

Maladies infectieuses

Estimation du nombre de diagnostics d'infection par le VIH chez les enfants par la méthode capture-recapture, France métropolitaine, 2003-2006

Rapport de master

Sommaire

Abréviations	2
1. Introduction	3
2. Objectifs	4
2.1 Objectif principal	4
2.2 Objectif spécifique	4
3. Matériels et méthodes	5
3.1 Population cible	5
3.2 Population d'étude	5
3.3 Définition d'un cas	5
3.4 Méthode capture-recapture	5
3.5 Recueil des données	8
3.6 Identification des doublons intrasource et des cas communs	10
3.7 Connaissances <i>a priori</i>	10
3.8 Réaffectation proportionnelle du pays de naissance	11
3.9 Stratégie d'analyse	11
3.10 Outils d'analyse statistique	11
4. Résultats	12
4.1 Enquête auprès du réseau LaboVIH	12
4.2 Doublons de la base LaboVIH	12
4.3 Nombre de diagnostics d'infection à VIH	12
4.4 Caractéristiques des enfants	12
4.5 Cas communs aux sources	13
4.6 Estimation du nombre total de diagnostics et évaluation de la dépendance	13
4.7 Réaffectation proportionnelle du pays de naissance dans la base LaboVIH	14
4.8 Modèles log-linéaires	14
5. Discussion	18
5.1 Estimations et exhaustivité	18
5.2 Obtention d'une troisième source	18
5.3 Réaffectation proportionnelle du pays de naissance	18
5.4 Hypothèses	19
6. Conclusion	20
Références bibliographiques	21
Annexe 1 – Questionnaire de l'enquête réalisée auprès des laboratoires participant à LaboVIH	23
Annexe 2 – Identification des cas communs entre deux sources	24

Estimation du nombre de diagnostics d'infection par le VIH chez les enfants par la méthode capture-recapture, France métropolitaine, 2003-2006

Rapport de master

Ce rapport a été rédigé et soutenu par Maxime Esvan, pour l'obtention du diplôme de master 2^e année de biostatistique de l'Université de Bordeaux 2, dans le cadre d'un stage effectué d'avril à août 2008 à l'Institut de veille sanitaire (InVS), au sein de l'unité VIH-IST-hépatites C et B chronique du Département des maladies infectieuses.

À l'InVS, ont participé à ce travail :

- Maxime Esvan (recueil et analyse des données, rédaction du rapport);
- Florence Lot (participation au recueil des données et à la rédaction du rapport);
- Anne Gallay (soutien méthodologique pour l'analyse des données et participation à la rédaction du rapport);
- Pascale Bernillon (soutien biostatistique);
- Vanina Bousquet (réalisation de l'imputation multiple);
- Françoise Cazein (aide à l'enquête auprès des biologistes).

À l'unité VIH/IST de l'Institut national de la santé et de la recherche médicale CESP 1018, ont participé à ce travail :

- Karima Hamrene (participation au recueil des données);
- Josiane Warszawski (responsable de l'Enquête périnatale française – EPF).

Remerciements

Nous remercions tous les biologistes qui participent à LaboVIH, et plus particulièrement ceux qui ont été sollicités pour renseigner des données sur les diagnostics d'infection à VIH chez les enfants.

Nous remercions tous les pédiatres qui participent à l'EPF, ainsi qu'au système de déclaration obligatoire de l'infection par le VIH chez les enfants.

L'enquête EPF-ANRS C01/C010/C011 est soutenue par l'Agence nationale de recherches contre le sida et les hépatites virales.

Abréviations

AIC	Critère d'information d'Akaike
ANRS	Agence nationale de recherches contre le sida et les hépatites virales
Ddass	Direction départementale des affaires sanitaires et sociales
DOVIH	Déclaration obligatoire du VIH
EPF	Enquête périnatale française
LaboVIH	Surveillance de l'activité de dépistage du VIH dans les laboratoires d'analyses de biologie médicale
TME	Transmission mère-enfant
VIH	Virus de l'immunodéficience humaine

1. Introduction

En France, les infections par le virus de l'immunodéficience humaine (VIH) chez les enfants sont principalement liées à une transmission mère-enfant (TME). Cette transmission concerne soit des enfants nés en France, soit des enfants nés à l'étranger qui arrivent ultérieurement en France.

Chez les femmes accouchant en France, grâce à des stratégies prophylactiques et thérapeutiques, la TME est désormais très faible, avec un taux de transmission de 1,3 % chez les femmes traitées pendant la grossesse et ayant accouché entre 1997 et 2004 [1]. Trois facteurs principaux ont été identifiés pour expliquer les échecs : une charge virale maternelle élevée au moment de l'accouchement, une durée de traitement antirétroviral courte et un accouchement très prématuré.

La prévention de la TME nécessite que la mère connaisse sa séropositivité. La loi du 27 janvier 1993 stipule qu'un test de dépistage de l'infection par le VIH soit systématiquement proposé aux femmes enceintes, lors du premier examen prénatal, après information sur les risques de contamination. Cependant, les données de l'Enquête nationale périnatale réalisée en 2003 semblent montrer que seules quatre femmes sur cinq déclarent avoir été testées pour le VIH durant la grossesse [2]. De plus, un dépistage négatif lors de la première consultation peut être suivi d'une contamination au cours de la grossesse, qui peut passer alors totalement inaperçue. Ces échecs du dépistage ont été retrouvés dans l'analyse des diagnostics d'un certain nombre d'enfants suivis à Necker [3] et empêchent les femmes de bénéficier des progrès thérapeutiques réalisés dans la prévention de la TME du VIH.

La France est historiquement un pays d'accueil pour les migrants, qu'ils soient originaires d'Europe, d'Afrique ou d'ailleurs. Les enfants, en provenance d'une région où la prévalence du VIH est élevée, sont particulièrement susceptibles d'être porteurs d'une infection à VIH transmise par leur mère lors de l'accouchement ou lors de l'allaitement, en raison de la difficulté d'accès à une prévention de la TME [4]. C'est le cas dans beaucoup de pays d'Afrique subsaharienne et les enfants qui sont nés dans cette région du monde représentent une part importante des diagnostics d'infection à VIH faits en France chez les moins de 13 ans [5].

En France, la surveillance épidémiologique du VIH/sida repose essentiellement sur le système de la déclaration obligatoire, du sida

depuis 1986 et du VIH depuis 2003 [6]. Cette déclaration obligatoire a pour but de recueillir l'exhaustivité des cas, afin de mettre en place des actions préventives et de conduire des programmes adaptés aux besoins de santé publique. Néanmoins, la sous-déclaration des diagnostics d'infection à VIH a été estimée globalement autour de 35 % et pourrait être encore plus importante pour les cas pédiatriques [5]. Conjointement, il existe une surveillance de l'activité de dépistage du VIH, permettant de recueillir des informations auprès de l'ensemble des laboratoires d'analyses biologiques et médicales (enquête LaboVIH). Cette surveillance recense le nombre de personnes testées pour le VIH chaque année et le nombre de personnes confirmées positives [7].

Parallèlement à ces deux systèmes de surveillance, existe depuis 1985 l'Enquête périnatale française (EPF) qui est une cohorte prospective de mères séropositives. Cette cohorte a joué un rôle essentiel dans l'acquisition des connaissances sur les mécanismes de transmission du VIH, sa prévention et sur l'évolution de l'infection chez l'enfant [1,8]. Elle repose sur un certain nombre de services hospitaliers participants, dont le recrutement représenterait environ 70 % des femmes séropositives accouchant en France.

Des chiffres circulent sur le nombre d'enfants vivant avec le VIH en France (environ 1 500) [9]. Quant au nombre de nouveaux diagnostics chaque année, il est estimé entre 15 et 20 chez les nouveau-nés en France, auquel il faudrait rajouter les diagnostics chez les enfants nés en pays de forte endémie qui arrivent en France après leur naissance, pour lesquels aucune estimation n'existe [9]. Pourtant, une surveillance épidémiologique et une cohorte à visée de recherche coexistent, même si l'exhaustivité de ces systèmes est mal connue. C'est pourquoi, la méthode de capture-recapture pouvait représenter une alternative intéressante pour estimer la taille de la population étudiée. Le terrain d'application initial de cette méthode était la zoologie et s'est par la suite développé dans le champ de l'épidémiologie [10-14]. Cette technique peut être utilisée lorsqu'au moins deux sources d'informations permettent d'identifier les cas d'une même maladie.

Cette étude a pour but d'estimer le nombre de nouveaux diagnostics d'infection à VIH chez les enfants de moins de 13 ans, en France métropolitaine, entre 2003 et 2006, par la méthode de capture-recapture à partir de trois sources de données : la déclaration obligatoire du VIH, l'enquête LaboVIH et l'EPF.

2. Objectifs

2.1 OBJECTIF PRINCIPAL

Déterminer le nombre d'enfants âgés de moins de 13 ans nouvellement diagnostiqués comme infectés par le VIH en France métropolitaine entre le 1^{er} janvier 2003 et le 31 décembre 2006.

2.2 OBJECTIF SPÉCIFIQUE

Évaluer l'exhaustivité du système de déclaration obligatoire de l'infection à VIH et celle de l'EPF pour les diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans.

3. Matériels et méthodes

Il s'agit d'une étude rétrospective, utilisant la méthode de capture-recapture, réalisée par la mise en commun des diagnostics d'infection par le VIH chez les enfants âgés de moins de 13 ans en France métropolitaine entre le 1^{er} janvier 2003 et le 31 décembre 2006, à partir de trois sources disponibles.

3.1 POPULATION CIBLE

Les enfants âgés de moins de 13 ans, pour lesquels un nouveau diagnostic d'infection par le VIH, a été réalisé en France métropolitaine, du 1^{er} janvier 2003 au 31 décembre 2006, constituent la population cible.

3.2 POPULATION D'ÉTUDE

Elle est constituée par tous les nouveaux diagnostics d'infection à VIH chez les enfants de moins de 13 ans, recensés pendant la période allant du 1^{er} janvier 2003 au 31 décembre 2006, par au moins un des trois systèmes d'information suivant :

- la déclaration obligatoire du VIH (DOVIH) ;
- l'EPF ;
- la surveillance de l'activité de dépistage des laboratoires d'analyse de biologie médicale (LaboVIH).

3.3 DÉFINITION D'UN CAS

Un cas est un enfant âgé de moins de 13 ans chez lequel un diagnostic d'infection à VIH a été réalisé en France métropolitaine entre le 1^{er} janvier 2003 et le 31 décembre 2006. Le diagnostic est basé sur un examen virologique (culture virale ou PCR) chez l'enfant de moins de 18 mois ou sur une sérologie confirmée positive (Western Blot) chez l'enfant de plus de 18 mois.

3.4 MÉTHODE CAPTURE-RECAPTURE

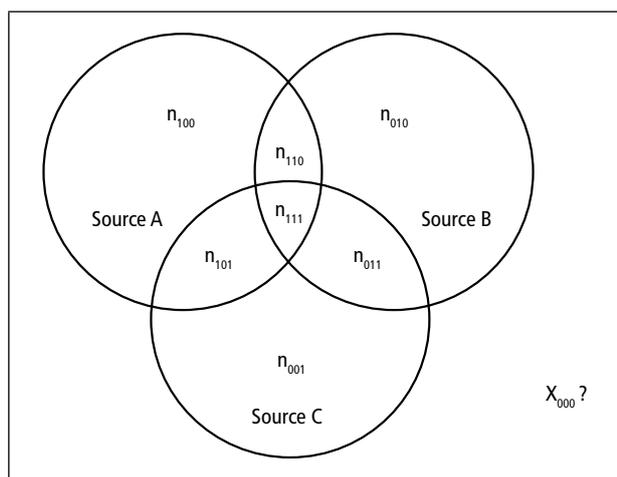
3.4.1 Principe

Cette méthode permet, en croisant différentes sources (figure 1) et après avoir identifié les cas communs entre ces sources, d'estimer, sous certaines hypothèses, le nombre de cas présents dans aucune des sources (x_{000}).

La méthode dite "capture-recapture" fut d'abord utilisée en zoologie. En épidémiologie, elle permet d'estimer le nombre total de cas d'une maladie donnée à partir de plusieurs sources, d'affiner la prévalence ou l'incidence à partir d'études portant sur la population générale, d'évaluer la complétude d'un système de surveillance et de fournir un intervalle de confiance sur le nombre de cas atteints une maladie [12-14].

| FIGURE 1 |

Répartition des cas d'une maladie recensés par trois sources A, B, C : principe de la méthode capture-recapture



Les indices 1 et 0 correspondent à la présence (1) et l'absence des cas (0) dans les sources A, B et C respectivement.

3.4.2 Conditions d'application

La fiabilité des résultats dépend du respect de différentes conditions d'applications :

- Identification des vrais cas
L'identification de faux cas par une source induit une surestimation de nombre total de cas et une sous-estimation de l'exhaustivité pour chacune des autres sources. Par ailleurs, une définition des cas différente selon les systèmes induit également une surestimation du total (N).
- Identification de tous les vrais cas communs
Une surestimation des cas communs entrainera une sous-estimation du nombre total de cas et inversement. S'il n'existe pas d'identifiant commun unique entre les différents systèmes, l'identification des cas communs est basée sur une combinaison de critères.
- Même période et même zone d'étude
Si l'une des sources considère une zone géographique ou une période de l'étude différente de celles des autres systèmes, les estimations obtenues peuvent être biaisées.
- Homogénéité de capture
Tous les cas de la population étudiée ont la même probabilité d'être identifiés par une même source, même si cette probabilité peut être différente selon les sources. Dans une source, la notification des cas ne doit pas être liée à une ou plusieurs caractéristiques individuelles.

- Indépendance entre les sources

Cette condition implique que les cas recensés dans une source ne dépendent pas du fait qu'ils soient recensés ou non dans une autre source. Si l'identification des cas par un système augmente la probabilité de capture par un autre système, une dépendance positive sera constatée, entraînant une sous-estimation du nombre total. À l'inverse, si l'identification des cas par une source diminue la capacité de capture par une autre source, cela entraînera une surestimation causée par une dépendance négative.

- Population d'étude close

Aucun mouvement de population ne doit être constaté sous peine de biaiser le nombre de cas communs et donc la taille de la population.

3.4.3 Estimation du nombre total de cas avec deux sources

L'estimation du nombre total de cas avec deux sources est possible sous l'hypothèse d'indépendance entre les sources.

Le tableau 1 présente la répartition des cas selon leur présence dans chacune des deux sources. Dans ce tableau, la présence dans une source est codée avec l'indice 1 et l'absence par l'indice 0. La cellule correspondant au nombre de cas identifiés par aucune des sources (n_{00}) est celle à estimer.

Sous l'hypothèse d'indépendance des sources, les estimateurs de Sekar et Deming (ou estimateur du maximum de vraisemblance (tableau 2) permettent d'estimer n_{00} ainsi que sa variance [15].

Lorsque les effectifs sont faibles, Chapman et Seber ont proposé d'autres estimateurs (ou estimateurs presque non biaisés (tableau 2) [16]. Un intervalle de confiance pour N se calcule dans les deux cas de la manière suivante :

$$IC\ 95\ \% \ N = \hat{N} \pm 1,96 \times \sqrt{Var(\hat{N})}$$

Le taux d'exhaustivité de la source A est $P_A = \frac{N_A}{N}$ où N_A est le nombre de cas recensés dans la source A et N le nombre de cas estimés.

| TABLEAU 1 |

Tableau de contingence avec deux sources

		Source A		Total
		Oui	Non	
Source B	Oui	n_{11}	n_{01}	N_B
	Non	n_{10}	n_{00}	
Total		N_A		$N = n_{11} + n_{10} + n_{01} + n_{00}$

| TABLEAU 2 |

Formules des estimateurs du nombre total de cas, méthode capture-recapture à deux sources

Valeurs estimées		Estimateur du maximum de vraisemblance	Estimateur presque non biaisé
Cellule vide	\hat{n}_{00}	$\frac{n_{01} \times n_{10}}{n_{11}}$	$\frac{n_{01} \times n_{10}}{(n_{11} + 1)}$
Nombre de cas	\hat{N}	$\frac{N_A \times N_B}{n_{11}}$	$\left[\frac{(N_A + 1) \times (N_B + 1)}{n_{11} + 1} \right] - 1$
Variance du nombre de cas	$Var(\hat{N})$	$\frac{N_A \times N_B \times n_{01} \times n_{10}}{n_{11}^3}$	$\frac{(N_A + 1) \times (N_B + 1) \times n_{01} \times n_{10}}{(n_{11} + 1)^2 \times (n_{11} + 2)}$

3.4.4 Évaluation de la dépendance avec plus de deux sources

Lorsque la méthode capture-recapture est appliquée avec plus de deux sources, la dépendance entre les sources peut être évaluée et prise en compte pour l'estimation du nombre total de cas. 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 couple de sources donne une

estimation très différente des autres couples, une dépendance entre ces deux sources peut être suspectée [10,11].

L'indépendance entre deux sources A et B peut être testée avec un test du χ^2 appliqué sur un tableau de contingence 2*2 qui répartit tous les cas notifiés au moins une fois par les sources autres que A et B. Un rapport de cotes (OR) et son intervalle de confiance à 95 % permettent d'indiquer le sens et la force des dépendances entre les sources deux à deux.

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

L'estimation du nombre total de cas avec plus de deux sources permet de s'affranchir de la dépendance entre les sources.

- Principe

Le modèle log-linéaire permet d'étudier les relations entre plusieurs 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'interaction.

Avec la méthode capture-recapture, le tableau de contingence a une cellule structurellement vide correspondant à l'absence de notification de cas dans les différentes sources. Le modèle log-linéaire utilise toutes les cellules non vides afin d'estimer les effectifs attendus, y compris celui correspondant à une absence totale de notification. La présence d'une cellule vide dans le tableau de contingence rend impossible l'ajustement du modèle en tenant compte de l'interaction d'ordre maximum entre toutes les sources.

Ainsi, le modèle complet pour trois sources :

$$\ln(F_{ijk}) = \theta + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{AC} + \lambda_{ik}^{BC}$$

F_{ijk} correspond à la fréquence attendue pour la cellule ijk du modèle considéré. λ_i^A , λ_j^B et λ_k^C correspondent aux effets principaux de la présence dans chaque source. λ_{ij}^{AB} , λ_{jk}^{AC} et λ_{ik}^{BC} sont les termes d'interactions d'ordre 2 de la présence dans chacune des deux sources indiquées. Ici, i , j et k varient tous les trois de 0 à 1.

Lorsque seule la dépendance entre les sources est prise en compte, huit modèles peuvent être ajustés aux données, du modèle saturé avec les trois interactions d'ordre un, au modèle le plus simple sans aucune interaction.

Les modèles log-linéaires permettent de calculer des estimations en tenant compte des dépendances entre les sources, mais également des variables d'hétérogénéité de capture, en les considérant dans les modèles.

- Validité et choix d'un modèle

La validité d'un modèle est basée sur l'adéquation du modèle aux données observées. Un modèle sera considéré comme adéquat si le test de la déviance (G^2) est non significatif et peu élevé. La sélection du modèle repose sur le principe de parcimonie. Le modèle s'adaptant aux connaissances épidémiologiques et ayant le plus faible critère d'information d'Akaike (AIC) pourra être retenu :

$$AIC = G^2 - 2 \times ddl \text{ (avec } ddl \text{ : nombre de degrés de liberté)}$$

Lorsque différents modèles construits à partir d'une base commune sont en concurrence, une méthode de pondération ("weighted AIC") peut fournir une moyenne pondérée tenant compte des estimations de chaque modèle. Ainsi, il suffit de pondérer, pour chaque modèle i , l'estimation du nombre de cas notifiés dans aucune source par $1/e^{(AIC_i/2)}$ [17].

Avec trois sources, le tableau 3 indique la répartition des cas selon leur présence ou non dans les sources.

| TABLEAU 3 |

Tableau de contingence avec trois sources

		Source A			
		Oui		Non	
		Source B		Source B	
		Oui	Non	Oui	Non
Source C	Oui	n_{111}	n_{101}	n_{011}	n_{001}
	Non	n_{110}	n_{100}	n_{010}	n_{000}

La présence dans une source est codée avec un indice 1 et l'absence par un indice 0.

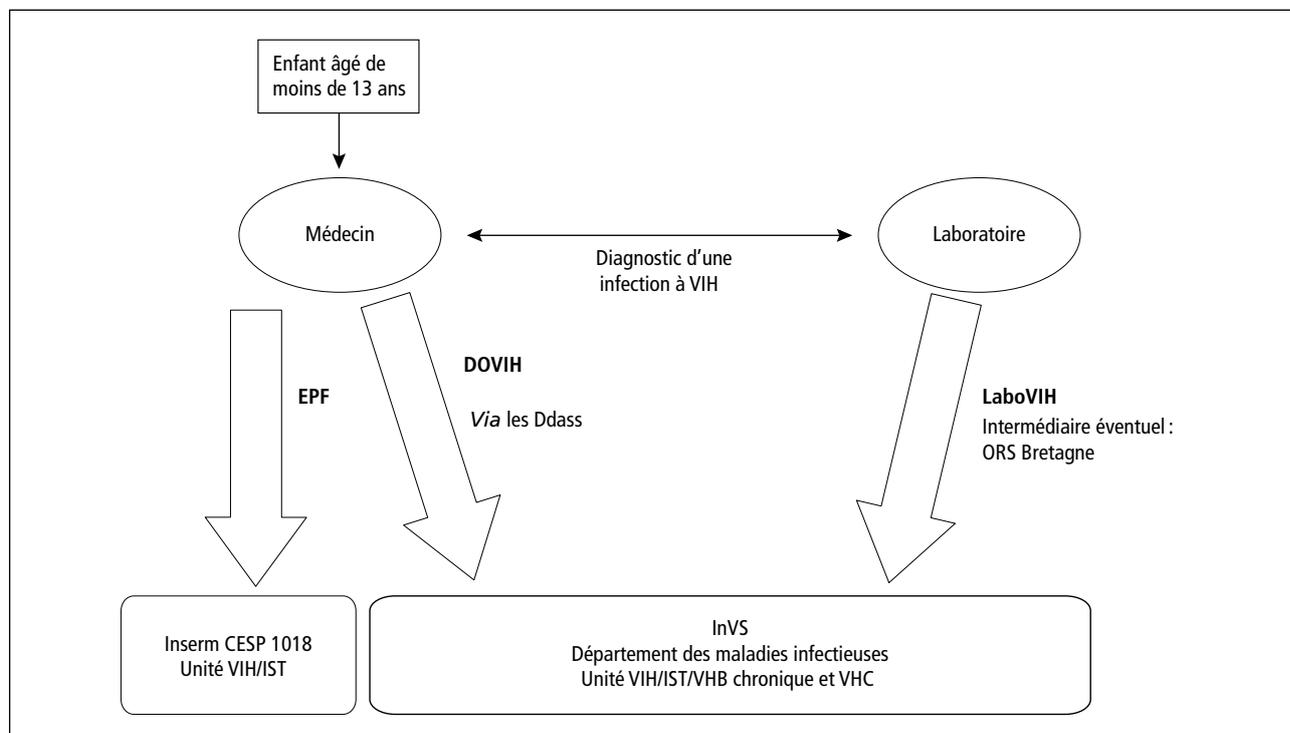
Le modèle retenu fournit une estimation du nombre de cas notifiés dans aucune source (\hat{n}_{000}), puis celle de \hat{N} grâce à la relation $\hat{N} = N_{obs} + \hat{n}_{000}$. Les intervalles de confiance des estimateurs du nombre total de cas peuvent être calculés d'après les formules de Bishop [18] ou être basés sur la déviance [19].

3.5 RECUEIL DES DONNÉES

La figure 2 décrit le circuit de recueil des diagnostics d'infection à VIH pour chacune des trois sources utilisées pour notre étude.

FIGURE 2 |

Circuit du recueil des diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans pour chacune des trois sources : le système de la déclaration obligatoire du VIH (DOVIH), l'Enquête périnatale française (EPF) et la surveillance de l'activité de dépistage du VIH (LaboVIH), France métropolitaine, 2003-2006



3.5.1 Déclaration obligatoire du VIH (DOVIH)

Le système de la déclaration obligatoire existe depuis plus d'un siècle en France dans le but de mieux connaître certaines maladies, leurs différentes formes et les populations qu'elles touchent. Ce dispositif de surveillance concerne 30 maladies qui nécessitent une intervention locale urgente, nationale ou internationale et/ou celles dont la surveillance est nécessaire à la conduite et à l'évaluation de la politique de santé publique. Les médecins et les biologistes (depuis 2003) sont tenus de transmettre aux autorités des informations concernant la personne chez laquelle ils diagnostiquent une de ces maladies.

L'une de ces maladies à déclaration obligatoire est l'infection par le VIH, quel qu'en soit le stade. Cette déclaration obligatoire a été mise en place en mars 2003 par l'InVS et recense les nouveaux diagnostics d'infection à VIH depuis janvier 2003. Cette déclaration obligatoire est venue compléter celle du sida qui existe depuis 1986 [6]. Une fiche de notification obligatoire spécifique s'intéresse aux enfants âgés de moins de 13 ans. Cette fiche est complétée par les pédiatres avec un code d'anonymat (construit sur la base de la date de naissance, du prénom, de l'initiale du nom et du sexe), puis adressée au médecin inspecteur de santé publique de la Direction départementale des affaires sanitaires et sociales (Ddass), puis à l'InVS. Jusqu'en juin 2007, la DOVIH chez l'enfant n'était réalisée que par les pédiatres et non sur

la base d'une double déclaration par le biologiste et le clinicien, comme chez l'adulte. En juillet 2007, le dispositif de surveillance de l'infection à VIH a été modifié à la demande des pédiatres, afin d'impliquer les biologistes et d'améliorer l'exhaustivité des données recueillies. La mise en place de ce nouveau dispositif est postérieure à la période de notre étude.

La base constituée pour l'étude a sélectionné les cas à partir du fichier disponible en mai 2008, qui incluait les fiches de notification reçues jusqu'au 30 septembre 2007.

3.5.2 Enquête périnatale française (EPF)

L'EPF est une cohorte prospective de mères séropositives incluses au plus tard à l'accouchement, ayant donné leur accord de participation pour leur inclusion et le suivi de leurs enfants infectés ou non. Cette cohorte initiée en 1985 est coordonnée par l'équipe VIH-IST de l'Inserm CESP 1018 [1,8]. Elle repose sur une centaine de services de maternité et de pédiatrie sur l'ensemble du territoire français. Son objectif est d'étudier les facteurs associés à la TME du VIH, la tolérance des antirétroviraux administrés en périnatal chez les mères et leurs enfants, et le pronostic de l'infection pédiatrique. Elle a évolué en trois parties distinctes à partir de 2005 (figure 3).

3.5.2.1 Volet ANRS-CO01

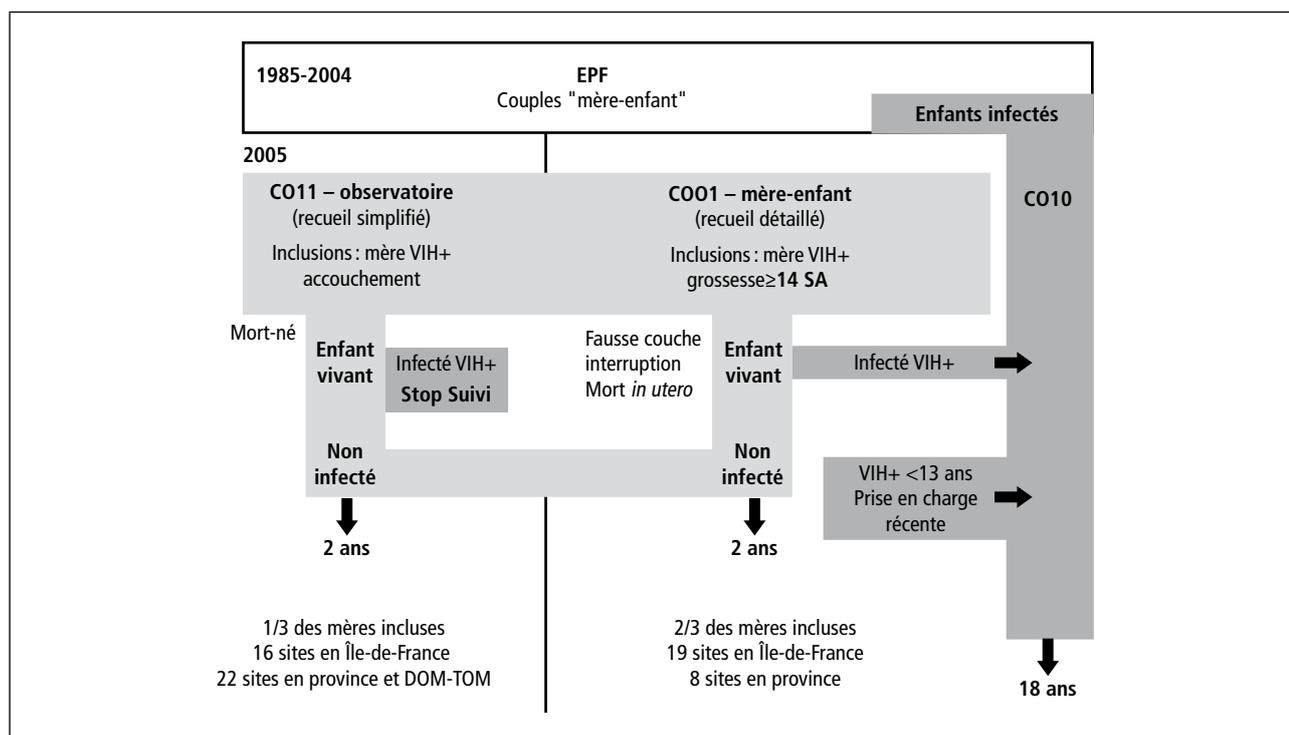
Il concerne l'étude de la transmission du VIH de la mère à l'enfant, grâce à l'inclusion de femmes enceintes infectées par le VIH et de leurs enfants nés en France dans l'un des 27 sites participants répartis sur l'ensemble du territoire français. Ce volet permet en particulier de décrire l'évolution des stratégies de prévention de la transmission mère-enfant du VIH et d'évaluer l'efficacité.

3.5.2.2 Volet ANRS-CO11

Il s'agit d'un observatoire de la TME et de pharmacovigilance qui vient compléter CO01 par un recueil simplifié dans un plus grand nombre de sites (38) ne participant pas à CO01. Son but est de constituer un dispositif d'alerte pour surveiller l'évolution du taux de TME et des pratiques préventives en France, et de repérer la survenue de toxicité chez les enfants exposés en période périnatale aux antirétroviraux.

| FIGURE 3 |

Fonctionnement de l'Enquête périnatale française, Inserm CESP 1018



3.5.2.3 Volet ANRS-CO10

C'est une cohorte d'enfants séropositifs suivis jusqu'à l'âge de 18 ans. Elle intègre la cohorte des enfants nés dans les principaux sites de CO01 et étend les inclusions à tous les enfants de moins de 13 ans nouvellement pris en charge dans les sites participants sous réserve qu'ils soient naïfs de traitement antirétroviral. Le recueil d'enfants a été rétrospectif entre 2003 et 2004 puis prospectif à partir d'octobre 2004. Son but est d'étudier le pronostic à court, moyen et long terme des enfants infectés en période périnatale.

Pour notre étude, une base a été constituée en mai 2008 et regroupait l'ensemble des enfants séropositifs de moins de 13 ans identifiés par l'un des trois volets de la cohorte EPF quel que soit le site participant (à l'exclusion des DOM), dont le diagnostic de l'infection à VIH avait été réalisé entre le 1^{er} janvier 2003 et le 31 décembre 2006.

3.5.3 Surveillance de l'activité de dépistage (Enquête LaboVIH)

Afin de surveiller l'activité de dépistage de l'infection à VIH en France, l'InVS a mis en place à partir de 2001 l'enquête LaboVIH [7]. Il s'agit

d'une surveillance nationale réalisée auprès de l'ensemble des 4 200 laboratoires d'analyses de biologie médicale de France, qu'ils soient de ville ou hospitalier, de métropole ou d'outre-mer.

Chaque laboratoire envoie directement à l'InVS des données agrégées par semestre sur le nombre de personnes testées pour le VIH et sur le nombre de personnes confirmées positives. Il précise aussi le nombre de diagnostics d'infections confirmés chez des enfants âgés de moins de 15 ans.

Une troisième source a été constituée spécifiquement à partir de ce réseau de surveillance pour les besoins de la méthode capture-recapture. Parmi les laboratoires participant chaque année (le taux de participation est d'environ 85 %), 137 laboratoires métropolitains avaient signalé au moins un diagnostic d'infection à VIH confirmé chez un enfant sur la période 2003-2006. Un questionnaire leur a été envoyé début avril 2008 afin de recueillir des données individuelles pour chacun de ces diagnostics. Pour chaque enfant infecté par le VIH, son année de naissance, son sexe, son département ou pays de domicile, la date et l'âge au diagnostic, le type de virus, les techniques du diagnostic (examen virologique ou sérologie) et les coordonnées du médecin prescripteur ont été recueillis (annexe 1).

Afin d'obtenir un meilleur taux de réponse, deux relances ont été effectuées auprès des laboratoires n'ayant pas répondu, la première par courrier fin avril et la seconde, par téléphone courant mai.

Dans le cadre de la surveillance de l'activité de dépistage, les données agrégées des laboratoires de la région Bretagne sont envoyées à l'Observatoire régional de la Santé de Bretagne, puis sont transmises globalement pour la région à l'InVS sans pouvoir identifier les laboratoires. Pour cette raison, il n'a pas été possible de cibler les laboratoires ayant réalisé un diagnostic chez un enfant et le questionnaire de recueil individuel d'information a donc été envoyé à l'ensemble des laboratoires bretons.

3.6 IDENTIFICATION DES DOUBLONS INTRASOURCE ET DES CAS COMMUNS

L'une des conditions d'application de la méthode de capture-recapture est que seuls les vrais cas communs doivent être identifiés.

3.6.1 Identification des doublons intrasource

La première étape a consisté à éliminer les doublons à l'intérieur d'une même source. Cette recherche n'avait de sens que pour la source LaboVIH, dans la mesure où les biologistes déclarent le nombre de nouveaux diagnostics VIH dans leur laboratoire, mais que des doublons interlaboratoires peuvent exister. Au niveau des deux autres sources, la recherche des doublons intrabase est réalisée par l'InVS (grâce au code d'anonymat de la DOVIH) et par l'Inserm au fur et à mesure de la saisie des cas dans l'EPF.

3.6.2 Identification des cas communs entre deux sources

Aucun identifiant commun et unique aux trois bases n'existait. L'identification des cas communs a nécessité plusieurs algorithmes. Du fait du faible nombre d'individus dans les différentes bases, les résultats fournis par les algorithmes ont pu être vérifiés manuellement. La validation de certains cas communs identifiés comme possibles avec les algorithmes a nécessité la consultation des fiches de DO et des dossiers EPF à la recherche d'informations non présentes dans les bases informatiques.

Cela a permis de classer les cas communs identifiés comme certains, possibles ou faux. L'ensemble des algorithmes, ainsi que les résultats détaillés de ceux-ci se trouvent en annexe 2.

3.6.2.1 Recherche des cas communs DOVIH – EPF

Les différents critères utilisés pour l'identification des cas communs entre les bases DOVIH et EPF étaient : l'année de naissance, le sexe, le pays de naissance ou la maternité de naissance (ou son département) pour les enfants nés en France, la date de diagnostic (ou de prise en charge) à plus ou moins un mois près, l'origine géographique de la mère, l'hôpital de suivi (ou son département) et le statut vital.

3.6.2.2 Recherche des cas communs LaboVIH – EPF et LaboVIH – DOVIH

Les algorithmes étaient basés sur les critères suivants : l'année de naissance ou l'âge au diagnostic, le sexe, la date de diagnostic (ou de prise en charge) et l'hôpital de suivi.

3.6.3 Identification des cas communs aux trois sources

L'identification des cas communs aux trois sources a été réalisée en deux étapes :

- lorsqu'un cas commun était identifié par chacun des trois croisements de sources, un triplet certain était constitué ;
- lorsqu'un cas commun était identifié par deux croisements de sources, mais pas pour le troisième croisement, l'éventuel triplet a été validé par un contrôle manuel.

3.7 CONNAISSANCES A PRIORI

L'interprétation des résultats et les choix de modèles sont subordonnés au respect des conditions d'application de la méthode capture-recapture. Afin de prendre en compte les limites liées au non-respect de ces conditions dans notre étude, différentes hypothèses ont été suggérées *a priori*.

3.7.1 Dépendances entre les sources

Les pédiatres des services hospitaliers participant à la cohorte EPF ont été sensibilisés à la DOVIH à plusieurs reprises. Ils pourraient donc mieux déclarer les nouveaux diagnostics d'infection à VIH chez les enfants à l'InVS, entraînant une dépendance positive entre les sources DOVIH et EPF.

3.7.2 Hétérogénéité

Trois variables ont été identifiées comme pouvant être potentiellement des variables d'hétérogénéité :

- La région (Île-de-France *versus* Province)

Il pourrait y avoir une moins bonne exhaustivité de la cohorte EPF pour les diagnostics réalisés en Province, en raison d'un nombre de sites participants plus faible comparativement à la région Île-de-France qui regroupe la plupart des sites.

Dans le système de la DOVIH, il existe une disparité selon les régions, avec une sous-déclaration plus importante dans certaines régions, qui peut atteindre plus de 50 % [20].

Le taux de participation des laboratoires au réseau LaboVIH varie également selon la région déclarante (de 78 % à 98 % en 2005) [7].

- L'année de diagnostic

Il existe un délai de notification dans la DOVIH pouvant être à l'origine d'une moins bonne complétude de la base au 30 septembre 2007 pour l'année 2006 [20]. De plus, l'année 2003 pourrait être incomplète car, malgré un rattrapage, la mise en place du système de DOVIH n'a eu lieu qu'en mars de cette année là.

- Le pays de naissance (France *versus* étranger)

Au niveau de la cohorte EPF, il y aurait une meilleure exhaustivité pour les enfants nés en France que pour ceux nés à l'étranger. En effet, l'inclusion d'enfants infectés concernait au départ uniquement les enfants nés de mères connues comme séropositives à l'accouchement et ayant accouché dans une maternité d'EPF (volet CO01). L'inclusion des enfants nés à l'étranger et des enfants nés en France de mères découvertes séropositives après l'accouchement est récente (volet CO10 mis en place en 2005); pour les années 2003 et 2004, l'inclusion de ces enfants a été rétrospective et certains cas ont pu échapper au recueil. De plus, la demande d'un consentement signé chez les parents nouvellement confrontés à l'infection par le VIH chez leurs enfants se heurte parfois à un refus de leur part.

3.8 RÉAFFECTATION PROPORTIONNELLE DU PAYS DE NAISSANCE

Les caractéristiques des populations touchées par l'épidémie de VIH/sida en France se sont modifiées depuis la fin des années 1990 avec une proportion croissante de personnes originaires de pays d'endémie, et en particulier d'Afrique subsaharienne [20]. Le nombre d'enfants nés de mères originaires d'Afrique subsaharienne a lui aussi augmenté, qu'il s'agisse d'enfants nés en France ou en Afrique [5,21].

La probabilité d'inclusion des cas dans une source peut varier selon le pays de naissance (né en France *versus* né à l'étranger). Cette variable est vraisemblablement une variable d'hétérogénéité de capture, notamment pour la source EPF. Cependant, cette variable n'était pas disponible auprès des biologistes du réseau LaboVIH.

Afin de prendre en compte cette variable dans l'analyse, nous avons fait le choix de réaffecter cette information pour les cas notifiés uniquement dans la source LaboVIH. Lorsqu'un cas du réseau LaboVIH était commun à au moins l'une des deux autres sources (DOVIH et EPF), le pays de naissance a été conservé pour ce cas dans la source LaboVIH.

La distribution de l'âge au diagnostic en fonction du pays de naissance dans les sources DOVIH et EPF suggérait que l'âge au diagnostic était lié au pays de naissance. Les enfants nés en France étaient diagnostiqués plus tôt que les enfants nés à l'étranger.

Par une réaffectation proportionnelle pour chaque classe d'âge (1 an ; >1 an), nous avons appliqué la distribution observée du pays de

naissance dans les sources DOVIH et EPF aux cas recensés uniquement dans LaboVIH, sous l'hypothèse d'une distribution similaire de l'âge au diagnostic en fonction du pays de naissance dans la source LaboVIH. Il a ainsi été possible de calculer les effectifs des enfants nés à l'étranger et en France, pour chacune des deux classes d'âge.

3.9 STRATÉGIE D'ANALYSE

La dépendance entre les sources considérées deux à deux a été approchée dans un premier temps par le calcul des rapports de cotes de Wittes, puis confirmée par les huit modèles log-linéaires disponibles avec trois sources, grâce à un module spécifique [22].

Pour chaque variable d'hétérogénéité supposée (pays de naissance, année et région de diagnostic), les interactions entre les sources et cette variable ont été testées grâce à une stratégie pas à pas descendante manuelle et sélectionnées d'après le test de Wald.

Le modèle final, comprenant les dépendances et les interactions source-variable présélectionnées, a été retenu selon la même stratégie. L'AIC et l'adéquation du modèle avec les données observées, basé sur la statistique du rapport de vraisemblance G^2 , ont permis de valider le choix du modèle retenu.

La méthode utilisée pour estimer les intervalles de confiance du nombre total de diagnostics a été inspirée par un travail de Gemmell [23]. Pour chaque strate i , 10 000 valeurs ont été générées d'après une simulation de distribution log-normale de moyenne μ_i et de variance σ^2_i , recueillies pour chaque strate. L'intervalle de confiance du nombre total de diagnostics d'infection à VIH chez les enfants de moins de 13 ans a été calculé en additionnant les 16 distributions log-normales puis en identifiant les 2,5^e et 97,5^e centiles.

L'exhaustivité a pu être obtenue dans chaque strate, en rapportant le nombre de cas observés au nombre de cas estimés dans le modèle final.

3.10 OUTILS D'ANALYSE STATISTIQUE

L'identification des cas communs a été réalisée par la procédure "SQL" de SAS[®] version 9.1. Les estimations ont été faites avec le logiciel STATA *via* le module "recap" et la commande "GLM" dédiée aux modèles linéaires généralisés, en spécifiant un lien logarithmique et une distribution de Poisson.

4. Résultats

4.1 ENQUÊTE AUPRÈS DU RÉSEAU LABOVIH

Cent trente-sept laboratoires d'analyses de biologie médicale de France métropolitaine (hors Bretagne) participant au réseau LaboVIH avaient signalé 287 diagnostics d'infection à VIH chez des enfants entre le 1^{er} janvier 2003 et le 31 décembre 2006.

Le questionnaire a été envoyé à ces laboratoires, ainsi qu'à 182 laboratoires bretons. Respectivement 82,5 % et 58,8 % d'entre eux ont répondu, soit 220 laboratoires (220=113+107). Huit laboratoires ont indiqué ne pas disposer de l'information demandée pour chaque enfant.

Au total, 132 nouveaux diagnostics d'infection à VIH chez des enfants âgés de moins de 13 ans ont été validés par les laboratoires sur l'ensemble du territoire métropolitain sur la période 2003-2006. La différence entre ce nombre de diagnostics validé par l'enquête et le nombre agrégé signalé dans le cadre de la surveillance de l'activité de dépistage du réseau LaboVIH était liée à : des diagnostics réalisés chez des enfants âgés de 13 à 15 ans qui n'ont pas été inclus dans l'étude ; des diagnostics faits chez des adultes et comptabilisés par erreur parmi les enfants de moins de 15 ans ; des diagnostics non confirmés (sérologies positives sans valeur diagnostique chez des enfants de moins de 18 mois).

4.2 DOUBLONS DE LA BASE LABOVIH

Trois doublons propres à la source LaboVIH ont été repérés. Pour chacun d'entre eux, l'une des deux unités constituant le doublon a été supprimée afin de ne pas le comptabiliser deux fois.

4.3 NOMBRE DE DIAGNOSTICS D'INFECTION À VIH

Les nombres de diagnostics d'infection à VIH chez des enfants âgés de moins de 13 ans en France métropolitaine entre le 1^{er} janvier 2003 et le 31 décembre 2006, recensés dans chacune des trois sources sont les suivants : 103 par le système de déclaration obligatoire, 101 par la cohorte EPF et 129 dans le réseau LaboVIH après avoir éliminé les trois doublons.

4.4 CARACTÉRISTIQUES DES ENFANTS

Les caractéristiques des enfants de moins de 13 ans nouvellement diagnostiqués comme infectés par le VIH et recensés en France métropolitaine entre 2003 et 2006 sont présentées dans le tableau 4.

Le nombre de diagnostics variait selon les sources et les années. En 2006, il était plus faible dans les sources DOVIH et EPF (respectivement 13 et 14) que dans la source LaboVIH (30), comparativement aux autres années.

La distribution des diagnostics d'infection à VIH n'était pas la même selon la région de diagnostic. La proportion de cas diagnostiqués en Île-de-France et recensés par la source EPF (78,2 %) était supérieure à celle de la source DOVIH (69,9 %) et de la source LaboVIH (63,6 %) ($p=0,04$).

Pour les sources où cette information était disponible (DOVIH et EPF), les enfants étaient plus souvent nés à l'étranger. Cependant, la proportion d'enfants nés à l'étranger était plus élevée dans la source DOVIH que dans la source EPF ($p=0,04$).

L'âge au diagnostic était supérieur à un an pour plus des deux tiers des enfants dans chaque source.

| TABLEAU 4 |

Répartition des nouveaux diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans dans chacune des sources (DOVIH, EPF et LaboVIH), France métropolitaine, 2003-2006

		DOVIH (n=103)		EPF (n=101)		LaboVIH (n=129)	
Année de diagnostic	2003	29	28,1 %	28	27,7 %	31	24,0 %
	2004	36	35,0 %	32	31,7 %	34	26,4 %
	2005	25	24,3 %	27	26,7 %	34	26,4 %
	2006	13	12,6 %	14	13,9 %	30	23,3 %
Sexe	Fille	57	55,3 %	54*	54,5 %	72*	56,3 %
	Garçon	46	44,7 %	45*	45,5 %	56*	43,8 %
Pays de naissance	France	35	34,0 %	47	46,5 %	-	-
	Étranger	68	66,0 %	54	53,5 %	-	-
Région de diagnostic	Île-de-France	72	69,9 %	79	78,2 %	82	63,6 %
	Province	31	30,1 %	22	21,8 %	47	36,4 %
Âge au diagnostic	Un an ou moins	21	20,4 %	34	33,7 %	37	28,7 %
	Plus d'un an	82	79,6 %	67	66,3 %	92	71,3 %

DOVIH : Système de déclaration obligatoire du VIH.

EPF : Enquête périnatale française.

LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires.

* Le sexe est inconnu pour 2 cas recensés par la source EPF et 1 cas recensé par la source LaboVIH.

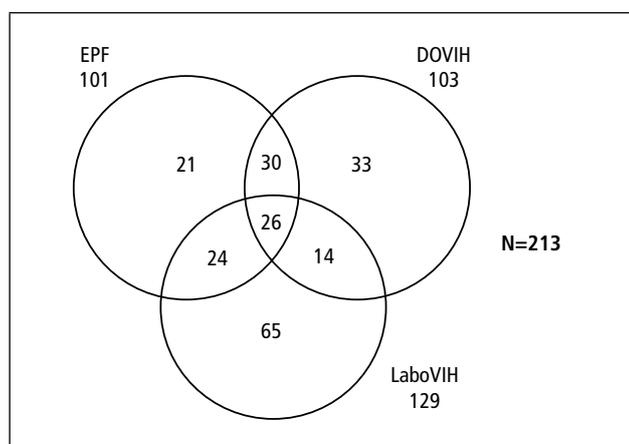
4.5 CAS COMMUNS AUX SOURCES

4.5.1 Situation 1 (cas certains)

Parmi les 103 cas notifiés dans la source DOVIH, 30 cas étaient communs à la source EPF et 14 à LaboVIH. Les sources EPF et LaboVIH avaient 24 nouveaux diagnostics en commun (figure 4). Au total, 68 diagnostics étaient recensés dans deux sources et 26 étaient communs aux trois sources. Après élimination des cas communs, 213 nouveaux diagnostics d'infection à VIH chez des enfants de moins de 13 ans ont été recensés par au moins l'une des trois sources entre le 1^{er} janvier 2003 et le 31 décembre 2006.

| FIGURE 4 |

Distribution du nombre de diagnostics d'infection à VIH selon la source, cas communs certains, France métropolitaine, 2003-2006



DOVIH: Système de déclaration obligatoire du VIH;
EPF: Enquête périnatale française;
LaboVIH: Surveillance de l'activité de dépistage du VIH des laboratoires.

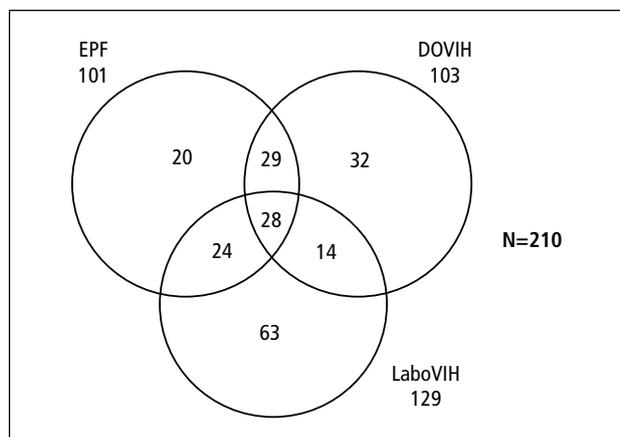
4.5.2 Situation 2 (cas certains+cas possibles)

La prise en compte des trois cas communs possibles qui n'ont pas pu être confirmés ou infirmés après consultation des fiches augmentait le nombre de nouveaux diagnostics communs dans les trois sources, de 26 à 28 (figure 5).

Les analyses ont été réalisées pour chacune des deux situations; les résultats étaient similaires. Les analyses suivantes sont donc présentées pour la situation n° 1, ne prenant en compte que les cas communs certains.

| FIGURE 5 |

Distribution du nombre de diagnostics d'infection à VIH selon la source, cas communs certains et probables, France métropolitaine, 2003-2006



DOVIH: Système de déclaration obligatoire du VIH;
EPF: Enquête périnatale française;
LaboVIH: Surveillance de l'activité de dépistage du VIH des laboratoires.

4.6 ESTIMATION DU NOMBRE TOTAL DE DIAGNOSTICS ET ÉVALUATION DE LA DÉPENDANCE

L'estimation était plus faible pour le couple de sources (DOVIH, EPF) ($\hat{N}=185$) que pour les deux autres couples de sources (tableau 5). Ces résultats suggéraient une dépendance positive entre la DOVIH et EPF. Cette dépendance était confirmée par les tests d'indépendance des sources (tableau 6). Le troisième rapport de cotes (couple LaboVIH, EPF) suggérait également une dépendance positive, bien qu'elle soit non significative à 5%.

| TABLEAU 5 |

Estimation du nombre total de diagnostics d'infection à VIH chez les enfants de moins de 13 ans, méthode capture-recapture en croisant les trois sources deux à deux, cas communs certains, France métropolitaine, 2003-2006

Source A, Source B	n _{source A}	n _{source B}	n _{source A, source B}	\hat{N}	IC 95 %
LaboVIH, DOVIH	129	103	40	329	[265-393]
EPF, DOVIH	101	103	56	185	[164-206]
LaboVIH, EPF	129	101	50	259	[220-298]

DOVIH: Système de déclaration obligatoire du VIH;
EPF: Enquête périnatale française;
LaboVIH: Surveillance de l'activité de dépistage du VIH des laboratoires.

Test d'indépendance entre les sources deux à deux, diagnostics d'infection à VIH chez les enfants de moins de 13 ans, cas communs certains, France métropolitaine, 2003-2006

(Source A, Source B)/Source C	OR	IC 95 %
(LaboVIH, DOVIH)/EPF	0,76	[0,35-1,67]
(EPF, DOVIH)/LaboVIH	5,03	[2,26-11,20]
(LaboVIH, EPF)/DOVIH	2,04	[0,90-4,62]

DOVIH : Système de déclaration obligatoire du VIH ;

EPF : Enquête périnatale française ;

LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ;

OR : Rapport de cotes ; IC 95 % : Intervalle de confiance à 95 %.

4.7 RÉAFFECTATION PROPORTIONNELLE DU PAYS DE NAISSANCE DANS LA BASE LABOVIH

Parmi les 148 diagnostics des sources EPF et DOVIH pour lesquels le pays naissance était renseigné, 91,3 % des enfants d'1 an ou moins et 19,6 % des enfants de plus d'1 an étaient nés en France.

Parmi les 129 diagnostics notifiés dans la source LaboVIH, 65 n'étaient pas communs aux deux autres sources et n'étaient donc pas renseignés pour le pays de naissance. Douze enfants avaient 1 an ou moins au moment du diagnostic et 53 avaient plus d'1 an.

La réaffectation proportionnelle du pays de naissance selon la classe d'âge a permis d'attribuer un lieu de naissance en France à 11 enfants d'1 an ou moins et à 11 enfants de plus d'1 an et une naissance à l'étranger à un enfant d'1 an ou moins et à 42 enfants de plus d'1 an.

4.8 MODÈLES LOG-LINÉAIRES

4.8.1 Dépendance entre les sources

Quatre modèles avaient une adéquation correcte et, parmi ces modèles, le plus parcimonieux était celui prenant en compte les dépendances entre les sources LaboVIH et EPF et entre DOVIH et EPF (tableau 7). Il estimait le nombre de nouveaux diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans à 366 (IC 95 %=[290-523]). Cette estimation était proche de celle pondérée par l'AIC (366 versus 336).

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans, méthode capture-recapture, cas communs certains, France métropolitaine, 2003-2006

Modèle	\hat{n}	\hat{N}	IC 95 %	ddl	G ²	p	AIC
LaboVIH*DOVIH, LaboVIH*EPF, DOVIH*EPF	116	329	[253-544]	0	0	1	0
LaboVIH*DOVIH, LaboVIH*EPF	23	236	[222-260]	1	16,70	<10 ⁻⁴	14,70
LaboVIH*DOVIH, DOVIH*EPF	57	270	[240-323]	1	3,01	0,08	1,01
LaboVIH*EPF, DOVIH*EPF	153	366	[290-523]	1	0,48	0,49	-1,52
LaboVIH*DOVIH, EPF	29	242	[227-265]	2	18,06	<10 ⁻⁴	14,06
LaboVIH*EPF, DOVIH	53	265	[242-298]	2	28,42	<10 ⁻⁴	24,42
DOVIH*EPF, LaboVIH	85	298	[265-345]	2	5,68	0,06	1,68
LaboVIH, DOVIH, EPF	51	264	[244-290]	3	28,45	<10 ⁻⁴	22,45
Weighted AIC		336					

DOVIH : Système de déclaration obligatoire du VIH ; EPF : Enquête périnatale française ; LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ; \hat{n} : Estimation du nombre de diagnostics recensés dans aucune des trois sources ; \hat{N} : Estimation du nombre de diagnostics ; IC 95 % : Intervalle de confiance à 95 % ; ddl : Degré de liberté ; G² : Déviance ; p : p-value ; AIC : Critère d'information d'Akaike ; Weighted AIC : Estimation de \hat{N} pondéré par l'AIC de tous les modèles ci-dessus.

4.8.2 Variables d'hétérogénéité

Les variables d'hétérogénéité ont été intégrées dans les modèles indépendamment les unes des autres (tableaux 8, 9 et 10). Ces étapes préliminaires ont servi de base pour le modèle final prenant en compte toutes les variables d'hétérogénéité.

En tenant compte du pays de naissance, trois modèles avaient une bonne adéquation et deux d'entre eux avaient un AIC négatif et égal (tableau 8). Suivant le principe de parcimonie, le modèle prenant en compte l'interaction entre le pays de naissance et la source EPF a été retenu. Celui-ci donnait une estimation du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans à 366.

Tous les modèles prenant en compte l'année de diagnostic avaient une bonne adéquation (tableau 9). Le modèle le plus parcimonieux prenait en compte l'interaction entre l'année de diagnostic et la source DOVIH. L'estimation obtenue était similaire à celle du modèle précédent et à l'estimation pondérée ($\hat{N}=366$).

Les quatre modèles tenant compte de la région de diagnostic étaient stables (tableau 10). Le modèle le plus parcimonieux prenait en compte une interaction entre la région de diagnostic et la source EPF ainsi que les deux dépendances initialement mises en évidence.

| TABLEAU 8 |

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans, méthode capture-recapture tenant compte du pays de naissance, cas communs certains, France métropolitaine, 2003-2006

Modèle	\hat{n}	\hat{N}	ddl	G ²	p	AIC
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Pays*DOVIH, Pays*EPF, Pays*LaboVIH	117	330	3	6,98	0,07	0,98
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Pays*DOVIH, Pays*EPF	116	329	4	7,38	0,12	-0,62
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Pays*EPF	116	329	5	10,90	0,05	0,90
LaboVIH*EPF, DOVIH*EPF, Pays*EPF	153	366	6	11,38	0,08	-0,62
Weighted AIC		342				

DOVIH : Système de déclaration obligatoire du VIH ; EPF : Enquête périnatale française ; LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ; Pays : Pays de naissance ; \hat{n} : Estimation du nombre de diagnostics recensés dans aucune des trois sources ; \hat{N} : Estimation du nombre de diagnostics ; IC 95 % : Intervalle de confiance à 95 % ; ddl : Degré de liberté ; G² : Déviance ; p : p-value ; AIC : Critère d'information d'Akaike ; Weighted AIC : Estimation de \hat{N} pondéré par l'AIC de tous les modèles ci-dessus.

| TABLEAU 9 |

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans, méthode capture-recapture tenant compte de l'année de diagnostic, cas communs certains, France métropolitaine, 2003-2006

Modèle	\hat{n}	\hat{N}	ddl	G ²	p	AIC
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Année*DOVIH, Année*EPF, Année*LaboVIH	116	329	9	13,27	0,15	-4,7
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Année*DOVIH, Année*LaboVIH	116	329	12	17,72	0,12	-6,28
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Année*DOVIH	116	329	15	21,88	0,11	-8,12
LaboVIH*EPF, DOVIH*EPF, Année*DOVIH	153	366	16	22,36	0,13	-9,65
Weighted AIC	137	350				

DOVIH : Système de déclaration obligatoire du VIH ; EPF : Enquête périnatale française ; LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ; Année : Année de diagnostic ; \hat{n} : Estimation du nombre de diagnostics recensés dans aucune des trois sources ; \hat{N} : Estimation du nombre de diagnostics ; IC 95 % : Intervalle de confiance à 95 % ; ddl : Degré de liberté ; G² : Déviance ; p : p-value ; AIC : Critère d'information d'Akaike ; Weighted AIC : Estimation de \hat{N} pondéré par l'AIC de tous les modèles ci-dessus.

| TABLEAU 10 |

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans, méthode capture-recapture tenant compte de la région de diagnostic, cas communs certains, France métropolitaine, 2003-2006

Modèle	\hat{n}	\hat{N}	ddl	G ²	p	AIC
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Région*DOVIH, Région*EPF, Région*LaboVIH	120	333	3	2,38	0,5	-3,62
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Région*DOVIH, Région*EPF	116	329	4	5,21	0,27	-2,8
LaboVIH*EPF, DOVIH*EPF, DOVIH*LaboVIH, Région*EPF	116	329	5	6,88	0,23	-3,14
LaboVIH*EPF, DOVIH*EPF, Région*EPF	153	366	6	7,33	0,29	-4,67
Weighted AIC		345				

DOVIH : Système de déclaration obligatoire du VIH ; EPF : Enquête périnatale française ; LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ; Région : Région de diagnostic ; \hat{n} : Estimation du nombre de diagnostics recensés dans aucune des trois sources ; \hat{N} : Estimation du nombre de diagnostics ; IC 95 % : Intervalle de confiance à 95 % ; ddl : Degré de liberté ; G² : Déviance ; p : p-value ; AIC : Critère d'information d'Akaike ; Weighted AIC : Estimation de \hat{N} pondéré par l'AIC de tous les modèles ci-dessus.

Sur la base des étapes précédentes de modélisation, un modèle prenant en compte les deux dépendances LaboVIH*EPF et DOVIH*EPF, les interactions entre le pays de naissance et la source EPF, l'année de diagnostic et la source DOVIH et la région de diagnostic et la source

EPF a été construit (tableau 11). L'estimation du nombre total de cas était de 364 pour ce modèle qui a permis de calculer l'exhaustivité des sources par strate pour les variables d'hétérogénéité. L'analyse pas à pas descendante n'a pas permis d'identifier d'autres modèles.

| TABLEAU 11 |

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans, méthode capture-recapture tenant compte du pays de naissance, de l'année de diagnostic et de la région de diagnostic, cas communs certains, France métropolitaine, 2003-2006

Modèle	\hat{n}	\hat{N}	IC 95 %	ddl	G ²	p	AIC
LaboVIH*EPF, DOVIH*EPF, Pays*EPF, Année*DOVIH, Région*EPF	151	364	[347-419]	96	119,45	0,06	-72,51

DOVIH : Système de déclaration obligatoire du VIH ; EPF : Enquête périnatale française ; LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ; Pays : Pays de naissance ; Année : Année de diagnostic ; Région : Région de diagnostic ; \hat{n} : Estimation du nombre de diagnostics recensés dans aucune des trois sources ; \hat{N} : Estimation du nombre de diagnostic ; IC 95 % : Intervalle de confiance à 95 % ; ddl : Degré de liberté ; G² : Déviance ; p : p-value ; AIC : Critère d'information d'Akaike ; Weighted AIC : Estimation de \hat{N} pondéré par l'AIC de tous les modèles ci-dessus.

Le modèle retenu a permis d'estimer à 364 le nombre total de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans sur la période d'étude, soit un nombre moyen d'environ 90 cas par an (tableau 12). L'exhaustivité globale des trois sources conjuguées était de 59 % et variait selon la source. La source LaboVIH avait la meilleure exhaustivité (35 %), tandis que celle des sources DOVIH et LaboVIH était plus faible (28 %).

L'exhaustivité variait selon le pays de naissance pour une même source et entre les trois sources. Elle était meilleure pour les enfants nés en France dans les sources EPF (34 %) et LaboVIH (40 %).

Pour les trois sources, l'exhaustivité était meilleure en Île-de-France qu'en Province. Cette différence était plus marquée pour la source EPF (37 % versus 15 %).

En ce qui concerne l'année de diagnostic, l'exhaustivité des sources DOVIH et EPF était particulièrement faible pour l'année 2006 (respectivement 15 % et 17 %).

| TABLEAU 12 |

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans et exhaustivité par strate selon l'année de diagnostic, le pays de naissance et la région de diagnostic, selon chaque source, méthode capture-recapture, cas communs certains, France métropolitaine, 2003-2006

Strate	\hat{N}	IC 95 %	Total		DOVIH		EPF		LaboVIH	
			N _{obs}	Exh (%)						
Année de diagnostic										
2003	104,2	[91,4-130,9]	61	58,5	29	27,8	28	26,9	31	30,0
2004	90,0	[80,3-110,3]	58	64,4	36	40,0	32	35,5	34	37,8
2005	85,2	[76,2-109,9]	51	60,0	25	29,3	27	31,7	34	39,9
2006	84,9	[72,5-110,7]	43	50,6	13	15,3	14	16,5	30	35,3
Pays de naissance										
France	137,9	[129,6-163,5]	88	63,8	35	25,4	47	34,1	55	40,0
Étranger	226,4	[203,2-269,9]	125	55,2	68	30,0	54	23,9	74	32,7
Région de diagnostic										
Île-de-France	214,0	[198,0-249,3]	136	63,6	72	33,6	79	36,9	82	38,3
Province	150,3	[136,9-187,0]	77	51,2	31	20,6	22	14,6	47	31,3
Total	364,3	[346,7-418,7]	213	58,5	103	28,3	101	27,7	129	35,4

DOVIH : système de déclaration obligatoire du VIH ; EPF : Enquête périnatale française ; LaboVIH : Surveillance de l'activité de dépistage du VIH des laboratoires ; \hat{N} : Estimation du nombre de diagnostic ; N_{obs} : Nombre de cas observé ; Exh : Exhaustivité.

Pour tenter d'expliquer la baisse de l'exhaustivité des sources DOVIH et EPF en 2006, des estimations ont été réalisées par strate de région de diagnostic, de lieu de naissance et d'année de diagnostic (tableau 13).

Cette stratification réduisait considérablement les effectifs dans chaque strate rendant difficile les interprétations. Néanmoins, on observait dans les trois sources une diminution du nombre de cas à partir de 2005 pour les enfants diagnostiqués en Île-de-France et nés à l'étranger.

Estimations du nombre de diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans et exhaustivité par strate de région de diagnostic, de pays de naissance et d'année de diagnostic, selon chaque source, méthode capture-recapture, cas communs certains, France métropolitaine, 2003-2006

Strate	N̂	IC 95 % (N̂)	Total		DOVIH		EPF		LaboVIH	
			N _{obs}	Exh (%)						
Région de diagnostic: province										
Pays de naissance: étranger										
2003	32,0	[24,5-48,2]	18	56,3	7	21,9	7	21,9	9	28,1
2004	19,4	[13,7-32,1]	9	46,4	6	30,9	1	5,2	5	25,8
2005	24,8	[18,4-38,8]	13	52,4	7	28,2	6	24,2	8	32,2
2006	24,6	[17,3-40,3]	11	44,7	3	12,2	0	0,0	10	40,7
Pays de naissance: France										
2003	13,2	[9,2-22,0]	6	45,5	2	15,2	1	7,6	3	22,7
2004	11,3	[8,3-18,2]	6	53,0	3	26,5	2	17,7	3	26,5
2005	12,0	[10,7-21,6]	8	66,7	1	8,3	4	33,3	5	41,7
2006	13,0	[9,1-21,5]	6	46,2	2	15,4	1	7,7	4	30,8
Région de diagnostic: Île-de-France										
Pays de naissance: étranger										
2003	40,5	[32,7-57,3]	26	64,2	16	39,5	14	34,6	13	32,0
2004	32,8	[26,9-45,9]	22	67,1	15	45,7	14	42,7	13	39,6
2005	28,2	[21,6-42,6]	16	56,7	9	31,9	7	24,8	9	31,9
2006	24,1	[16,5-30,4]	10	41,5	4	16,6	4	16,6	7	29,0
Pays de naissance: France										
2003	18,5	[14,4-27,6]	11	59,5	5	27,0	7	37,8	6	32,4
2004	26,5	[23,4-33,6]	21	79,2	11	41,5	14	52,8	13	49,1
2005	20,2	[16,8-28,1]	14	69,3	8	39,6	10	49,5	12	59,4
2006	23,2	[19,2-32,1]	16	69,0	4	17,2	9	38,8	9	38,8
Total	364,3	[346,7-418,7]	213	58,5	103	28,3	101	27,7	129	35,4

DOVIH: Système de déclaration obligatoire du VIH; EPF: Enquête périnatale française; LaboVIH: Surveillance de l'activité de dépistage du VIH des laboratoires; N̂: Estimation du nombre de diagnostic; IC 95%: Intervalle de confiance à 95%; N_{obs}: Nombre de cas observé; Exh: Exhaustivité.

5. Discussion

À notre connaissance, il s'agit de la première étude utilisant la méthode capture-recapture pour estimer le nombre d'enfants nouvellement diagnostiqués comme infectés par le VIH, que ce soit en France ou dans un pays développé.

5.1 ESTIMATIONS ET EXHAUSTIVITÉ

Les estimations du nombre de diagnostics d'infection à VIH chez les enfants de moins de 13 ans variaient légèrement autour de 364 sur la période 2003-2006 selon le modèle considéré.

Sur la base du modèle prenant en compte les deux dépendances intersources DOVIH*EPF et LaboVIH*EPF, les interactions année de diagnostic*DOVIH, région*EPF et pays de naissance*EPF, l'exhaustivité de la déclaration obligatoire était de 28 %, celle de la cohorte EPF de 28 % et celle du réseau LaboVIH de 35 %.

L'exhaustivité de la déclaration obligatoire du VIH était peu satisfaisante pour les enfants. Elle est meilleure sur l'ensemble de la population (65 %) et pourrait être en partie expliquée par la participation des biologistes à la déclaration obligatoire depuis sa mise en place en 2003, ce qui permet de relancer le clinicien en cas de non envoi de son feuillet médical [6]. L'exhaustivité de la déclaration des diagnostics d'infection à VIH chez les enfants âgés de moins de 13 ans devrait s'être améliorée depuis 2007, suite à l'implication des biologistes dans le système de surveillance.

L'exhaustivité relativement faible de la source EPF pouvait être expliquée par un défaut de couverture géographique et une concentration de sites participants en Île-de-France. Par ailleurs, l'exhaustivité étonnamment faible pour les enfants nés en France pouvait s'expliquer si la séropositivité des mères n'était pas connue à la naissance. Cette situation implique le plus souvent un diagnostic tardif chez l'enfant et une probabilité d'inclusion moindre dans la cohorte EPF. Le même phénomène était également à l'origine de la faible exhaustivité observée chez les enfants nés à l'étranger.

L'exhaustivité de la source LaboVIH était peu satisfaisante, alors que le taux de participation des laboratoires au réseau LaboVIH est de 85 % et que les laboratoires participants représentent 98 % des diagnostics positifs d'infection à VIH (information validée systématiquement auprès des laboratoires) (données non publiées, InVS, F. Cazein). Cependant, l'information concernant spécifiquement le nombre de diagnostics chez les enfants âgés de moins de 15 ans était mal renseignée et non validée. L'enquête complémentaire pour constituer la troisième source n'ayant ciblé que les laboratoires qui avaient signalé au moins un diagnostic chez un enfant âgé de moins de 15 ans n'a pas concerné les laboratoires qui avaient réalisé ces diagnostics sans les signaler.

5.2 OBTENTION D'UNE TROISIÈME SOURCE

Cette étude a nécessité la création d'une troisième source afin de fournir des estimations par des modèles log-linéaires tenant compte des dépendances entre les sources.

Le taux de réponse à notre enquête a été satisfaisant : supérieur à 80 % pour les laboratoires qui avaient indiqué avoir nouvellement diagnostiqué un cas d'infection à VIH chez un enfant de moins de 15 ans et proche de 60 % sur l'ensemble des laboratoires bretons. Par ailleurs, les laboratoires d'analyse de biologie médicale qui n'ont pas répondu avaient une faible activité de dépistage et la Bretagne est une région où le nombre d'infections à VIH diagnostiquées annuellement est le plus faible [7].

Cette étude a permis de valider les données agrégées recueillies par la surveillance de l'activité de dépistage des laboratoires du réseau LaboVIH. Ainsi, parmi les 287 diagnostics VIH signalés chez des enfants par les laboratoires répondants, 132 (59,2 %) ont été validés. Les résultats de cette étude pourraient être utilisés pour évaluer la qualité des données du réseau LaboVIH.

5.3 RÉAFFECTATION PROPORTIONNELLE DU PAYS DE NAISSANCE

L'épidémiologie de l'infection à VIH diffère selon l'origine géographique. Il était donc nécessaire de pouvoir fournir des estimations du nombre de cas selon le pays de naissance de l'enfant. Par ailleurs, la répartition des cas selon le pays de naissance dans chaque source laissait penser qu'il pouvait être une variable d'hétérogénéité.

Cependant, le pays de naissance n'était pas disponible auprès des laboratoires d'analyses biologiques et médicales. Les proportions observées d'enfants nés en France et à l'étranger dans les sources EPF et DOVIH ont été appliquées à la source LaboVIH par classe d'âge et sous l'hypothèse d'une distribution de l'âge au diagnostic identique selon le pays de naissance. Cette hypothèse était raisonnable compte tenu d'une distribution d'âge similaire entre les trois sources. Par ailleurs, la proportion d'enfants nés à l'étranger dans les deux classes d'âge était similaire dans les deux sources DOVIH et EPF ($p=0,4$).

Une réaffectation proportionnelle a été réalisée en considérant la distribution de l'âge au diagnostic selon le pays de naissance dans les deux sources pour lesquelles l'information était disponible [24]. Il existe, par ailleurs, des méthodes spécifiques à la méthode capture-recapture pour l'imputation de variables, qui n'ont pas été appliquées ici [25]. Les résultats de la réaffectation proportionnelle étaient comparables à ceux obtenus par imputation multiple (respectivement 57 % et 62 % des 129 enfants diagnostiqués dans la base LaboVIH étaient nés en France).

5.4 HYPOTHÈSES

5.4.1 Identification des vrais cas

Cette définition était respectée grâce à une définition de cas commune aux trois sources, reposant sur un diagnostic de confirmation biologique (examen virologique ou sérologie) sur deux prélèvements différents, conformément aux recommandations en vigueur [9].

5.4.2 Identification de tous les vrais cas communs

Il est rare en épidémiologie de disposer d'identifiants communs et uniques dans différentes bases. L'identification des cas communs à plusieurs sources repose le plus souvent sur des combinaisons de critères [26].

Alors que l'identification des cas communs aux sources DOVIH et EPF a été effectuée sans trop de difficultés du fait d'un nombre satisfaisant de variables communes disponibles, l'identification des cas communs concernant la source LaboVIH s'est faite sur un nombre restreint de variables communes.

Cependant, le faible nombre de cas a permis de vérifier manuellement les cas communs potentiels proposés par les algorithmes pour les confirmer ou non. Cette procédure a permis d'exclure 24 cas communs faussement positifs d'après les algorithmes. Seuls trois cas potentiellement communs non pas pu être confirmés. Cependant, il est possible que les algorithmes n'aient pas permis d'identifier des cas communs, notamment avec la source LaboVIH, ce qui pourrait être à l'origine d'une surestimation du nombre total de cas.

5.4.3 Même période et même zone d'étude

La période d'étude 2003-2006 était la même pour les trois sources.

Le recueil de données était prospectif pour les sources DOVIH et EPF, alors qu'il était rétrospectif pour la troisième source. Cependant, le nombre de cas identifiés dans la source LaboVIH était supérieur au nombre de cas recensés dans les sources DOVIH et EPF.

L'existence d'un délai de déclaration pour la DOVIH pouvait être à l'origine d'une incomplétude de la base en 2006. Néanmoins, le redressement pour tenir compte de ce délai de déclaration a permis d'estimer un ou deux cas supplémentaires par rapport aux treize déjà inclus dans la base pour 2006.

Malgré une forte prévalence du VIH dans les départements français d'Amérique (Guyane et Guadeloupe notamment), la zone d'étude a été restreinte à la France métropolitaine pour les trois sources, en raison du faible nombre d'établissements participant à EPF dans les DOM. Par ailleurs, il existait un défaut de couverture de la source EPF avec une probabilité d'inclusion nulle de cas dans certaines zones en Province.

5.4.4 Homogénéité de capture

La probabilité de capture des diagnostics d'infection à VIH selon le pays de naissance, la région de diagnostic et l'année de diagnostic variait

au sein de chaque source. Ces résultats indiquaient une hétérogénéité de capture liée à ces variables.

La prise en compte de ces trois variables a peu modifié les estimations du nombre total de cas : de 366 pour le modèle ne prenant en compte que les dépendances (entre les sources EPF et DOVIH, d'une part, et entre LaboVIH et EPF, d'autre part) à 364 pour le modèle final retenu prenant en compte les interactions entre EPF et pays de naissance, EPF et région de diagnostic, et DOVIH et année de diagnostic.

La probabilité d'inclusion des cas dans la cohorte EPF était très liée au pays de naissance et à la région de diagnostic. D'une part, les sites participants à EPF étaient plus nombreux en Île-de-France et, d'autre part, l'inclusion des enfants nés en France était meilleure, même si l'exhaustivité pour cette variable restait basse en Île-de-France.

La diminution de l'exhaustivité des sources DOVIH et EPF en 2006, amorcée en 2005, pouvait être expliquée par une réduction du nombre d'enfants diagnostiqués en Île-de-France et nés à l'étranger. Cette baisse observée à partir de 2005 dans les trois sources pouvait être une conséquence de la politique d'immigration avec un nombre d'enfants étrangers infectés entrants inférieur à celui des années précédentes.

5.4.5 Indépendance entre les sources

Une dépendance positive, suspectée entre les sources DOVIH et EPF, a été confirmée dans les modèles log-linéaires. Elle s'expliquait par la sensibilisation des médecins participant déjà à la cohorte EPF, au système de la déclaration obligatoire du VIH lors de sa mise en place en 2003.

Bien que cela n'ait pas été envisagé *a priori*, une dépendance positive a également été retrouvée entre LaboVIH et EPF. Cette dépendance a pu être créée par la relance spécifique de deux laboratoires hospitaliers parisiens (Necker et Bichat) participant à la cohorte EPF et ayant déclaré un nombre important de diagnostics chez les enfants via l'enquête LaboVIH.

5.4.6 Population d'étude close

Cette condition suppose qu'il n'y ait pas de mouvement de population (ni entrée, ni sortie) afin d'assurer une équiprobabilité de capture des cas dans chaque source. Cette condition a pu affecter la probabilité d'inclusion dans la source EPF. Les enfants repartant dans leur pays d'origine rapidement après un diagnostic d'infection à VIH avaient une probabilité moindre d'être inclus dans l'EPF. Cependant, ces situations sont probablement exceptionnelles en raison d'une faible disponibilité des traitements antirétroviraux dans certains pays, notamment en Afrique subsaharienne.

5.4.7 Alternative aux modèles log-linéaires

Les modèles écologiques sont une alternative intéressante aux modèles log-linéaires pour la prise en compte des variables d'hétérogénéité [27,28]. Par ailleurs, la méthode dite Sample Coverage Approach permet d'estimer la taille d'une population d'étude à partir de l'exhaustivité des sources [29]. Ces alternatives sont présentées et discutées par Chao [30].

6. Conclusion

Cette étude originale a permis de fournir pour la première fois une estimation du nombre annuel de diagnostics d'infection à VIH chez les enfants de moins de 13 ans. Il est néanmoins possible que ce nombre soit surestimé, ce qui nécessiterait de le confronter aux données de terrain.

Les recommandations en matière de dépistage des femmes enceintes, répété en fin de grossesse en cas de situations à risque, ont toute leur place pour optimiser la prévention de la TME.

La prévalence élevée dans certaines régions du monde, notamment en Afrique subsaharienne, et la proportion importante de diagnostics chez des enfants nés à l'étranger dans notre étude, justifierait de proposer un dépistage aux enfants qui arrivent en France, afin qu'ils bénéficient des traitements nécessaires le plus tôt possible.

Par ailleurs, une incitation à la déclaration des nouveaux diagnostics d'infection par le VIH chez l'enfant est nécessaire afin d'améliorer l'exhaustivité du système de déclaration obligatoire et de mieux décrire l'épidémiologie de l'infection chez les enfants.

Références bibliographiques

- [1] Warszawski J, Tubiana R, Le Chenadec J, Blanche S, Teglas JP, Dollfus C, *et al.* Mother-to-child HIV transmission despite antiretroviral therapy in the ANRS French perinatal cohort. *AIDS* 2008;22:289-99.
- [2] Vilain A, de Peretti C, Herbet JB, Blondel B. La situation périnatale en France en 2003. Premiers résultats de l'Enquête nationale périnatale. *Études et résultats (Drees)* 2005;383.
- [3] Macassa E, Burgard M, Veber F, Picard C, Neven B, Malhaoui N, *et al.* Characteristics of HIV-infected children recently diagnosed in Paris, France. *Eur J Pediatr* 2006;165:684-7.
- [4] Onusida. Rapport sur l'épidémie mondiale de sida. 2006.
- [5] Institut de veille sanitaire. Données de la déclaration obligatoire sida-VIH chez les enfants/adolescents (données du 31 décembre 2006). Disponible sur : www.invs.sante.fr/publications/2007/vih_sida_jeunes/notification_enfants_ado.ppt
- [6] Guide "déclarer, agir, prévenir". Le nouveau dispositif de surveillance des maladies à déclaration obligatoire, 2003. Saint-Maurice (Fra) : Institut de veille sanitaire, janvier 2003, 64 p. Disponible sur : www.invs.sante.fr
- [7] Cazein F. Activité de dépistage du VIH – Lutte contre le VIH/sida et les infections sexuellement transmissibles en France – 10 ans de surveillance, 1996-2005. Saint-Maurice (Fra) : Institut de veille sanitaire, mars 2007, 156 p. Disponible sur : www.invs.sante.fr
- [8] Blanche S, Rouzioux C, Moscato ML, Veber F, Mayaux MJ, Jacomet C, *et al.* A prospective study of infants born to women seropositive for human immunodeficiency virus type 1. HIV Infection in newborns french collaborative study group. *N Engl J Med* 1989;320:1643-8.
- [9] Yeni P. Prise en charge médicale des personnes infectées par le VIH. Médecine-Sciences Flammarion, 2008.
- [10] Wittes J, Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *J Chronic Dis* 1968;21(5):287-301.
- [11] Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chronic Dis* 1974;27(1):25-36.
- [12] Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;17(2):243-64.
- [13] Capture-recapture and multiple-record systems estimation I: history and theoretical development. International working group for disease monitoring and forecasting. *Am J Epidemiol* 1995;142(10):1047-58.
- [14] Capture-recapture and multiple-record systems estimation II: applications in human diseases. International working group for disease monitoring and forecasting. *Am J Epidemiol* 1995;142(10):1059-68.
- [15] Sekar CC, Deming WE. On a method of estimating birth and death rates and the extent of registration. *American statistical association journal* 1949.
- [16] Chapman DG. Some properties of the hypergeometric distribution. *University of California* 1951;1(7):131-60.
- [17] 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.
- [18] Bishop YMM, Fienberg SE, Holland PW. Estimating the size of a closed population. *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, 1975. p. 229-56.
- [19] Regal RR, Hook EB. Goodness-of-fit based confidence intervals for estimates of the size of a closed population. *Stat Med* 1984;3(3):287-91.
- [20] Cazein F. Surveillance du VIH : notification obligatoire et surveillance virologique – Lutte contre le VIH/sida et les infections sexuellement transmissibles en France – 10 ans de surveillance, 1996-2005. Saint-Maurice (Fra) : Institut de veille sanitaire, mars 2007, 156 p. Disponible sur : www.invs.sante.fr

- [21] Jasseron C, Mandelbrot L, Tubiana R, Teglas JP, Faye A, Dollfus C, *et al.* Prevention of mother-to-child HIV transmission: similar access for sub-Saharan African immigrants and for French women? *AIDS* 2008;22(12):1503-11.
- [22] Stata module to perform capture-recapture analysis for three sources with goodness-of-fit based confidence intervals [computer program]. 2007.
- [23] Gemmell I, Millar T, Hay G. Capture-recapture estimates of problem drug use and the use of simulation based confidence intervals in a stratified analysis. *J Epidemiol community health* 2004;58(9):758-65.
- [24] Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of *ad hoc* methods of correcting for missing values for confounding variables. *Am J Epidemiol* 1991;134(8):895-907.
- [25] Zwane EN, Van Der Heijden PG. Analysing capture-recapture data when some variables of heterogeneous catchability are not collected or asked in all registrations. *Stat Med* 2007;26(5):1069-89.
- [26] Brenner H. Application of capture-recapture methods for disease monitoring: potential effects of imperfect record linkage. *Methods Inf Med* 1994;33(5):502-6.
- [27] Tilling K, Sterne JA. Capture-recapture models including covariate effects. *Am J Epidemiol* 1999;149(4):392-400.
- [28] Alho JM. Logistic regression in capture-recapture models. *Biometrics* 1990;46(3):623-35.
- [29] Lee SM, Chao A. Estimating population size *via* sample coverage for closed capture-recapture models. *Biometrics* 1994;50:88-97.
- [30] Chao A, Tsay PK, Lin SH, Shau WY, Chao DY. The applications of capture-recapture models to epidemiological data. *Stat Med* 2001;20(20):3123-57.

Annexe 2 – Identification des cas communs entre deux sources

RECHERCHE DES CAS COMMUNS DOVIH – EPF

Algorithme 1a:

- Sexe
- Année de naissance
- Pays de naissance ou maternité si la naissance a eu lieu en France
- Date de diagnostic ou de prise en charge (mois-année) à un mois près
- Hôpital de suivi
- Origine de la mère
- Statut vital

25 cas communs identifiés ; 24 validés comme certains.

Algorithme 1b:

- Année de naissance
- Date de diagnostic (mois-année) à trois mois près
- Hôpital de suivi
- Statut vital = décédé

2 cas communs identifiés ; 2 validés comme certains.

Algorithme 2:

- Pays de naissance ou maternité si la naissance a eu lieu en France (5 doublons)
- Date de diagnostic ou de prise en charge (mois-année) à trois mois près (1 doublon)
- Hôpital de suivi (5 doublons)
- Origine de la mère (5 doublons)

16 cas communs identifiés ; 16 validés comme certains.

Algorithmes 3a:

- Sexe
 - Année de naissance
 - Pays de naissance ou maternité si la naissance a eu lieu en France
- Deux critères non remplis parmi :
- Date de diagnostic ou de prise en charge (mois-année) à un mois près
 - Hôpital de suivi
 - Origine de la mère

10 cas communs identifiés ; 9 validés comme certains.

Algorithme 3b:

- Sexe
 - Année de naissance
 - Date de diagnostic ou de prise en charge (mois-année) à un mois près
- Deux critères non rempli parmi :
- Origine de la mère
 - Pays de naissance
 - Hôpital de suivi

4 cas communs identifiés ; 3 validés comme certains.

Algorithme 3c:

- Année de naissance
 - Date de diagnostic ou de prise en charge (mois-année) à un mois près
 - Hôpital de suivi
- Deux critères non rempli parmi :
- Sexe
 - Origine de la mère
 - Pays de naissance

3 cas communs identifiés; 2 validés comme certains.

RECHERCHE DES CAS COMMUNS LABOVIIH – EPF

Algorithme 1 :

- Sexe
- Année de naissance
- Date de diagnostic ou de prise en charge (jour-mois-année)
- Hôpital de suivi
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

27 cas communs identifiés; 27 validés comme certains.

Algorithmes 2a:

- Sexe
- Année de naissance
- Date de diagnostic ou de prise en charge (jour-mois-année)
- **Département de l'hôpital de suivi ou du laboratoire**
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

5 cas communs identifiés; 5 validés comme certains.

Algorithme 2b :

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année)**
- Hôpital de suivi
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

2 cas communs identifiés; 2 validés comme certains.

Algorithmes 3a:

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année)**
- **Département de l'hôpital de suivi ou du laboratoire**
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

3 cas communs identifiés; 3 validés comme certains.

Algorithme 3b:

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois année) à un mois près**
- Hôpital de suivi
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

5 cas communs identifiés ; 3 validés comme certains.

Algorithmes 4:

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année) à un mois près**
- **Département de l'hôpital de suivi ou du laboratoire**
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

6 cas communs identifiés ; 5 validés comme certains.

Algorithme 5a:

- Sexe
- Année de naissance
- Hôpital de suivi
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

6 cas communs identifiés ; 2 validés comme certains.

Algorithme 5b:

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année)**
- Âge au diagnostic plus ou moins un mois pour les moins de 24 mois

6 cas communs identifiés ; 3 validés comme certains.

RECHERCHE DES CAS COMMUNS DOVIH – LABOVIH

Algorithme 1:

- Sexe
- Année de naissance
- Date de diagnostic ou de prise en charge (mois-année)
- Hôpital de suivi

21 cas communs identifiés ; 21 validés comme certains.

Algorithmes 2:

- Sexe
- Année de naissance
- Date de diagnostic ou de prise en charge (mois-année)
- **Département de l'hôpital de suivi ou du laboratoire**

8 cas communs identifiés ; 7 validés comme certains.

Algorithme 2b :

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année) à un mois près**
- Hôpital de suivi

5 cas communs identifiés; 5 validés comme certains.

Algorithme 3a :

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année) à un mois près**
- **Département de l'hôpital de suivi**

7 cas communs identifiés; 2 validés comme certains.

Algorithme 3b :

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année)**

2 cas communs identifiés; 1 validé comme certain.

Algorithmes 3c :

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année) à un mois près**

2 cas communs identifiés; 2 validés comme certains.

Algorithme 3d :

- Sexe
- Année de naissance
- **Date de diagnostic ou de prise en charge (mois-année) à deux mois près**
- Hôpital de suivi

5 cas communs identifiés; 2 validés comme certains.

Estimation du nombre de nouveaux diagnostics d'infection par le VIH chez les enfants par la méthode capture-recapture, France métropolitaine, 2003-2006

Cette étude propose d'utiliser la méthode capture-recapture pour fournir une estimation du nombre de nouveaux diagnostics d'infection à VIH chez les enfants, en France métropolitaine, sur la période 2003-2006. Trois sources d'information ont été utilisées : le système de déclaration obligatoire du VIH (DOVIH), l'Enquête périnatale française (EPF) et la surveillance de l'activité de dépistage des laboratoires (LaboVIH). Les estimations ont été produites par des modèles log-linéaires prenant en compte les dépendances entre les sources et les variables d'hétérogénéité.

Deux cent treize cas ont été recensés par l'une au moins des trois sources. Le modèle log-linéaire retenu a pris en compte les dépendances entre les sources DOVIH et EPF, EPF et LaboVIH, et les interactions entre EPF et la région de diagnostic, EPF et le pays de naissance, et DOVIH et l'année de diagnostic. Le nombre de nouveaux diagnostics d'infection à VIH chez les enfants de moins de 13 ans a ainsi été estimé à 364 sur la période 2003-2006, dont 62 % étaient nés à l'étranger. L'exhaustivité de la DOVIH était estimée à 28 %.

Malgré une surestimation possible, qui nécessiterait de confronter les résultats aux données de terrain, cette étude a permis d'estimer pour la première fois un nombre annuel moyen de 90 nouveaux diagnostics d'infection à VIH chez l'enfant en France métropolitaine. Les recommandations pour le dépistage des femmes enceintes, répété si besoin en fin de grossesse, ont toute leur place pour optimiser la prévention de la transmission materno-foetale. La prévalence élevée dans certaines régions du monde, notamment en Afrique subsaharienne, justifierait de proposer un dépistage du VIH aux enfants qui arrivent en France, afin qu'ils bénéficient d'une prise en charge adaptée. L'implication des biologistes depuis 2007 et la sensibilisation des pédiatres devrait augmenter l'exhaustivité du système de la déclaration obligatoire du VIH chez l'enfant.

Mots clés : capture-recapture, modèles log-linéaires, VIH, diagnostics, enfants

Estimation of the number of new HIV diagnoses in children using the capture-recapture method, France, 2003-2006

This study aimed to estimate the number of new HIV diagnoses in children less than 13 years old using the capture-recapture method during the 2003-2006 period in mainland France. Three sources were used: the mandatory notification system (DOVIH), the French Perinatal Cohort (EPF) and the Laboratory Network (LaboVIH). The estimations were obtained from Log Linear modelling including dependencies between sources and variable of heterogeneity.

The three sources collected 213 new HIV diagnoses in children. The selected model took into account dependencies between sources DOVIH and EPF, EPF and LaboVIH and the interactions between EPF and region of diagnosis, EPF and place of birth, and DOVIH and year of diagnosis. The estimated number of new HIV diagnoses in children less than 13 years was 364 between 2003 and 2006, 62% of whom were born in a foreign country. The exhaustiveness of DOVIH was estimated to be 28%.

Even if an overestimation is possible which would require comparing the results with field data, this study estimated for the first time the average annual number of 90 new HIV diagnoses in children under than 13 years old in mainland France. Recommendations for screening of pregnant women, repeated if necessary at the end of pregnancy, are important to optimize the prevention of mother-to-child transmission. As HIV prevalence is quite high in some regions in the world like sub-Saharan Africa, HIV screening should be proposed to children who emigrate in France in order to access to care. The exhaustiveness of the mandatory notification system should be improved with the biologist notifications implemented in 2007 and the sustained effort of heightening of paediatricians.

Citation suggérée :

Estimation du nombre de diagnostics d'infection par le VIH chez les enfants par la méthode capture-recapture, France métropolitaine, 2003-2006 – Rapport de master. Saint-Maurice (Fra) : Institut de veille sanitaire, juillet 2010, 27 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 : 978-2-11-099281-9

Réalisé par Diadeis-Paris

Dépôt légal : juillet 2010