entre le trimestre et chacune des trois sources. La probabilité d'être capturé dans chacune des sources ne variait pas selon le trimestre de diagnostic.

3.6.3 La zone géographique

Après introduction de la zone géographique découpée selon 5 zones avec les trois sources, le meilleur modèle

le retenu incluait un terme d'interaction entre zone géographique et EPIBAC. Il existait une probabilité différente d'être capturé par le système EPIBAC selon la zone géographique d'hospitalisation du cas.

3.6.4 Le sérotype

L'introduction du sérotype, divisé en trois sous-groupes B, C et sérotypes rares ou inconnus, avec les trois sources montrait, par une stratégie pas à pas descendante, qu'il existait une interaction entre groupe et CNR et entre groupe et DO. L'interaction entre CNR et DO disparaissait.

3.7 ESTIMATION DU NOMBRE TOTAL DE CAS AVEC LES VARIABLES D'HÉTÉROGÉNÉITÉ DE CAPTURE

Selon les résultats de l'analyse précédente, la zone géographique (R) pour EPIBAC et le sérotype (G) pour la DO et CNR étaient des facteurs qui modifiaient la probabilité de capture d'un cas dans la source considérée et il convenait de réaliser une analyse stratifiée sur ces variables.

Par une stratégie pas à pas descendante à partir du modèle saturé (modèle 1) incluant les trois sources de déclaration, le sérotype et la zone géographique on obtenait un modèle qui présentait une bonne adéqua-

tion avec les données tout en prenant en compte les interactions entre les sources et le séro groupe et entre la zone géographique et EPIBAC : ce modèle incluait la dépendance CNR*EPIBAC et trois termes d'interaction : CNR*groupe, DO*groupe et EPIBAC*zone géographique et avait le plus petit BIC et AIC (modèle 4) (tableau 11).

Avec ce modèle, le nombre total de cas identifiés par aucune des sources était de 45, et l'estimation du nombre total de cas survenus en 1996 était de 460. L'introduction du séro groupe faisait disparaître l'interaction entre les sources DO et CNR.

3.8 EXHAUSTIVITÉ SELON LES SOURCES

3.8.1 Résultats de l'analyse des modèles log-linéaires

En l'absence des variables de stratification, le taux d'exhaustivité de la DO était de 61 % (285 cas déclarés / 470 cas estimés), celui du CNR de 61 % (329 / 470), et celui du système EPIBAC de 49 % (231/470) (tableau 12).

Après introduction des variables d'hétérogénéité de capture les taux d'exhaustivité des trois sources augmentaient de manière non significative.

TABLEAU 11 : Analyse log-linéaire avec les variables d'hétérogénéité de capture, modèles, tests d'adéquation, AIC et BIC, France 1996.

Numéro	Modèles	N	IC95 %	G ²	ddl	p	AIC	BIC
1	DC,CE,DE,GD,GC,GE,RE,RC,RD	481	424-540	48,41	74	0,99	-100	-262
2	CE, GD,GC,RD,RE	461	453-469	62,13	82	0,95	-102	-281
3	CE, GC, GE, RD, RE	447	441-453	78,05	82	0,6	-86	-266
4	CE, GD, GC, RE	460	452-468	70,34	86	0,88	-102	-290
5	CE, GD, GC, R	461	453-469	91,35	90	0,44	-89	-286
6	CE, D, GC, RE	447	441-453	87,5	88	0,49	-89	-281
7	CE, GD, RE	444	439-449	129,7	88	0	-46	-239
8	GD, GC, RE	442	434-450	86,57	87	0,49	-87	-278
9	CE,GD, GC	460	452-468	110,8	94	0,11	-77	-283
10	CE,D,GC,R	447	441-453	108,5	92	0,11	-75	-277
11	E,GD,GC,R	442	434-450	107,6	91	0,11	-74	-274
12	CE,D,GC	447	441-453	128	96	0,016	-64	-274
13	E,GD,GC,R	442	434-450	127	95	0,01	-63	-271

N : nombre de cas total estimés, G² : test du rapport de vraisemblance, ddl : degré de liberté, p : adéquation du modèle, D : DO, C : CNR, E : EPIBAC, G : groupe, R : zone géographique, AIC : Akaike information criteria, BIC : Bayesian information criteria

TABLEAU 12 : Estimation du nombre total de cas d'IM confirmés et exhaustivité globale et stratifiée sur le séro groupe et la zone géographique des sources avec les modèles retenus, France, 1996.

	x	N (IC 95 %)	DO	CNR Exhaustivité en % (IC95 %)	EPIBAC
Total brut	55	470	61	70	49
Modèle DC, CE		(430-510)	(56-66)	(65-77)	(45-54)
Total après stratification	45	460	62	72	50
Modèle CE, DG, CG, RE		(452-468)	(61-63)	(70-73)	(49-51)

X : estimation du nombre de cas déclarés dans aucune des sources, N : nombre total de cas estimés.

3.8.2 Résultats des analyses stratifiées

Selon l'analyse stratifiée sur chacune des variables d'hétérogénéité de capture, on observait pour le séro-groupe que dans toutes les sources l'exhaustivité était meilleure pour les cas de séro-groupe C, puis pour les

cas de séro-groupe B. Elle était inférieure à 40 % dans les trois sources pour les cas de sérogroupes inconnus ou rares. Deux zones géographiques présentaient une exhaustivité inférieure à 40 % pour EPIBAC, le nord-ouest et le sud-est (tableau 13).

TABLEAU 13 : Estimation du nombre total de cas d'IM confirmés et exhaustivité après stratification sur le séro-groupe et la zone géographique des sources avec les modèles retenus, France, 1996.

Séro-groupe Modèle : CE, C*groupe, D*groupe	x	N (IC95 %)	Exhaustivité en %		
			DO	CNR	EPIBAC
B	12	292 (289-295)	67	82	55
C	3	78 (76-80)	73	84	56
Rare ou inconnu	29	89 (72-106)	36	27	30
Total stratifié	44	459 (437-481)	62	72	50
Zone géographique Modèle : zone* E, DC, CE					
Zone nord-est	13	121 (102-140)	60	73	54
Zone nord-ouest	17	118 (90-146)	66	73	37
Zone sud-ouest	4	62 (54-70)	71	69	69
Zone sud-est	14	94 (74-114)	53	62	36
Zone Ile de France	15	75 (61-89)	53	72	60
Total	55	470 (381-559)	61	70	49

3.8.3 Evolution de l'exhaustivité selon les sources

L'exhaustivité de la DO a augmenté de manière non significative entre 1989-1990 et 1991-1992, ($p=0,11$), et de manière significative entre 1991-1992 et 1996, $p=0,02$ (tableau 14) [1,2]. Ces études avaient été réalisées à partir des deux sources DO et CNR. L'étude de 1989-1990 avait exploré l'existence de dépendance entre les sources par la méthode du calcul du coefficient de corrélation des taux d'exhaustivité après stratification sur les départements [8]. Aucune dépendance

entre les sources DO et CNR n'avait été mise en évidence.

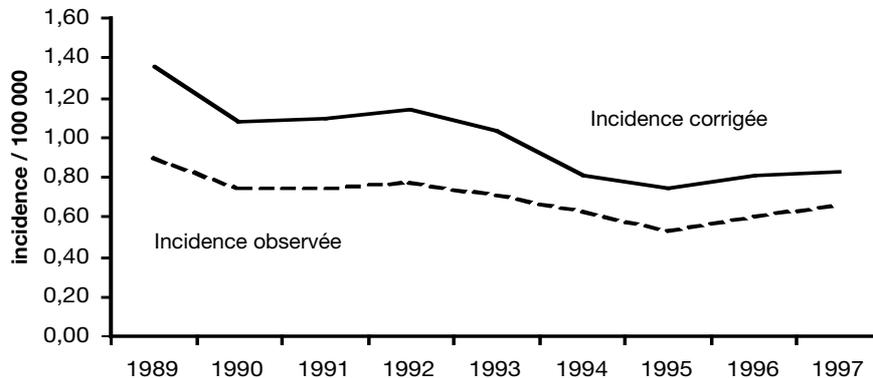
Pour le CNR, on observe des augmentations significatives entre 1989-1990 et 1991-1992 ($p<0,001$), et entre 1991-1992 et 1996 ($p<0,001$).

A partir des résultats des différentes études, nous avons calculé les incidences pour chaque année en prenant en compte de la sous-déclaration rapportée l'année de l'étude et en extrapolant aux années suivantes. Les courbes des taux d'incidence, observée et corrigée pour la sous-déclaration, se rapprochent avec le temps (figure 4).

TABLEAU 14 : Exhaustivités des sources DO et CNR et leur évolution depuis 1989 (1-2).

Année	Exhaustivité en % (IC à 95 %)	
	DO	CNR
1989-1990	53 (50-56)	55 (52-58)
1991-1992	56 (54-58)	63 (61-66)
1996	62 (61-63)	72 (70-73)

FIGURE 4 : Taux d'incidence observé et taux d'incidence corrigé pour la sous-déclaration des infections à méningocoques, 1989-97, France, déclaration obligatoire.



3.9 ETUDE DE LA REPRÉSENTATIVITÉ DES SOURCES

3.9.1 La zone géographique

Les cas déclarés par DO et par le CNR étaient représentatifs par zones géographiques de l'ensemble des cas estimés (tableau 15). Pour EPIBAC, on observait que dans le nord-ouest et le sud-est les cas étaient sous-déclarés alors que dans le sud-ouest les cas

étaient sur-déclarés ; la différence était significative pour le sud-ouest ($p = 0,03$).

3.9.2 Le sérotype

Le système de la DO sous-représentait les cas de sérotype rare ou inconnu ($p=0,01$) (tableau 16). Le CNR sur-représentait les cas de sérotype B ($p = 0,01$) et sous-représentait les cas de sérotype rare ou inconnu ($p=0,001$). EPIBAC sous-représentait les cas de sérotype rare ou inconnu ($p=0,04$). Au total, les cas de sérotype rare ou inconnu étaient sous-représentés dans les trois sources

TABLEAU 15 : Représentativité de la DO, du CNR et de EPIBAC selon la zone géographique, IM, France 1996.

Zone géographique	Proportion estimée %	Proportion déclarée (%) et p					
		DO		CNR		EPIBAC	
		%	p	%	p	%	p
Nord-est	26	26	NS	27	NS	28	NS
Nord-ouest	25	27	NS	26	NS	19	NS
Sud-ouest	13	15	NS	13	NS	19	0,03
Sud-est	20	18	NS	18	NS	15	NS
Ile de France	16	14	NS	16	NS	19	NS

TABLEAU 16 : Représentativité de la DO, du CNR et de EPIBAC selon le sérotype, IM, France 1996.

Sérotype	Proportion estimée %	Proportion déclarée (%) et p					
		DO		CNR		EPIBAC	
		%	p	%	p	%	p
B	64	69	NS	73	0,01	70	NS
C	16	20	NS	20	NS	19	NS
Rare ou inconnu	18	11	0,01	7	0,001	11	0,04

Discussion

4.1 LES PRINCIPAUX RÉSULTATS

Le nombre total de cas estimés d'infections à méningocoque confirmés par l'isolement de la souche en France métropolitaine en 1996, était de 460 cas avec un intervalle de confiance variant de 438 à 482 cas. En 1996, l'exhaustivité des systèmes de surveillance était meilleure pour le CNR 72 %, puis la DO 62 % et enfin pour EPIBAC 50 %. Au total les trois systèmes permettaient d'identifier 90 % des cas d'IM confirmés biologiquement.

L'analyse montrait qu'il existait une dépendance entre les systèmes CNR et EPIBAC, que les cas de séro-groupe rare ou inconnu étaient sous-estimés dans les trois systèmes, que les cas de séro-groupe B étaient sur-représentés par le CNR, et que les cas survenus dans la zone sud-ouest étaient sur-représentés par EPIBAC. Les trois sources étaient représentatives des cas par groupe d'âge et par trimestre de diagnostic. L'analyse stratifiée sur les variables d'hétérogénéité de capture, le séro-groupe et la zone géographique, modifiait très peu l'estimation du nombre total de cas survenus en 1996 et des taux d'exhaustivité obtenus en l'absence de stratification.

4.2 COMPARAISON AVEC LES ÉTUDES ANTÉRIEURES

L'amélioration de l'exhaustivité des sources DO et CNR depuis 1989 est liée à une meilleure participation des cliniciens et des biologistes aux systèmes de surveillance. Le méningocoque est une maladie parmi les mieux déclarées en France car des mesures de prophylaxie doivent être appliquées en urgence dès le diagnostic d'un cas pour prévenir la survenue d'autres cas parmi les contacts du malade. Il s'agit aussi d'une maladie qui a toujours été largement relayée dans la presse. Depuis 1998, cette tendance s'accroît en particu-

liers pour les décès et les cas groupés, confirmés ou non. Ce fait devrait permettre d'augmenter encore la sensibilisation des cliniciens à la déclaration des IM.

La campagne active du CNR au début des années 1990 pour proposer l'expertise des souches de méningocoques isolées en France a, sans aucun doute, fortement contribué à l'amélioration de l'exhaustivité du CNR.

4.3 DÉFINITION DE CAS ET VALIDITÉ EXTERNE

Notre étude a estimé qu'en 1996, 38 % des cas d'IM pour lesquels une souche avait été isolée du sang ou du LCR n'étaient pas déclarés par le système de la DO. Certains cas déclarés à la DO, confirmés uniquement par la présence d'antigènes solubles ou suspectés par la présence d'un purpura fulminans, avaient été exclus de l'étude selon la condition implicite n°1 de la méthode capture-recapture : la présence de cas définis selon des critères non admis dans les autres sources entraînerait une surestimation du nombre total de cas et une sous estimation de l'exhaustivité des autres sources [19, 20].

Les taux d'exhaustivité calculés ne concernent que les cas confirmés par culture du méningocoque en 1996. L'analyse réalisée ne nous permet pas de connaître le nombre d'IM pour lesquels l'isolement de la souche a été impossible, soit parce que l'état hémodynamique trop précaire ou le décès précoce du patient n'ont pas permis de réaliser les prélèvements bactériologiques nécessaires, soit parce que le méningocoque, très sensible aux conditions de transport et aux antibiotiques, n'a pas poussé en culture. Au Canada, les cas sans confirmation biologique représentaient 10 % des cas déclarés en 1997 et 1998 [21]. En Mars 2000, des nouveaux critères de déclaration des cas d'IM, élargissant la définition aux formes cliniques sans isolement de souche, ont été adoptés et diffusés aux DDASS et aux

médecins. Cela permettra d'inclure l'ensemble des cas cliniques ou confirmés biologiquement et de mieux connaître le nombre d'IM survenant en France, en particulier les formes rapidement fatales.

En 1996, le réseau EPIBAC est représentatif de 60 % de l'ensemble des admissions dans les services prenant en charge les IM [22]. Notre étude montre que le réseau EPIBAC déclarait 50 % des cas d'IM confirmés biologiquement en 1996, avec de fortes variations selon les zones géographiques, de 36 % dans le sud-est à 69 % dans le sud-ouest. La représentativité d'EPIBAC par zones géographiques était correcte sauf pour la zone sud-ouest qui sur-représentait les cas d'IM.

La représentativité d'EPIBAC calculée à partir des admissions dans les services hospitaliers couverts par un laboratoire participant à EPIBAC, ne peut pas être directement comparée à l'exhaustivité de la déclaration des cas d'infections à méningocoques calculée dans notre étude pour l'ensemble des laboratoires français. Une étude de l'exhaustivité de l'ensemble des laboratoires participants à EPIBAC, en comparant uniquement les cas DO et CNR hospitalisés dans les services couverts par les laboratoires participants à EPIBAC, permettrait de savoir si tous les cas hospitalisés dans ces services sont bien déclarés à EPIBAC. Cependant nos résultats corroborent le fait que EPIBAC sous-estime les cas d'IM survenus en France en 1996 et indiquent que le redressement du nombre total de cas survenus en France à partir des admissions dans les services couverts par EPIBAC n'est pas très éloigné de la réalité.

4.4 SENSIBILITÉ DE LA RECHERCHE DE DOUBLONS

Selon les critères de sélection des doublons utilisés, le nombre de cas déclarés dans les trois sources variait de 452 cas avec la définition a, à 415 cas avec la définition b. Le nombre de cas total estimé à l'aide des modèles log-linéaires variait de 536 (IC 95 % : 485-587) pour la définition a, à 470 (IC95 % : 430-510) pour la définition b, soit une réduction significative de 14 % du nombre total de cas estimés.

La présence de faux-positifs et de faux-négatifs peut entraîner des variations importantes de l'estimation du nombre total de cas [23]. Dans notre étude, avec le premier tri, on excluait beaucoup de doublons pour lesquels une analyse plus fine prenant en compte les erreurs de transcription et les variations de calculs de l'âge, indiquaient qu'ils étaient des vrais doublons. Les variables utilisées pour l'identification des cas communs telles que le prénom, la date de naissance et la date de diagnostic, étaient très spécifiques des cas et rendaient le risque de faux positifs faible. Brenner a démontré qu'en l'absence d'identifiant commun à l'ensemble des fichiers, les faux négatifs et les faux positifs

surviennent en même temps et que leurs effets antagonistes sur l'estimation du nombre total de cas doivent s'annuler en partie ou complètement [24].

Un des objectifs de l'étude était de décrire l'évolution de l'exhaustivité de la DO et du CNR. Les résultats obtenus devaient être comparés aux études réalisées en 1991 et en 1993 dans lesquelles les critères de sélection des doublons étaient moins spécifiques : la différence d'âge acceptée entre deux cas communs était de 2 ans, et seuls les mois de survenus des cas devaient être compatibles. En sélectionnant les cas communs selon la sélection a, on n'observait aucune amélioration de l'exhaustivité de la DO et du CNR depuis 1990. Ce résultat contrastait avec l'augmentation objective depuis 1990 du nombre de laboratoires qui envoyaient leurs souches au CNR, ce qui laissait supposer que l'exhaustivité du CNR avait augmentée depuis 1990. Avec la sélection b, dont les critères de sélection des cas communs étaient plus proches de ceux des études antérieures tout en restant plus spécifiques, nos résultats étaient compatibles avec l'évolution de la participation des laboratoires à la surveillance des IM.

4.5 DÉPENDANCE ENTRE LES SOURCES ET VARIABLES D'HÉTÉROGÉNÉITÉ DE CAPTURE

L'analyse log-linéaire appliquée à la capture-recapture ne peut tester les interactions d'ordre 3 entre les trois sources [25]. Nous avons montré qu'en l'absence des facteurs d'hétérogénéité de capture il existait une dépendance d'une part entre les sources DO et CNR et d'autre part entre CNR et EPIBAC, mais pas entre DO et EPIBAC. La dépendance entre DO et CNR disparaissait lorsque l'on introduisait le séro groupe comme variable de stratification dans le modèle log-linéaire. Les cas de séro groupe rares ou inconnus étaient responsables de cette dépendance. Avant la mise en commun des trois fichiers, le séro groupe était inconnu pour 23 % (54) et 18 % (51) de cas EPIBAC et DO respectivement. Après l'identification des doublons nous avons complété les informations manquantes dans une source à partir des informations de l'autre source. Ainsi 74 % (40) des cas EPIBAC et 59 % (30) des cas DO dont le séro groupe était inconnu avaient un doublon de séro groupe connu, 2 cas communs DO-EPIBAC restaient de séro groupe inconnu, et les cas sans doublon (14 cas EPIBAC et 21 cas DO) restaient de séro groupe inconnu. Il n'y avait pas de cas de séro groupe inconnu dans le fichier CNR. Au total, après la mise en commun des fichiers, un cas de séro groupe inconnu était toujours, sauf pour deux doublons DO-EPIBAC, un cas isolé. Il existait une dépendance négative entre les 3 fichiers deux à deux pour les cas de séro groupe inconnu (annexe 5).

Les cas de sérogroupes rares (A, W135, Y, E19) nécessitaient l'expertise du CNR pour être identifiés comme tels. Sur les 27 cas de séro-groupe rare, 3 étaient des triplets, 8 étaient des doublons CNR-EPIBAC, 6 DO-CNR, 1 DO-EPI, 7 étaient isolés CNR, 1 était isolé EPI-BAC, 1 était isolé DO. Au total, 8 cas de séro-groupe rare n'avait pas de correspondant CNR ; il peut s'agir de cas identifié par les antigènes solubles et non signalés comme tels dans les fichiers ou de cas pour lesquels les informations disponibles dans les fichiers n'ont pas permis de retrouver le doublon correspondant. Malgré ces 8 cas, il existait une dépendance positive entre le CNR et la DO et entre le CNR et EPIBAC pour les séro-groupes rares (annexe 5).

A cause du recodage des doublons dont un était de séro-groupe inconnu avec l'information du doublon correspondant, nous avons artificiellement engendré ces dépendances. L'interprétation des résultats est donc difficile compte tenu de ce biais d'information.

La dépendance CNR-EPIBAC, qui persistait après l'introduction du séro-groupe comme variable de stratification, était liée en partie à la dépendance positive observée pour les cas de séro-groupe rares et à la dépendance négative liée aux cas de séro-groupe inconnu recodés ou non selon l'existence d'un doublon. Cette dépendance apparaissait dans tous les groupes d'âge, pour les sérogroupes B et C, pour toutes les zones géographiques et pour tous les trimestres. Les systèmes de surveillance CNR et EPIBAC reposent sur les laboratoires hospitaliers et on peut extrapoler que le fait de participer à un des systèmes de surveillance était lié positivement à la participation à l'autre système.

L'interaction EPIBAC*zone géographique fait apparaître le fait que certaines zones géographiques déclarent moins bien leurs cas que d'autres et nous avons mis en évidence une sur-représentation des cas survenant dans la zone sud-ouest. Les modalités de déclaration sont indépendantes d'un département à l'autre et d'un hôpital à l'autre, voire d'un service hospitalier à l'autre, notre regroupement par larges zones géographiques correspondantes aux zones de numérotation télépho-

nique était donc très imprécise. Cependant, les zones qui présentaient une mauvaise exhaustivité étaient celles pour lesquelles le taux de participation des laboratoires hospitaliers par rapport au nombre d'entrées dans les services de médecine couverts était faible.

4.6 REPRÉSENTATIVITÉ DES DIFFÉRENTES SOURCES

Les trois sources étaient représentatives pour l'âge divisé en trois groupes. Dans l'étude de 1990, les sujets de 5 à 24 ans avaient plus de chance d'être déclarés dans la DO que les autres classes d'âge. Nous n'avons pas retrouvé de différence de représentativité liée à l'âge pour les trois sources.

Les différences de représentativité selon les zones géographiques pour EPIBAC étaient en accord avec la différence de couverture du réseau EPIBAC selon les régions : dans la zone sud-est comme dans la zone nord-ouest les hôpitaux participant au réseau EPIBAC représentent, suivant les régions, de 26 à 75 % des admissions dans les services de médecine prenant en charge les infections à méningocoques (annexe 4). Une analyse par département était impossible vu le nombre de département et le faible nombre de cas dans la majorité des départements : trop de cellules vides ne permettait pas des estimations valides des effectifs manquants [26].

Les cas de séro-groupe rare ou inconnu représentaient 18 % du nombre total estimé de cas en 1996 et 14 % des cas observés après mise en commun des trois sources et élimination des doublons ($p=0,09$). Avant appariement les cas rares ou inconnus représentaient 21 % des cas DO, 8 % des cas CNR et 29 % des cas EPIBAC et après appariement, 11 %, 7 % et 12 % respectivement. Il existait un problème de report de l'information sur les sérogroupes dans EPIBAC et DO qui confirme la spécificité du CNR dans l'analyse bactériologique des souches qui lui sont envoyées.

Conclusion

En 1996, plus d'un tiers des infections à méningocoques survenues en France métropolitaine n'ont pas été notifiées au système de surveillance nationale de la déclaration obligatoire. Les résultats de notre étude estiment que 460 cas (IC à 95 % 438-482) d'IM confirmés par isolement de la souche sont survenus en 1996. Il s'agit d'une estimation à minima de l'importance des IM en France, car l'étude a été réalisée à partir des cas confirmés biologiquement uniquement.

Les systèmes de surveillance EPIBAC et CNR présentaient une dépendance liée au fait que ces deux réseaux s'appuient sur les laboratoires hospitaliers qui participent volontairement à la surveillance des IM. Cette dépendance était renforcée par la présence des cas de sérogroupe rares ou inconnus.

La méthode de capture-recapture peut être appliquée

pour estimer les cas non déclarés à travers les systèmes de surveillance passifs. Cette méthode permet d'évaluer les systèmes de surveillance passifs sans augmenter le coût ni les efforts nécessaires pour un recueil exhaustif des cas. Les méthodes de capture-recapture offrent un complément aux systèmes de surveillance passive en permettant de corriger les taux d'incidence observés en fonction de la sous-déclaration calculée.

Recommandations :

L'analyse de l'évolution de l'incidence doit toujours tenir compte de l'exhaustivité, et de son évolution, du système de surveillance.

Une sensibilisation des médecins à la déclaration obligatoire est nécessaire afin d'améliorer l'exhaustivité du système.

Références

1. HUBERT B, DESENCLOS JC. Evaluation de l'exhaustivité et de la représentativité d'un système de surveillance par la méthode capture-recapture. Application à la surveillance des infections à méningocoques en France en 1989 et 1990. *Rev. Epidém. et Santé Publ.*, 1993 ; 41 : 241-249.
2. Michard V, Capek I, Guibourdenche M, Riou JY, Lepoutre A. Evaluation de la surveillance des infections à méningocoques en 1991 et 1992. données non publiées disponibles sur demande.
3. Direction générale de la Santé. Circulaire D.G.S./P.G.E./1 C, n°79 du 5 février 1990. *BEH* 1990 ; 7 : 25-27.
4. Papoz L, Balbau B, Lellouch J. Case counting in Epidemiology : Limitations of Methods Based on Multiple Data Sources. *International Journal of Epidemiology* 1996 ; 25 (3) : 474-8.
5. 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.
6. Cormak RM. Log-linears Models for Capture-Recapture. *Biometrics* 1989 ;45 :395-413.
7. Bishop YMM, Fienberg SE, Holland PW. Discrete Multivariate Analysis : Theory and Practice. Cambridge : MIT press, 1975. pp220-56.
8. Sekar CC, Deming WE. On a method of estimating birth and death rates and the extend of registration, *Amer Stat Ass J.* 1949 ; 44 : 100-15.
9. Chapman DG. Some properties of the hypergeometric distribution with applications to the zoological samples census. *Uni Calif Public Stat* 1951 ; 1 :1060-7.
10. Seber GAF. The effect of trap response in tag capture estimates. *Biometrics* 1970 ; 26 :13-22.
11. Wittes J, Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *J Chron Dis* 1968 ; 21 :287-301.
12. Wittes JT, Colton T, Sidel VW . Capture-recapture Methods for assessing the completeness of case ascertainment when using multiple information source. *J Chron Dis* 1974 ; 27 :25-36.
13. Rice J. Loglinear analysis : analysis of categorical variables in the logit setting. *Advances in Social Science Methodology*, Volume 2, pages 1-52.
14. 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.
15. Drapper D. Assessment and propagation of model uncertainty. *J R Stat Soc [B]* 1995 ; 57 : 45-70.
16. Hook EB, Regal RR. Capture-recapture methods in epidemiology : methods and limitations. *Epidemiol reviews* 1995 ; 17 (2) : 243-64.
17. Hook EB, Regal RR. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. *J Clin Epidemiology* 1999 ; 52 (10) : 917-26.
18. Hubert B. Les infections à méningocoques en France en 1996. *Annual Epidemiological Report. Infectious Diseases Epidemiology in France in 1996.* Réseau National de Santé Publique, Saint Maurice, France, December 1997.
19. Desenclos JC, Hubert B. Limitations to the universal use of capture-recapture methods. *Intern J Epidemiol* 1994 ; 23 (6) : 1322-3.
20. Chang YF, Laporte RE, Aaron DJ, Songer TJ. The importance of source selection and pilot study in the capture-recapture application. *J Clin Epidemiology* 1999 ; 52 (10) : 927-8.
21. Invasive meningococcal disease in Canada, 1 january 1997 to 31 december 1998. *CDCR* 2000 ; 26-21 : 177-182.
22. AC de Benoist, V goulet, E Laurent. Infections invasives à Haemophilus influenzae, Listeria monocytogenes, méningocoque, pneumocoque, streptocoques A et B en France en 1997. *Annual Epidemiological Report. Infectious Diseases Epidemiology in France in 1996.* Réseau National de Santé Publique, Saint Maurice, France, December 1997.
23. Mastro DT, Kitayapon D, Weniger BC et al. Estimating the number of HIV-infected injection drug users in Bangkok : a capture-recapture method. *Am J Public Health* 1994 ; 84 (7) : 1094-9.
24. Brenner H. Application of Capture-Recapture Methods for Disease Monitoring : Potential Effects of Imperfect Record Linkage. *Meth Inform Med* 1994 ; 33 : 502-6.
25. Cormak RM. Log-linears Models for Capture-Recapture. *Biometrics* 1989 ; 45 :395-413.
26. Domingo-Salvany A, Hartnoll R, et al. Analytical considerations in the use of capture-recapture to estimate prevalence : cases studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *American Journal of Epidemiology*, 1998 ; 148 (8) :732-40.

Annexe 1

FICHE

DE DÉCLARATION, DÉCLARATION OBLIGATOIRE

Questionnaire à retourner à la DDASS de

MALADIE À DÉCLARATION OBLIGATOIRE
(Décret du 10 juin 1986)

MÉNINGITE À MÉNINGOCOQUE ET MÉNINGOCOCCÉMIES

*Droit d'accès et de rectification par l'intermédiaire
du médecin déclarant (loi du 6 janvier 1978)*

Centralisation des informations au Réseau National de Santé Publique

Critères de déclaration :

Méningocoque isolé dans le sang ou le L.C.R.

Ou bien, présence d'antigène soluble dans le L.C.R., le sang ou les urines.

Important : Pour les cas survenus dans une collectivité, la DDASS doit être alertée dans les plus brefs délais, indépendamment de cette feuille de déclaration, en raison des mesures précoces de prévention à prendre dans l'entourage du malade (chimioprofylaxie avec ou sans vaccination).

Caractéristiques du malade :							
Initiale du nom : ____	Prénom : _____						
Sexe : <input type="checkbox"/> M <input type="checkbox"/> F							
Date de naissance : / ____ / ____ / ____ /	ou Âge : ____						
Code postal du domicile : _____							
Purpura fulminans (purpura extensif avec collapsus grave) ? : <input type="checkbox"/> Oui <input type="checkbox"/> Non							
Confirmation du diagnostic :							
- Méningocoque isolé dans	: <input type="checkbox"/> Sang <input type="checkbox"/> L.C.R. <input type="checkbox"/> Non isolé						
- Groupe	: <input type="checkbox"/> A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/> Autre <input type="checkbox"/> Non groupé						
- Antigène soluble	: <input type="checkbox"/> Absence <input type="checkbox"/> Présence <input type="checkbox"/> Non recherché						
Hospitalisation (phase aiguë) :	Date : / ____ / ____ / ____ / Durée : ____ jours						
Évolution :	<input type="checkbox"/> Guérison <input type="checkbox"/> Décès <input type="checkbox"/> Séquelles						
Si séquelles, préciser :							
Prophylaxie des sujets contacts :							
Chimioprofylaxie, nombre de personnes (Précisez l'antibiotique)	<table border="1"><thead><tr><th>Collectivité</th><th>Milieu familial</th></tr></thead><tbody><tr><td></td><td></td></tr><tr><td></td><td></td></tr></tbody></table>	Collectivité	Milieu familial				
Collectivité	Milieu familial						
Vaccination, nombre de personnes							
Autres cas dans l'entourage : <input type="checkbox"/> Oui <input type="checkbox"/> Non <input type="checkbox"/> Inconnu							
Pour chaque autre cas, indiquer les initiales et la date de diagnostic :							
-							
-							
-							

Médecin déclarant

Nom :

Adresse :

Téléphone :

Semaine de déclaration sur Minitel [____] [____]
(partie à remplir par la DDASS) SS AA

Date de déclaration : / ____ / ____ / ____ /

Signature et tampon :

Annexe 2

FICHE DE DÉCLARATION, CNR

IMPORTANT : Complétez les parties encadrées et joignez la liasse de 3 feuillets à chaque culture que vous adressez à l'état pur. Ces examens sont effectués à titre gratuit si les études bactériologiques et antigéniques minimales ont été effectuées ou si la souche bactérienne est atypique. Dans le cas contraire, ils seront facturés (suivant nomenclature).

N° Enregistrement

UNITÉ DES NEISSERIA
CENTRE NATIONAL DE REFERENCE
DES MENINGOCOQUES ET NEISSERIA APPARENTÉES
 Téléphone : 45 68 83 30
INSTITUT PASTEUR
 25-28, rue du Docteur Roux - 75724 PARIS Cedex 15

RENSEIGNEMENTS CLINIQUES

Malade Porteur

Nom : Prénom :

Sexe : Age :

ISOLEMENT Date :

Site :

LCR Sang Rhinopharynx

Crachat Liq. articulaire

Autres :

SIGNES CLINIQUES :

Méningite Méningococcémie

Autres :

Evolution :

ORIGINE GEOGRAPHIQUE DE LA CONTAMINATION

Cas isolé Epidémie

Adresse complète du laboratoire expéditeur (dans le cadre)
(écrire en MAJUSCULES, reproduire sur les 3 feuillets)
 Cette adresse sera utilisée telle quelle pour le retour de notre réponse
 (indiquer le **code postal**)
Personne du laboratoire à prévenir en cas d'urgence du résultat
 Nom : Tél.

RENSEIGNEMENTS BACTERIOLOGIQUES

ANTIGENES SOLUBLES : LCR Sang Urines

Date de la culture envoyée Milieu d'isolement utilisé

Morphologie bactérienne Oxydase Catalase Glucose Maltose Fructose Saccharose

Croissance milieu sélectif Neisseria :

Oui Non

Polysac ONPG Tributyrine DNase α GT NO₃

AGGLUTINATIONS : A B C

Autres :

Diagnostic proposé : Difficultés rencontrées :

Cadre réservé

Commentaires :

REPONSE DU CENTRE DE REFERENCE :

Diagnostic : Sérotype : Date d'arrivée :

Sérotype : Date de Réponse :

Sensibilité aux antibiotiques :

N° LNP

Annexe 3

FICHE

DE DÉCLARATION, EPIBAC

CODE DU LABORATOIRE : «**codelab**»

EPIBAC

RESEAU NATIONAL DE SANTE PUBLIQUE
14, rue du Val d'Osne - 94415 Saint-Maurice Cedex

 Tél : (1) 43.96.65.04

Année : 199 Mois : ou Trimestre :

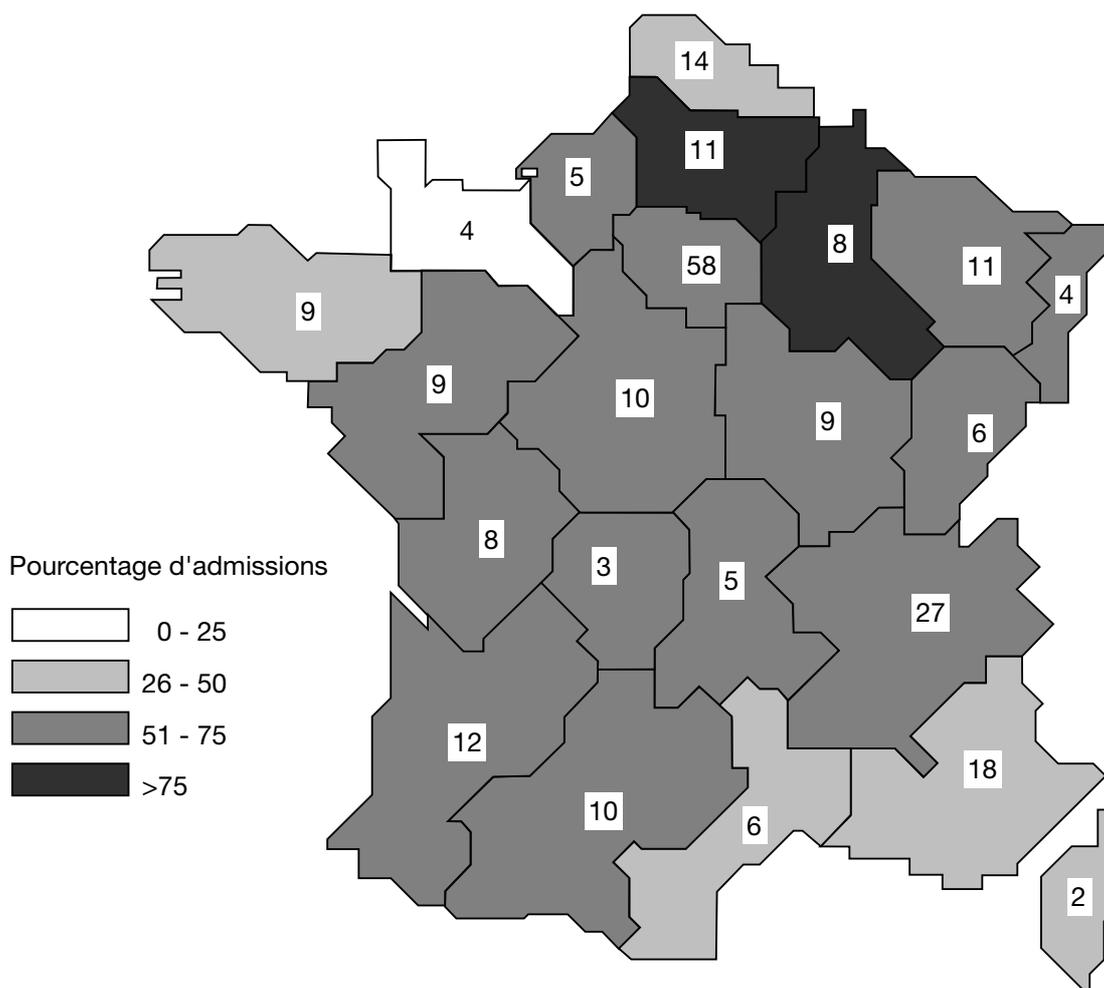
ISOLEMENTS (et/ou présence d'antigènes solubles) à partir de sites normalement stériles

	Date de Prélèvement	Date de Naissance		ou Age	Sexe	Prélèvements +			Ag solubles	Groupe Sérotype Biotype Lysovar
		Jour /Mois	Jour/Mois/An			ou Age	M ou F	LCR (+ si oui)		
1 - HAEMOPHILUS INFLUENZAE	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
2 - NEISSERIA MENINGITIDIS	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
3 - STREPTOCOCCUS PNEUMONIAE	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
	/	/ /								
4 - STREPTOCOCCUS groupe A	/	/ /								
	/	/ /								
5 - STREPTOCOCCUS groupe B	* /	/ /								
	* /	/ /								
	* /	/ /								
	* /	/ /								
6 - LISTERIA MONOCYTOGENES	* /	/ /								
	* /	/ /								
	* /	/ /								

Annexe 4

DISTRIBUTION

**GÉOGRAPHIQUE D'EPIBAC ET TAUX D'EXHAUSTIVITÉ
(NOMBRE D'ADMISSION EN MÉDECINE DES HÔPITAUX
COUVERTS PAR EPIBAC / NOMBRE TOTAL D'ADMISSION
EN MÉDECINE DE LA RÉGION)**

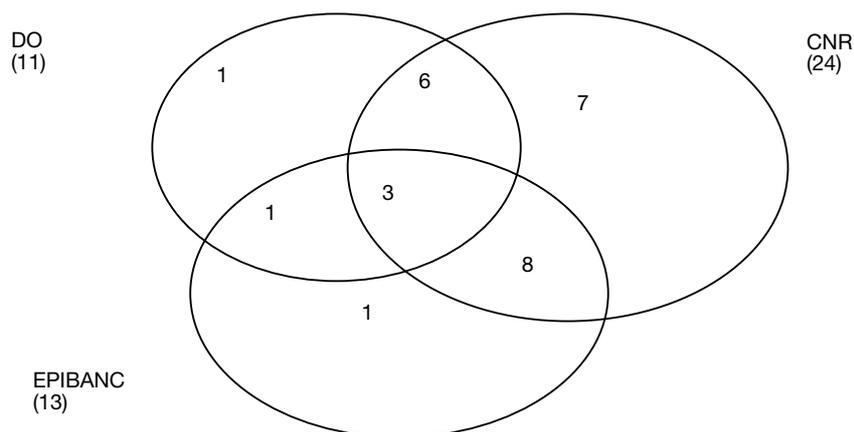


Annexe 5

RÉPARTITION

DES CAS D'INFECTION À MÉNINGOCOQUE DE SÉROGROUPE RARE ET DE SÉROGROUPE INCONNU, SELON LES TROIS SOURCES

Distribution des cas de sérotype rare après mise en commun des 3 sources et dépendance positive entre les sources CNR et DO et CNR et EPIBAC



Distribution des cas de sérotype inconnu après mise en commun des 3 sources et dépendance négative entre CNR et DO et CNR et EPIBAC

