

Evaluation du système de surveillance de la tuberculose par la méthode capture-recapture Région Rhône-Alpes, 1999-2000



MINISTERE DE LA SANTE ET DE LA
PROTECTION SOCIALE
Drass Rhône-Alpes
Cire



Institutions ayant contribué à l'investigation

Cire Rhône-Alpes

S. Rey, B. Helynck, F. Long

Institut de Veille Sanitaire

B. Decludt, D. Che

Société Régionale de santé publique

C. Ducos, M. Fortin, Y. Hemon, L. Tiberghien, M.C. Bruel

Rapport rédigé par Sylvie Rey

Remerciements : les auteurs remercient de leur collaboration les services de microbiologie des départements de la région Rhône-Alpes, le Dr Schorle de la direction régionale du service médical de la CNAMTS, les médecins des services médicaux des régimes d'assurance maladie et les médecins inspecteurs et les infirmières de santé publique des Ddass de la région Rhône-Alpes.

SOMMAIRE

1	Contexte	6
1.1	La surveillance de la tuberculose en France	6
1.2	Justification de l'étude	6
1.3	Diagnostic biologique	7
2	Objectifs de l'étude.....	7
3	Matériel et méthodes.....	7
3.1	Population d'étude et recueil des données.....	7
3.2	Matériel	7
3.2.1	La déclaration obligatoire	7
3.2.2	Enquête auprès des laboratoires	8
3.2.3	Enquête auprès des caisses d'assurance maladie	9
3.3	Méthode capture-recapture	9
3.3.1	Principes	9
3.3.2	Estimation du nombre total de cas avec deux sources	11
3.3.3	Conditions d'application de la méthode	12
3.3.4	Evaluation de la dépendance en utilisant la méthode de Wittes	13
3.3.5	Estimation du nombre total de cas en utilisant les modèles log-linéaires	14
3.3.6	Identification des cas communs intra-sources	15
3.3.7	Identification des cas communs entre les sources	15
3.3.8	Analyse statistique.....	16
3.3.9	Aspects éthiques	16
4	Résultats.....	17
4.1	Etude descriptive	17
4.1.1	Description des sources	17
4.1.2	Caractéristiques des cas notifiés par les trois sources	20
4.2	Distribution des cas déclarés dans les trois sources et identification des cas communs	21
4.2.1	Ensemble des cas en Rhône-Alpes, période 1999-2000	21
4.2.2	Cas bactériologiquement confirmés en Rhône-Alpes, années 1999-2000	21
4.3	Estimation du nombre total de cas et évaluation de la dépendance entre les sources en utilisant les méthodes de Wittes	22
4.3.1	Estimation du nombre total de cas et de l'exhaustivité en croisant les sources deux à deux	22
4.3.2	Evaluation de la dépendance entre les sources par estimation des rapports des cotes ..	26
4.3.3	Estimation du nombre total de cas après fusion des deux sources dépendantes	28
4.4	Estimation du nombre total de cas et évaluation de la dépendance entre les sources en utilisant les modèles log-linéaires pour l'ensemble des cas de tuberculose	29
4.4.1	Sélection du meilleur modèle par stratégie pas à pas descendante pour l'ensemble des cas déclarés par les 3 sources	29
4.4.2	Introduction des variables d'hétérogénéité pour l'ensemble des cas déclarés par les 3 sources	30
4.4.3	Modèles retenus et estimations finales du nombre total de cas et de l'exhaustivité des sources pour l'ensemble des cas déclarés par les 3 sources	32

4.5 Estimation du nombre total de cas et évaluation de la dépendance entre les sources en utilisant les modèles log-linéaires pour les cas bactériologiquement confirmés.....	33
4.5.1 Sélection du meilleur modèle par stratégie pas à pas descendante pour les cas bactériologiquement confirmés, déclarés par les 3 sources	33
4.5.2 Introduction des variables d'hétérogénéité pour les cas bactériologiquement confirmés déclarés par les 3 sources.....	34
4.5.3 Modèles retenus et estimations finales du nombre total de cas et de l'exhaustivité des sources pour les cas bactériologiquement confirmés, déclarés par les 3 sources	35
4.6 Analyse de l'hétérogénéité d'exhaustivité pour les sources DO et CAM	37
5 Discussion.....	37
5.1 Synthèse des principaux résultats.....	37
5.2 Conditions d'application de la méthode capture-recapture	38
5.3 Analyse des résultats	41
6 Conclusion	42

RESUME

La surveillance nationale des cas de tuberculose est basée sur le système de déclaration obligatoire (DO). En 1992-1993, une estimation de l'exhaustivité de la déclaration obligatoire de cette maladie avait été réalisée par la méthode capture-recapture sur un échantillon de 15 départements. Elle estimait à 43 % l'exhaustivité du système. L'objectif de cette étude était de suivre l'évolution du système de surveillance par une nouvelle estimation de son exhaustivité par la même méthode et de ses caractéristiques pour l'ensemble de la région Rhône-Alpes.

La méthode capture-recapture à 3 sources a été appliquée en utilisant les modèles log-linéaires et en prenant en compte des variables d'hétérogénéité pour l'ensemble des cas diagnostiqués et pour le sous-ensemble des cas bactériologiquement confirmés du fait des différences de définition des cas entre les trois sources. Cette méthode a permis d'estimer le nombre total de cas et l'exhaustivité de chacune des 3 sources de données. Pour ce faire, un recueil rétrospectif des cas de tuberculose ayant conduit à une exonération du ticket modérateur a été réalisé auprès des services médicaux des caisses d'assurance maladie (CAM) ainsi qu'un recueil des cas identifiés auprès de l'ensemble des laboratoires publics et privés d'analyses de biologie médicale. Ces données ont été croisées avec les cas déclarés aux Ddass.

Après élimination des doublons, un total de 1306 cas de tuberculose diagnostiquée et de 947 cas bactériologiquement positifs ont été identifiés. Selon le meilleur modèle log-linéaire incluant des interactions entre les sources DO et CAM et DO et LABM, l'estimation totale du nombre de cas diagnostiqués était de 1767 cas. L'exhaustivité de la DO, probablement sous-estimée du fait des différences de définition de cas entre les sources, était de 44 % [41-49]. Pour les cas à localisation pulmonaire, elle était de 51 % [47-55] et elle était meilleure dans les départements les plus importants en terme de population, atteignant 65 % [56-77] dans le Rhône. Pour les cas bactériologiquement confirmés, l'estimation du nombre de cas selon le meilleur modèle prenant en compte une interaction entre les sources DO et CAM était de 1122 [1077 – 1167] cas et l'exhaustivité de la DO atteignait 54 % [52-56]. Cette exhaustivité variait selon la zone géographique dans le même sens que pour l'analyse sur tous les cas.

L'analyse menée sur les cas bactériologiquement confirmés était plus fiable, car la définition de cas était identique pour les 3 sources. Les résultats obtenus montraient que l'exhaustivité de la DO demeurait globalement médiocre, sans évolution sensible depuis la première étude menée en 1992-1993. La dépendance observée entre la zone géographique et la DO suggérait que la déclaration obligatoire variait selon des caractéristiques liées à la zone (urbanisation, offre de soins). Ces éléments pouvaient permettre d'orienter de nouvelles actions de sensibilisation. L'introduction des responsables de laboratoires comme déclarants devra être évaluée dans le futur comme facteur d'amélioration de la déclaration de cette maladie, en ce qui concerne les cas bactériologiquement confirmés.

1 CONTEXTE

1.1 La surveillance de la tuberculose en France

La tuberculose figure parmi les maladies faisant l'objet d'une transmission obligatoire de données individuelles à l'autorité sanitaire (article L. 3113-1 du code de la santé publique). Cette déclaration obligatoire (DO) est actuellement une des principales sources d'information.

La surveillance au niveau national permet de suivre les tendances de la maladie et l'évolution des groupes à risque. Au niveau départemental, la DO fait le lien entre les services déconcentrés du ministère de la santé et les services des conseils généraux, et cette collaboration permet d'adapter les actions de lutte antituberculeuse.

En 1999 et 2000, seuls les médecins étaient concernés par la déclaration obligatoire des cas de tuberculose.

Le système de déclaration obligatoire est basé sur une surveillance passive. Comme pour les autres maladies de ce système, la déclaration obligatoire n'est pas exhaustive.

1.2 Justification de l'étude

L'exhaustivité de la DO de la tuberculose a été étudiée en 1992-1993 par le Réseau national de santé publique dans 15 départements, par la méthode capture-recapture à partir de deux autres sources d'information : elle était estimée à 50 % [48 %-52 %] pour l'ensemble des départements, variant de 27 à 68 % suivant les départements. Dans le Rhône, l'exhaustivité était estimée à 43 %, mais les autres départements de Rhône-Alpes n'avaient pas participé à cette étude.

En 2000, 6539 cas de tuberculose avaient été déclarés en France, dont 396 en Rhône-Alpes. Cependant la perception d'une sous-déclaration de la tuberculose persistait. La mise en œuvre du décret du 6 mai 1999, élargissant l'obligation de notification aux responsables de laboratoires, et non plus aux seuls médecins, devrait entraîner une amélioration de la déclaration.

Dix ans après l'enquête nationale, le besoin apparaissait en Rhône-Alpes d'évaluer à nouveau l'exhaustivité et de faire le point de la situation : estimation de l'incidence vraie de la maladie, évolution de l'exhaustivité depuis 1993 dans le Rhône, comparaison de la performance de la DO entre les différents départements, point de départ avant l'entrée en jeu des laboratoires dans le processus de notification.

1.3 Diagnostic biologique

Le diagnostic biologique de la tuberculose repose soit sur la mise en évidence de bacilles acido-alcoolo-résistants (BAAR) dans un prélèvement, soit sur l'identification du bacille de Koch après mise en culture d'un prélèvement. La présence de germes dans un prélèvement d'expectoration bronchique témoigne du risque de diffusion et de contamination de personnes dans l'entourage.

2 OBJECTIFS DE L'ETUDE

- Objectif général : évaluation quantitative du système de surveillance de la tuberculose par la déclaration obligatoire en région Rhône-Alpes.
- Objectifs spécifiques :
 - Estimer le nombre de cas de personnes mises sous traitement anti-tuberculeux dans la région Rhône-Alpes ;
 - Evaluer l'exhaustivité de la déclaration obligatoire en région Rhône-Alpes et dans chacun des départements de la région ;
 - Décrire la situation épidémiologique de la tuberculose dans chaque département et pour l'ensemble de la région et les caractéristiques des cas qui échappent à la DO de la tuberculose.

3 MATERIEL ET METHODES

3.1 Population d'étude et recueil des données

L'étude rétrospective a porté sur les années 1999 et 2000. Elle a été réalisée sur l'ensemble des cas domiciliés dans les 8 départements de la région Rhône-Alpes.

Deux sources complémentaires à la déclaration obligatoire ont été utilisées : les laboratoires d'analyses de biologie médicale (LABM) et les caisses d'assurance maladie (CAM).

3.2 Matériel

3.2.1 La déclaration obligatoire

La tuberculose fait partie de la liste des maladies à déclaration obligatoire (DO) depuis 1964. Tout médecin ayant diagnostiqué une tuberculose est tenu de remplir un formulaire de

notification officiel (annexe1). Cette fiche de notification est envoyée au médecin inspecteur de santé publique de la Direction départementale des affaires sanitaires et sociales (Ddass). Celui-ci valide les fiches puis une saisie est réalisée à la Ddass sur un logiciel spécifiquement conçu pour cet usage (logiciel BK version 3). Une analyse des données départementales peut être alors réalisée et utilisée par les acteurs locaux de la lutte contre la tuberculose. Le fichier est adressé annuellement à l'Institut de veille sanitaire (InVS) pour le suivi des tendances nationales.

Les critères de déclaration sont les cas de tuberculose maladie ayant donné lieu à la mise en route d'un traitement anti-tuberculeux avec au moins trois anti-tuberculeux. La primo-infection est exclue de cette définition.

Le formulaire de DO utilisé en 1999 et 2000 était indirectement nominatif. Il recueillait des informations sur les caractéristiques du malade, identifié par l'initiale de son nom et son prénom, sa date de naissance, son sexe, le code postal de son domicile. Le formulaire contenait également des informations sur la date de mise sous traitement, les dates et résultats des examens biologiques et la localisation de la tuberculose.

Les Ddass de la région Rhône-Alpes ont fourni les fichiers informatisés des années 1999 et 2000.

3.2.2 Enquête auprès des laboratoires

Tous les laboratoires d'analyses de biologie médicale publics et privés de la région ont été contactés par courrier pour participer au recueil des données.

Une fiche de recueil standardisée (annexe 2) devait être renseignée pour tout diagnostic biologique positif de tuberculose chez un patient domicilié dans un des départements de la région.

Deux relances téléphoniques ont été effectuées auprès des laboratoires par un enquêteur.

Les fiches ont été validées avec recherche des informations manquantes auprès des services cliniques demandeurs si besoin, en excluant les cas ne répondant pas aux critères de définition, puis saisies à l'aide du logiciel Epi-info.

Les informations recueillies étaient les suivantes : initiale du nom, prénom, date de naissance, code postal du domicile, dates et résultats des analyses bactériologiques et localisation du prélèvement.

3.2.3 Enquête auprès des caisses d'assurance maladie

La tuberculose « active » est l'une des trente maladies donnant lieu à l'exonération du ticket modérateur. Ces trente maladies comportant un traitement prolongé et une thérapeutique coûteuse ont été fixées par décret. L'exonération du ticket modérateur est accordé par la caisse de l'assuré sur avis du médecin conseil. Seul le motif de la première demande d'exonération est renseigné par les services médicaux. Selon le code de la Sécurité sociale, les « tuberculoses actives » sont définies comme « celles qui sont confirmées soit par la présence du bacille de Koch, dont la recherche doit être faite systématiquement, soit par l'existence de lésions histologiques spécifiques ». Le médecin conseil peut par ailleurs demander une exonération du ticket modérateur sur la constatation de la mise en œuvre d'un traitement spécifique.

Les services médicaux des principaux régimes régionaux (régime général, Mutualité sociale agricole, Caisse médicale régionale et Service médical régional des mines) ont été sollicités pour participer à l'étude.

Il était demandé aux médecins conseils de notifier les cas de tuberculose domiciliés dans un des départements de la région ayant fait l'objet d'une demande d'exonération du ticket modérateur en 1999 et 2000, en excluant les primo-infections.

Une fiche de recueil standardisée (annexe 3) était renseignée par les médecins conseils pour chaque cas. Cette fiche comportait les informations suivantes : initiale du nom, prénom, date de naissance, code postal, date de la demande, date de mise en route du traitement, localisation clinique de la maladie, résultats des examens bactériologiques.

Les fiches ont été validées en excluant celles ne répondant pas aux critères de définition de cas puis saisies à l'aide du logiciel Epi-info.

3.3 Méthode capture-recapture

La méthode capture-recapture initialement utilisée en zoologie est actuellement utilisée en épidémiologie pour estimer le nombre total de cas d'une maladie dans une population donnée, en particulier pour estimer l'exhaustivité d'un système de surveillance.

3.3.1 Principes

Cette méthode permet, en croisant les cas d'une maladie recensée par plusieurs sources de données dans une population et pour une période définie, et après avoir identifié les cas communs entre les différentes sources et sous certaines hypothèses, d'estimer le nombre de

cas identifiés par aucune des sources. L'estimation ainsi obtenue permet d'estimer le nombre total de cas de maladie et l'exhaustivité de chaque système.

La distribution des cas communs est représentée par un diagramme de Venn (Figure 1) ou par une table de contingence (tableaux 1 et 2). Les indices 1 ou 2 correspondent à la présence (1) ou absence (2) des cas dans les sources A, B et C dans cet ordre.

Figure 1 : répartition des cas selon les trois sources présentée sur un diagramme de Venn

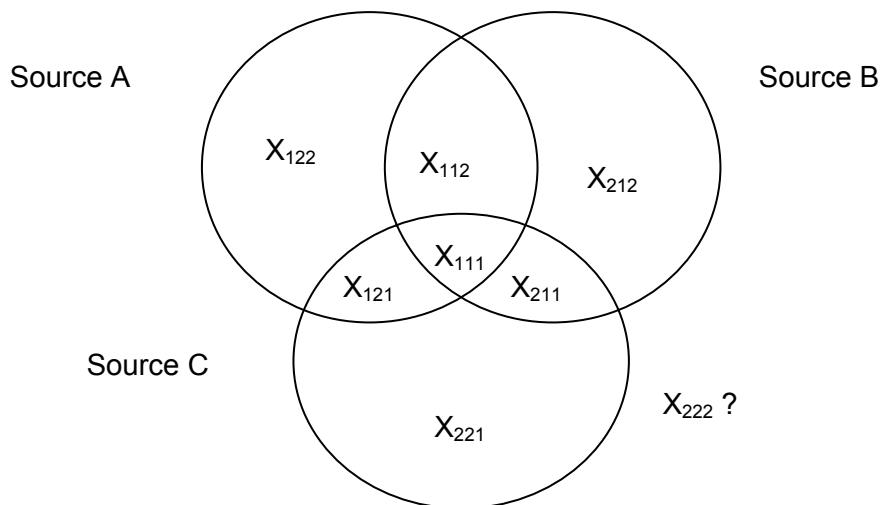


Tableau 1 : répartition des cas selon les trois sources présentée dans un tableau de contingence

		Source C			
		Oui		Non	
		Source B		Source B	
		Oui	Non	Oui	Non
Source A	Oui	X ₁₁₁	X ₁₂₁	X ₁₁₂	X ₁₂₂
	Non	X ₂₁₁	X ₂₂₁	X ₂₁₂	X ₂₂₂ ?

3.3.2 Estimation du nombre total de cas avec deux sources

Avec deux sources, le tableau de contingence répartit les cas en fonction de leur présence ou absence dans l'une ou l'autre source.

Tableau 2 : répartition des cas selon deux sources présentée dans un tableau de contingence

		Source B		
Source A		Oui	Non	
	Oui	X_{11}	X_{12}	N_A
	Non	X_{21}	$X_{22} ?$	
		N_B		

Sous l'hypothèse d'indépendance des sources, c'est à dire que la probabilité d'être (A+,B+) est la même que d'être (A+, B-), les estimateurs de Sekar et Demming permettent d'estimer le nombre de cas recensés par aucune des sources (X_{22}), le nombre total de cas (N), sa variance (VarN) et son intervalle de confiance à 95 % (IC) selon les formules suivantes :

$$X_{22} = \frac{X_{12} \cdot X_{21}}{X_{11}} \quad (\text{formule 1})$$

$$N = \frac{N_A \cdot N_B}{X_{11}} \quad \text{où } N = N_{\text{obs}} + X_{22} ? \quad (\text{formule 2})$$

$$\text{VarN} = \frac{N_A \cdot N_B \cdot X_{12} \cdot X_{21}}{(N_{11})^3} \quad (\text{formule 3})$$

$$\text{IC 95 \%} = N \pm 1,96 \text{VarN} \quad (\text{formule 4})$$

Chapman a proposé d'autres estimateurs non biaisés dans le cas d'effectifs faibles selon les formules suivantes :

$$N = \frac{(N_A + 1)(N_B + 1)}{(X_{11} + 1)} \quad \text{VarN} = \frac{(N_A + 1)(N_B + 1)X_{12} \cdot X_{21}}{(X_{11})^2 (X_{11} + 1)}$$

Le taux d'exhaustivité d'une source est : $E = \frac{N_A}{N}$ où N_A est le nombre de cas notifiés par la source A et N le nombre total de cas estimés.

3.3.3 Conditions d'application de la méthode

3.3.3.1 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 de la zone A dans la source B sera nulle. Si les cas sont issus de populations 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 identifiés 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 communs repose sur une combinaison de critères. Plusieurs définitions de cas communs peuvent ainsi être proposées selon les sources considérées. Une analyse de leur sensibilité permet d'orienter le choix sur une définition des cas communs.

3.3.3.2 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 identifiés 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 surestimation de N.

Avec plus de deux sources, la dépendance entre les sources peut être évaluée et prise en compte dans l'estimation de N grâce 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'homogénéité, dans les sources. Par ailleurs, la présence de variables d'homogénéité de capture peut induire une dépendance positive ou négative entre les sources.
- 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. 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 : 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 donc une surestimation de N. Cette condition est très liée à celle de l'homogénéité de capture.

3.3.4 Evaluation de la dépendance en utilisant la méthode de Wittes

Wittes a proposé trois méthodes pour identifier et évaluer la présence de dépendance entre différentes sources, s'il y a au moins trois sources.

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

Pour trois sources ou plus, la dépendance peut être évaluée en comparant les estimations pour chaque couple de sources. N est estimé pour chaque combinaison des sources couplées deux à deux. Si l'estimation pour un couple de sources est significativement différente de celle des autres couples, on peut suspecter une dépendance entre ces deux sources.

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

Pour trois sources, on peut évaluer le sens et la force de la dépendance entre deux sources. A partir d'un diagramme de Venn, on utilise les chiffres de deux sources qui apparaissent dans la troisième source, on calcule le rapport des cotes (RC) et les intervalles de confiance à 95 %. Un RC à 1 indique que les sources sont indépendantes. Un RC supérieur à 1 implique une dépendance positive à l'origine d'une sous-estimation du nombre de cas. Un RC inférieur à 1 indique une dépendance négative à l'origine d'une surestimation du nombre de cas.

3.3.4.3 Estimation du nombre total de cas après fusion des deux sources dépendantes

Le nombre total de cas peut être estimé en regroupant les sources dépendantes en une source unique. L'analyse est alors réalisée en considérant l'ensemble comme deux sources indépendantes, les deux sources fusionnées et la troisième.

3.3.5 Estimation du nombre total de cas en utilisant les modèles log-linéaires

L'utilisation de modèles log-linéaires permet d'estimer le nombre total de cas qui est calculé en prenant en compte la dépendance entre les sources ainsi que les variables d'hétérogénéité de capture. 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.

Le tableau de contingence a une cellule structurellement vide correspondant à l'absence de notification de cas dans l'ensemble des trois 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. Cependant, 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.

La sélection du meilleur modèle est choisie en accord avec deux principes : la parcimonie et une bonne adéquation avec les données observées.

L'application du principe de parcimonie retient le modèle qui a le moins de termes d'interactions. En partant du modèle saturé, par une procédure pas à pas descendante, on choisit le meilleur modèle avec le moins de termes d'interaction et qui a une bonne adéquation avec les données observées.

Les critères de choix du meilleur modèle retenu étaient basés sur la statistique du rapport de vraisemblance G^2 . Un rapport de vraisemblance non significatif indique une bonne adéquation du modèle. L'Akaike information criterion (AIC) et le Bayesian information criterion (BIC) ont également été utilisés comme critères d'adéquation du modèle, en retenant le modèle pour lequel les valeurs de l'AIC et du BIC sont les plus basses.

3.3.6 Identification des cas communs intra-sources

Les critères d'identification des cas communs doivent être suffisamment sensibles et spécifiques pour satisfaire à la condition d'application relative à l'identification des seuls vrais cas communs. Le choix des critères est réalisé avec les variables communes aux différentes sources, en tenant compte de la proportion renseignée de ces variables et de la validité de l'information.

La première étape consiste à éliminer les cas communs à l'intérieur même de chaque source car il peut arriver qu'un même cas soit à l'origine de plusieurs déclarations. Dans chaque source, les cas avec les mêmes initiales, date de naissance, sexe ont été identifiés. Avant d'être déclarés cas communs, les dates de mise sous traitement (DO et CAM) ou les dates de diagnostic bactériologique (LABM) ont été vérifiées et devaient être cohérentes (même mois et même année).

3.3.7 Identification des cas communs entre les sources

Les cas recensés par la DO, les LABM et les CAM ont été comparés à l'intérieur de chaque département.

Pour cette comparaison, cinq variables ont été utilisées : l'initiale du nom, le prénom, la date de naissance, le sexe et le code postal. Pour les deux premiers tris, un code a été établi en utilisant la première lettre du nom, la première lettre du prénom, la date de naissance, le sexe et le code postal.

Plusieurs tris successifs ont été réalisés :

- un premier tri pour identifier les cas strictement identiques sur l'ensemble des variables c'est à dire possédant le même code ;
- un second tri pour identifier les cas différent sur l'une des cinq variables composant le code ;
- un troisième tri pour identifier les cas différent sur la date de naissance ou le sexe ou le code postal mais avec initiales du nom et prénom identiques ;

- un quatrième tri pour identifier des cas communs entre caisses et DO ou LABM, différent sur l'initiale du nom et le prénom mais identiques sur les trois autres variables (essentiellement les cas pour lesquels l'initiale du nom et le prénom de l'ayant droit avaient été renseignés dans la source CAM).

3.3.8 Analyse statistique

La saisie, la validation et l'identification des cas communs, l'estimation du nombre total de cas avec les sources croisées deux à deux et les tests d'indépendance entre les sources ont été réalisés à l'aide du logiciel Epi-info, version 6.4c (CDC Atlanta, Georgie). La modélisation a été réalisée sur le logiciel BMDP, module 4F (BMDP Statistical Software Inc, Los Angeles, Californie).

3.3.9 Aspects éthiques

La Commission nationale informatique et libertés avait émis un avis favorable à la réalisation de l'étude selon le protocole qui lui avait été adressé.

4 RESULTATS

4.1 Etude descriptive

4.1.1 Description des sources

4.1.1.1 La déclaration obligatoire

En 1999 et 2000, le nombre de cas déclarés validés, pour la région Rhône-Alpes, était de 786. Le nombre de cas déclarés était stable pour les deux années. Il était de 390 en 1999 et de 396 en 2000. Le tableau 3 indique le nombre de cas déclarés pour les deux années et par département et l' incidence moyenne calculée à partir de ces données.

Tableau 3 : répartition par département des cas déclarés et taux d'incidence moyen par département de la tuberculose, 1999-2000, région Rhône-Alpes

Département	Population (1999)	Cas déclarés 1999-2000	Incidence moyenne / an (/100000 hab)
Ain	515 270	66	6,4
Ardèche	286 023	30	5,2
Drôme	437 778	34	3,9
Isère	1 094 006	150	6,9
Loire	728 524	68	4,7
Rhône	1 578 869	317	10,0
Savoie	373 258	50	6,7
Haute-Savoie	631 679	71	5,6
Région	5 645 047	786	7,0
France	58 520 688	13 047	11,1

Les résultats des examens bactériologiques directs étaient renseignés pour 90,3 % des cas. Le résultat était positif pour 46,8 % des déclarations (332). Pour 81,8 % des personnes, la déclaration indiquait si une mise en culture des prélèvements bactériologiques avait été réalisée. Elle avait été effectivement réalisée pour 631 personnes (80,3 %). Les résultats étaient renseignés pour 55,8 % des personnes (352) pour lesquelles cette culture avait été réalisée.

Les résultats étaient positifs pour 203 personnes soit 57,6 % des cas renseignés. Pour 57,7 % des résultats positifs, le résultat de l'examen direct était négatif ou inconnu.

Les délais observés entre la date de mise sous traitement et la date de déclaration pour la région et pour les deux années variaient de - 4 semaines à 59 semaines. La médiane était à 1 semaine, le mode à 0 et la moyenne à 3,4 semaines.

4.1.1.2 Les diagnostics des laboratoires d'analyses de biologie médicale

Sur l'ensemble de la région, 388 laboratoires publics et privés ont été contactés et 311 ont participé à l'étude. Le taux de participation variait selon les départements de 66 % à 91 %.

Au total, 434 cas de tuberculose diagnostiqués par les laboratoires ont été déclarés. Le tableau 4 indique la répartition des cas par département de domicile. Le nombre de cas déclarés par les LABM était proportionnellement faible dans le Rhône.

Tableau 4 : nombre de cas de tuberculose domiciliés déclarés par les laboratoires d'analyses de biologie médicale, 1999-2000

Département de domicile	Nombre de cas 1999	Nombre de cas 2000	Nombre de cas 1999-2000
Ain	23	17	40
Ardèche	15	12	27
Drôme	16	13	29
Isère	73	55	128
Loire	40	59	99
Rhône	57	16	73
Savoie	2	12	14
Haute-Savoie	14	10	24
Région	240	194	434

Parmi les 434 cas signalés par les laboratoires, 80,4 % présentaient un diagnostic biologique positif en examen direct ou en culture d'un prélèvement bronchique. Les autres cas étaient positifs soit sur un autre type de prélèvement soit à partir d'une biopsie tissulaire.

4.1.1.3 Les caisses d'assurance maladie

Les services médicaux des caisses d'assurance maladie (CAM) des trois principaux régimes (Caisse primaires d'assurance maladie (CPAM), Mutualité sociale agricole (MSA) et caisses médicales régionales des assurés non salariés et non agricoles (CMR) et du régime spécifique des travailleurs des mines (SSM)) ont été sollicités. Vingt trois caisses étaient concernées, toutes ont participé à l'étude. Il s'agissait de 12 CPAM, 8 MSA et 2 CMR et 1 SSM.

Les CAM ont déclaré 901 cas de tuberculose maladie ayant donné lieu à une demande d'exonération du ticket modérateur. Le tableau 5 indique la répartition des cas déclarés par département de domicile et par année.

Tableau 5 : nombre de cas de tuberculose maladie par département de domicile et par année déclarés par les caisses d'assurance maladie

département	Nombre de cas 1999	Nombre de cas 2000	Nombre de cas 1999-2000
Ain	51	42	93
Ardèche	26	21	47
Drôme	23	24	47
Isère	88	69	157
Loire	43	57	100
Rhône	173	129	302
Savoie	26	20	46
Haute-Savoie	57	52	109
région	487	414	901

Parmi ces 901 cas, 68,5 % avaient un résultat d'examen direct renseigné et parmi ceux-ci, 199 (32,3 %) avaient un résultat positif. Concernant les résultats des mises en culture des prélèvements, 662 (73,4 %) avaient un résultat renseigné et, parmi ceux-ci, 312 (47,1 %) avaient un résultat positif.

Au total, 49,3 % des cas déclarés par les CAM avaient un diagnostic bactériologique positif connu. La tuberculose avait une localisation pulmonaire pour 71,8 % des cas.

4.1.2 Caractéristiques des cas notifiés par les trois sources

Avant élimination des doublons, un total de 2121 cas avec un diagnostic de tuberculose avait été identifié par les trois sources.

Les comparaisons sur les cas validés selon la source pouvaient être effectuées selon l'âge et le sexe (tableau 6). La répartition par groupe d'âge des cas différait significativement selon la source ($\chi^2 = 45,2$, ddl =6, $p<0,001$). La répartition selon le sexe ne différait pas en fonction de la source.

La proportion de cas de localisation pulmonaire était similaire quelle que soit la source. La proportion de cas bactériologiquement confirmés était significativement plus importante pour la source DO que par la source CAM ($\chi^2 = 11,29$, $p<0,001$). Pour la source LABM, selon la définition, les cas étaient tous bactériologiquement confirmés (tableau 6).

Tableau 6 : caractéristiques des cas selon les trois sources

		D.O.		LABM		CPAM	
		N	%	N	%	N	%
sexe	Masculin	445	56,6	243	56,0	503	55,8
	Féminin	341	43,4	188	43,3	397	44,1
	Inconnu	0	0	3	0,7	1	0,1
Age	< 25 ans	65	8,3	8	1,8	27	3,0
	25-44 ans	182	23,1	112	25,8	266	29,5
	45-64 ans	206	26,2	116	26,8	258	28,6
	≥ 65 ans	333	42,4	198	45,6	350	38,9
Localisation	pulmonaire	572	72,8	337	77,6	646	71,7
	extra-pulmonaire	214	27,2	97	22,4	255	28,3
Confirmation bactériologique	oui	605	77,0	434	100,0	628	69,7
	non	181	23,0	0	0,0	273	30,3
Total		786	100,0	434	100,0	901	100,0

4.2 Distribution des cas déclarés dans les trois sources et identification des cas communs

L'élimination des doublons et des triplets a permis d'identifier 1306 cas de tuberculose. L'analyse globale a inclus tous les cas validés identifiés dans les trois sources.

4.2.1 Ensemble des cas en Rhône-Alpes, période 1999-2000

Parmi les 1306 cas, 159 étaient communs aux trois sources et 494 étaient communs à deux sources (figure 2, tableau 7).

Figure 2 : répartition du nombre de cas de tuberculose selon les sources (N=1306)

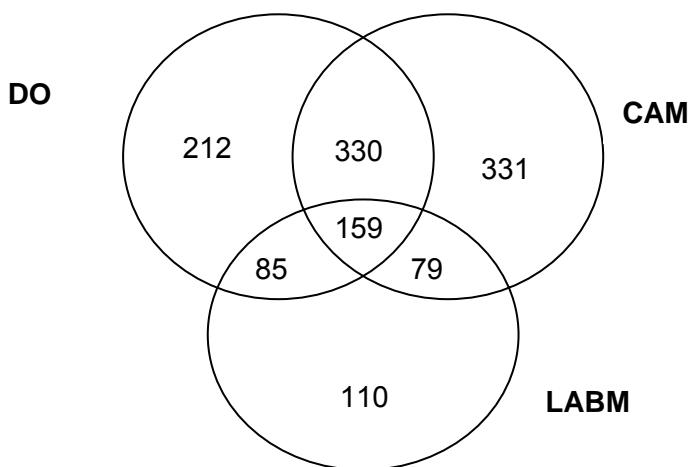


Tableau 7 : répartition du nombre de cas selon les sources (N =1306)

		CAM			
		Oui		Non	
		LABM		LABM	
		Oui	Non	Oui	Non
DO	Oui	159	330	85	212
	Non	79	331	110	X ₂₂₂ ?

4.2.2 Cas bactériologiquement confirmés en Rhône-Alpes, années 1999-2000

L'élimination des doublons et triplets a permis d'identifier 947 cas bactériologiquement confirmés de tuberculose. Parmi ces cas, 159 étaient communs aux trois sources, 401 étaient communs à deux sources et 387 étaient uniques (figure 3, tableau 8).

Figure 3 : répartition du nombre de cas bactériologiquement confirmés selon les sources (N=947)

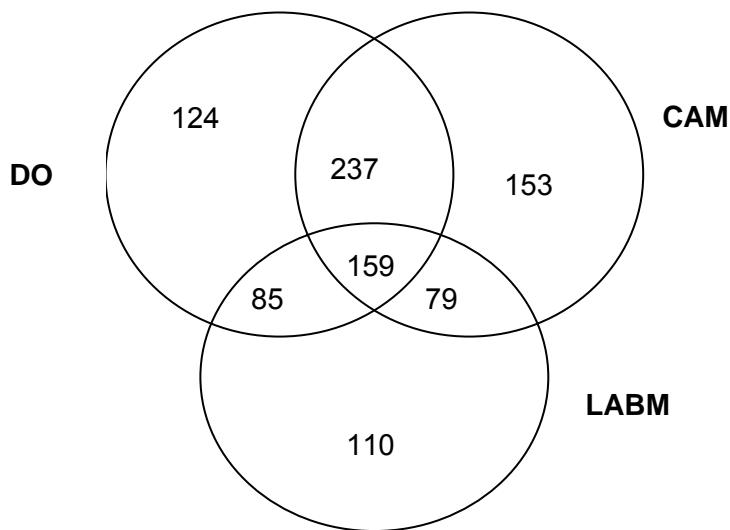


Tableau 8 : répartition du nombre de cas bactériologiquement confirmés selon les sources (N = 947)

CAM					
		Oui		Non	
		LABM		LABM	
		Oui	Non	Oui	Non
DO	Oui	159	237	85	124
	Non	79	153	110	$X_{222} ?$

4.3 Estimation du nombre total de cas et évaluation de la dépendance entre les sources en utilisant les méthodes de Wittes

4.3.1 Estimation du nombre total de cas et de l'exhaustivité en croisant les sources deux à deux

Les croisements des fichiers deux à deux permettent d'estimer le nombre total de cas, le nombre de cas non recensés par aucune des sources selon les estimateurs de Sekarg et Demming et de calculer à partir de ces estimations un taux d'exhaustivité pour chacune des sources.

Selon Wittes, la dépendance entre trois sources peut être identifiée par l'estimation du nombre total de cas en utilisant deux sources.

4.3.1.1 Ensemble des cas pour l'ensemble de la région pour la période 1999-2000 et par année

L'estimation du nombre total de cas était de 1395 en croisant les sources DO et LABM, de 1445 en croisant les sources DO et CAM et de 1636 en croisant les sources LABM et CAM (tableau 9).

L'intervalle de confiance à 95 % de l'estimation obtenue en croisant les sources CAM et LABM ne se superposait pas aux intervalles de confiance des deux autres estimations. Ce résultat, ajouté au nombre inférieur de cas estimé en croisant les sources DO et CAM et DO et LABM par rapport au nombre estimé en croisant les sources CAM et LABM indiquait soit une dépendance positive entre les sources DO et LABM et entre les sources DO et CAM, soit une dépendance négative entre les sources CAM et LABM.

L'exhaustivité du système de surveillance DO était estimée à 56,4 % [52,6-60,6] en croisant les sources DO et LABM et à 54,4 % [52,5-56,5] en croisant les sources DO et CAM.

Tableau 9 : estimation du nombre total de cas en croisant les sources deux à deux, exhaustivité des sources et intervalles de confiance, région Rhône-Alpes, 1999-2000

S1, S2	X ₁₁	X ₁₂	X ₂₁	N	IC (95 %)	Exh1	IC (95 %)	Exh2	IC (95 %)
DO, LABM	244	542	189	1395	1299-1491	56,4 %	52,6-60,6	31,1 %	29,1-33,4
DO, CAM	489	297	410	1445	1392-1498	54,4 %	52,5-56,5	62,2 %	60,0-64,6
CAM, LABM	238	661	195	1636	1516-1755	55,1 %	51,3-59,4	26,5 %	24,7-28,6

L'estimation du nombre total de cas a également été réalisée par année pour l'ensemble de la région.

Pour l'année 1999, le nombre de cas estimé était de 701 en croisant les sources DO et LABM, de 741 en croisant les sources DO et CAM et de 804 en croisant les sources CAM et LABM.

L'exhaustivité du système de surveillance DO était estimée à 55,7 % [51,1-61,3] en croisant les sources DO et LABM et à 52,6 % [50,2-55,3] en croisant les sources DO et CAM.

Pour l'année 2000, le nombre de cas estimé s'élevait à 691 en croisant DO et LABM, à 700 en croisant DO et CAM et à 864 en croisant CAM et LABM.

L'exhaustivité du système de surveillance DO était estimée à 57,5 % [52,1-64,0] en croisant les sources DO et LABM et à 56,6 % [53,6-59,9] en croisant les sources DO et CAM.

La sous-estimation du nombre de cas pour les sources DO et CAM et pour les sources DO et LABM évoquait également soit une dépendance positive entre DO et LABM et entre DO et CAM, soit une dépendance négative entre les sources CAM et LABM.

4.3.1.2 Cas bactériologiquement confirmés

L'estimation du nombre total de cas et de l'exhaustivité en croisant les sources deux à deux a également été réalisée pour le sous-ensemble des cas bactériologiquement confirmés (tableau 10).

Le nombre de cas estimé en croisant les sources deux à deux était de 1074 pour les sources DO-LABM, de 959 pour les sources DO-CAM et de 1143 pour les sources LABM-CAM. Les résultats observés laissaient supposer soit une dépendance positive entre DO et CAM, soit une dépendance négative entre DO et LABM et entre LABM et CAM.

Tableau 10 : estimation du nombre total de cas en croisant les sources deux à deux, exhaustivité des sources et intervalles de confiance, région Rhône-Alpes, cas bactériologiquement confirmés, 1999-2000

S1, S2	X ₁₁	X ₁₂	X ₂₁	N	IC (95 %)	Exh1	IC (95 %)	Exh2	IC (95 %)
DO, LABM	244	361	189	1074	1005-1142	56,4	53,1-60,3	40,4	38,0-43,1
DO, CAM	396	209	232	959	926-993	63,1	60,9-65,4	65,5	63,3-67,9
CAM, LABM	238	390	195	1143	1066-1219	55,0	51,6-59,0	37,9	35,6-40,7

4.3.1.3 Ensemble des cas par département pour la période 1999-2000

Compte tenu des effectifs faibles dans certains départements, les estimateurs de Chapman ont été utilisés. Par ailleurs, les données des deux départements du sud de la région, Drôme et Ardèche ont été fusionnées.

Les deux tableaux suivants présentent les estimations du nombre total de cas en croisant les sources DO et LABM (tableau 11) et en croisant les sources DO et CPAM (tableau 12).

Tableau 11 : estimation du nombre total de cas par département en croisant les sources DO et LABM, exhaustivité des sources et intervalles de confiance, 1999-2000

Département	X ₁₁	X ₁₂	X ₂₁	N	IC (95 %)	Exh1	IC (95 %)	Exh2	IC (95 %)
Ain	20	46	20	130	97-163	50,8 %	40,8-67,1	30,8 %	24,7-40,7
Drôme-Ardèche	24	40	32	147	112-182	43,5 %	35,5-56,4	38,1 %	31,0-49,3
Isère	89	61	38	214	198-229	70,1 %	65,4-75,5	59,3 %	55,4-63,9
Loire	36	32	63	185	153-218	36,8 %	31,3-44,5	53,5 %	45,6-64,7
Rhône	54	263	19	427	374-479	74,2 %	66,2-84,5	17,1 %	15,3-19,4
Savoie	9	41	5	76	49-102	65,8 %	50,3-95,1	18,4 %	14,1-26,6
Haute-Savoie	12	59	12	142	88-187	51,8 %	38,9-77,6	17,5 %	13,2-26,2

L'exhaustivité des cas déclarés par la source DO est meilleure dans les deux départements les plus importants de la région (Rhône et Isère).

Tableau 12 : estimation du nombre total de cas par département en croisant les sources DO et CAM, exhaustivité des sources et intervalles de confiance, 1999-2000

Département	X ₁₁	X ₁₂	X ₂₁	N	IC (95 %)	Exh1	IC (95 %)	Exh2	IC (95 %)
Ain	44	22	48	137	121-154	48,2 %	43,0-54,8	67,2 %	59,9-76,3
Drôme-Ardèche	34	30	60	175	143-207	36,6 %	31,1-44,5	53,7 %	45,6-65,3
Isère	89	61	68	264	241-287	56,8 %	52,3-62,2	59,7 %	54,8-65,1
Loire	30	38	70	224	174-274	30,4 %	25,0-38,7	44,6 %	36,7-56,9
Rhône	220	97	81	434	417-450	73,0 %	70,4-75,9	69,4 %	66,8-72,1
Savoie	25	25	21	91	74-108	54,9 %	46,6-66,9	50,5 %	42,9-61,6
Haute-Savoie	47	24	62	165	144-184	43,3 %	38,6-49,3	66,5 %	59,2-75,7

4.3.2 Evaluation de la dépendance entre les sources par estimation des rapports des cotes.

La deuxième méthode proposée par Wittes pour évaluer la présence et la force de la dépendance entre deux sources est d'estimer le rapport des cotes en utilisant les données de deux sources incluses dans la troisième source. Cette méthode a été utilisée pour tester la dépendance des sources deux à deux pour l'ensemble des cas (région et département) et pour les cas de tuberculose pulmonaire.

4.3.2.1 Ensemble des cas de tuberculose

Pour l'ensemble de la région et de la période d'étude, on observait (tableau 13) une forte dépendance positive entre les sources DO et CAM ($RC = 2,6, p<0,001$) et une dépendance positive un peu moins importante entre DO et LABM ($RC = 2,02, p< 0,001$). Le nombre total de cas était sous-estimé. Les sources LABM et CAM étaient indépendantes ($RC=1,20, p=0,25$).

Tableau 13 : test de dépendance par estimation des rapports de cotes, méthode de Wittes, ensemble des cas de tuberculose, Rhône-Alpes, 1999-2000

en LABM					en DO					en CAM				
DO					CAM					LABM				
CAM	+	159	79	238	LABM	+	159	85	244	DO	+	159	330	489
	-	85	110	195		-	330	212	542		-	79	331	410
		244	189	433			489	297	786			238	661	899

RC=2,6 [1,76-3,85]

RC =1,2 [0,88-1,65]

RC=2,02 [1,48-2,75]

Les mêmes estimations ont été appliquées par année et par département. Les tableaux 14 et 15 synthétisent les rapports de cotes estimés. Les mêmes dépendances positives étaient retrouvées pour chacune des deux années entre les sources DO et CAM et entre les sources DO et LABM, la dépendance étant plus forte entre les sources DO et CAM.

Tableau 14 : test d'indépendance par estimation des rapports de cotes, méthode de Wittes, ensemble des cas de tuberculose, région Rhône-Alpes, par année

année	DO-CAM / en LABM	CAM-LABM / en DO	DO-LABM / en CAM
	RC [IC, 95 %]	RC [IC, 95 %]	RC [IC, 95 %]
1999	2,88 [1,68-4,92]	1,50 [0,95-2,36]	2,14 [1,43-3,18]
2000	2,49 [1,38-4,50]	0,92 [0,59-1,44]	1,96 [1,19-3,23]

Les effectifs plus faibles par département rendaient plus difficile l'analyse des dépendances entre les sources. On retrouvait une dépendance positive entre les sources DO et CAM dans l'Ain, la Savoie (rapport de cotes non calculable) et dans le Rhône. Dans ce dernier département, la dépendance était très forte. Une dépendance positive était également retrouvée entre les sources DO et LABM dans la Loire, le Rhône, l'Isère et la Savoie (rapport de cotes non calculable). Dans le Rhône, on observait également une dépendance positive entre les sources LABM et CAM.

Tableau 15 : test d'indépendance par estimation des rapports de cotes, méthode de Wittes, ensemble des cas de tuberculose par département, période 1999-2000

Département	DO-CAM / en LABM	CAM-LABM / en DO	DO-LABM / en CAM
	RC [IC, 95 %]	RC [IC, 95 %]	RC [IC, 95 %]
Ain	4,33 [1,15-16,32]	1,24 [0,40-3,86]	2,73 [0,98-7,60]
Ardèche	1,20 [0,26-5,59]	0,87 [0,20-3,90]	1,50 [0,42-5,41]
Drôme	1,94 [0,44-8,61]	1,28 [0,32-5,13]	2,80 [0,75-10,45]
Isère	1,56 [0,73-3,35]	0,91 [0,47-1,77]	3,90 [1,97-7,74]
Loire	1,42 [0,62-3,24]	1,67 [0,63-4,39]	2,54 [1,06-6,10]
Rhône	14,55 [4,16-50,85]	3,49 [1,52-8,04]	3,40 [1,39-8,29]
Savoie	NC (p=0,04)	1,31 [0,31-5,60]	NC (p=0,03)
Haute-Savoie	2,50 [0,36-17,32]	2,97 [0,60-14,83]	1,82 [0,66-5,06]

4.3.2.2 Cas bactériologiquement confirmés

Les tests d'indépendance par calcul des rapports de cotes ont également été réalisés pour les sous-ensemble des cas bactériologiquement confirmés (tableau 16). On observait une dépendance positive forte entre les sources DO et CAM ($RC=2,60$, $p<0,001$) induisant une sous-estimation du nombre total de cas.

Tableau 16 : test d'indépendance par estimation des rapports de cotes, méthode de Witten, cas bactériologiquement confirmés, période 1999-2000

en LABM					en DO					en CAM				
DO					CAM					LABM				
+		-				+		-				+		
CAM	+	159	79	238	LABM	+	159	85	244	DO	+	159	237	396
	-	85	110	195		-	237	124	361		-	79	153	232
		244	189	433			396	209	605			238	390	628

4.3.3 Estimation du nombre total de cas après fusion des deux sources dépendantes

4.3.3.1 Ensemble des cas de tuberculose, période 1999-2000

La fusion de deux sources dépendantes et l'estimation du nombre total de cas en utilisant deux sources permet de prendre en compte la dépendance. Cependant, il existait deux interactions : DO et LABM et DO et CAM. Les estimations calculées dans le tableau 17 allaient de 1161 à 1543 mais étaient donc biaisées car cette méthode ne permet de contrôler qu'une interaction à la fois : en effet, en regroupant DO et LABM, la dépendance entre DO et CAM persistait et en regroupant DO et CAM, la dépendance entre DO et LABM persistait. Le nombre total de cas serait sous-estimé et la méthode n'était pas applicable.

4.3.3.2 Cas de tuberculose bactériologiquement confirmés, période 1999-2000

Pour les cas bactériologiquement confirmés, seule persistait une dépendance positive entre les sources DO et CAM. L'estimation du nombre total de cas en regroupant ces deux sources était donc la plus fiable.

Tableau 17 : estimation du nombre total de cas après fusion des deux sources dépendantes, cas bactériologiquement confirmés, région Rhône-Alpes, 1999-2000

$(S_1 + S_2) / S_3$	X_{11}	N_A	N_B	X_{22}	N_{est}	IC 95 %
(DO+LABM) / CAM	475	794	628	103	1050	1020-1079
(CAM+LABM) / DO	481	823	605	88	1035	1008-1062
(DO+CAM) / LABM	323	837	433	175	1122	1074-1170

4.4 Estimation du nombre total de cas et évaluation de la dépendance entre les sources en utilisant les modèles log-linéaires pour l'ensemble des cas de tuberculose

4.4.1 Sélection du meilleur modèle par stratégie pas à pas descendante pour l'ensemble des cas déclarés par les 3 sources

L'analyse pas à pas descendante à partir du modèle saturé, prenant en compte toutes les interactions d'ordre deux entre les trois sources (DL, DC, CL), montrait qu'un seul modèle présentait une bonne adéquation avec les données. Ce modèle incluait les interactions entre les sources DO et LABM (DL) et entre les sources DO et CAM (DC). Ce modèle avait une bonne adéquation ($G^2 = 1,32$) et les valeurs de l'AIC et du BIC étaient négatives. Ce modèle donnait une estimation du nombre total de cas de 1767, inférieure à la valeur estimée pour le modèle saturé de 1860. Cette valeur était proche de celle de la moyenne pondérée (1772), calculée à partir de toutes les estimations des différents modèles selon la formule :

$$N_{wBIC} = \sum (N_i * e^{-(BIC_i/2)} / \sum e^{-(BIC_i/2)}).$$

Tableau 18 : Analyse log-linéaire et modèles, estimation du nombre de cas de tuberculose selon les modèles utilisés, région Rhône-Alpes, 1999-2000

Modèles	G ²	p	ddl	AIC	BIC	Nobs	X	Nest	IC (95 %)
DC, CL, LD	0	1	0	0,0	0,0	1306	554	1860	1611-2108
DC, CL	20,47	0	1	18,5	12,8	1306	274	1580	1488-1672
LD, DC	1,32	0,25	1	-0,7	-6,3	1306	461	1767	1619-1915
CL, LD	23,63	0	1	21,6	16,0	1306	213	1519	1467-1570
DC,L	21	0	2	17,0	5,7	1306	297	1603	1528-1679
LD,C	27,1	0	2	23,1	11,8	1306	237	1543	1493-1593
CL,D	28	0	2	24,0	12,7	1306	192	1498	1454-1542
D,C,L	33,2	0	3	27,2	10,3	1306	184	1490	1376-1604

G² : test du rapport de vraisemblance, p : adéquation du modèle, ddl : degré de liberté, Nest : nombre de cas, IC : intervalle de confiance du nombre estimé à 95 %.

4.4.2 Introduction des variables d'hétérogénéité pour l'ensemble des cas déclarés par les 3 sources

Différentes variables d'hétérogénéité ont été tour à tour introduites dans le modèle : le département, la localisation de la maladie et l'âge en classes. Le tableau 19 présente les modèles retenus et les statistiques obtenues pour ces modèles, lorsque la probabilité d'être capturé par chacune des sources variait selon chacune des variables d'hétérogénéité.

4.4.2.1 La localisation de la maladie

La localisation était définie comme pulmonaire ou extra pulmonaire. L'introduction de cette variable faisait apparaître des interactions entre la localisation et chacune des trois sources. Le modèle le plus ajusté prenait en compte ces interactions. L'estimation du nombre de cas était alors de 1816.

4.4.2.2 La zone géographique de domicile

Compte tenu du faible nombre de cas par département, hormis pour le Rhône et l'Isère, trois départements (Loire, Drôme, Ardèche) ont été regroupés en "zone sud" de la région et trois autres (Ain, Savoie, Haute-Savoie) en "zone nord". L'introduction de cette variable "zone" faisait également apparaître des interactions entre la zone géographique et chacune des trois sources. Le meilleur modèle prenait en compte ces interactions. Cependant, ce modèle

n'avait pas une bonne adéquation du fait de la présence d'interactions d'ordre 3, ne pouvant être prises en compte. Le nombre total de cas estimé était alors de 1729.

4.4.2.3 L'âge

L'âge a été reclassé en trois groupes : les moins de 25 ans, les 25-64 ans et les plus de 64 ans. L'introduction de cette variable montrait que le meilleur modèle retenu prenait en compte les deux interactions entre la source DO et la source CAM et entre la source DO et la source LABM 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.

4.4.2.4 La localisation de la maladie et la zone géographique

En prenant en compte dans le modèle la localisation de la maladie et la zone géographique, il n'existait pas d'interaction entre la localisation et la zone géographique. Le meilleur modèle était DC, DL, PC, PL, PD, RC, RL, RD avec des valeurs AIC et BIC négatives. Ce modèle, comme celui prenant en compte la zone géographique seule n'avait pas une bonne adéquation du fait de la présence d'interaction d'ordre 3. Le meilleur modèle restait DL,DC pour l'estimation totale.

Tableau 19 : estimation du nombre total de cas de tuberculose et des statistiques obtenues en introduisant des variables d'hétérogénéité (zone géographique et localisation)

MODELE	Nest	IC (95 %)	ddl	G ²	p	AIC	BIC
Modèle initial (DC, DL)	1767	1619-1915	1	1,32	0,25	-0,7	-5,7
Localisation de la maladie							
DC, DL, LC, PC, PL, PD	1874	1693-2055	3	0,9	0,8	-5,1	-20,2
DC,DL,PC,PL,PD	1816	1653-1979	4	1,39	0,8	-6,6	-26,7
Zone géographique							
DC, DL, CL, RC, RL, RD	1872	1691-2053	9	23,3	0,005	5,3	-39,9
DC, DL, RC, RL, RD	1729	1592-1866	10	26,7	0,03	6,7	-43,5
Localisation et zone géographique							
DC, DL, PC, PL, PD, RC, RL, RD	1775	1624-1926	34	63,9	0,001	-4,1	-174,9

4.4.3 Modèles retenus et estimations finales du nombre total de cas et de l'exhaustivité des sources pour l'ensemble des cas déclarés par les 3 sources

Dans le modèle initial, l'exhaustivité différait significativement selon les sources. L'exhaustivité de la DO n'atteignait pas 50 % et l'exhaustivité des LABM était très faible. En stratifiant selon la localisation de la maladie, l'exhaustivité était meilleure pour les cas de localisation pulmonaire, et cela quelle que soit la source.

En stratifiant selon la zone de la région, l'exhaustivité de la DO variait selon les zones. Elle était plus élevée dans les départements de plus grande taille, plus urbanisés comme le Rhône et l'Isère. Dans le Rhône, l'exhaustivité de la DO atteignait 65 %. Elle était particulièrement faible dans le sud de la région (27 %). L'exhaustivité des CAM restait toujours plus élevée que celle de la DO.

Tableau 20 : estimation de l'exhaustivité des trois sources pour le modèle initial (DC, DL) et stratifié selon la zone géographique et la localisation de la maladie (DC, DL, PC, PL, PD, RC, RL, RD)

Modèle	N _{est}	IC (95 %)	Exhaustivité des sources		
			DO	LABM	CAM
Initial	1767	1619-1915	44 [41-49]	25 [23-27]	51 [47-56]
Stratifié par localisation					
Pulmonaire	1206	1116-1296	51 [47-55]	30 [28-32]	56 [52-60]
Extra-pulmonaire	569	418-720	30 [24-41]	13 [10-17]	40 [32-55]
Total	1775				
Stratifié par zone					
Nord	500	388-612	37 [31-48]	16 [13-20]	49 [40-64]
Sud	493	409-577	27 [23-32]	31 [27-38]	39 [34-47]
Isère	292	252-332	51 [45-60]	43 [38-50]	54 [47-62]
Rhône	490	410-570	65 [56-77]	15 [13-18]	61 [53-73]
Total	1775				

4.5 Estimation du nombre total de cas et évaluation de la dépendance entre les sources en utilisant les modèles log-linéaires pour les cas bactériologiquement confirmés

Compte tenu des différences dans la définition de cas entre les trois sources, l'analyse de l'exhaustivité en utilisant les modèles log-linéaires a été réalisée sur les cas bactériologiquement confirmés (N=947).

4.5.1 Sélection du meilleur modèle par stratégie pas à pas descendante pour les cas bactériologiquement confirmés, déclarés par les 3 sources

De même que pour l'analyse de l'ensemble des cas déclarés, l'analyse pas à pas descendante à partir du modèle saturé montrait qu'un seul modèle présentait une bonne adéquation avec les données (tableau 21). Ce modèle incluait une interaction entre les sources DO et CAM (DC). Ce modèle avait une bonne adéquation ($G^2 = 2,83$) et les valeurs de l'AIC et du BIC étaient négatives. Ce modèle donnait une estimation du nombre total de cas de 1122, inférieure à la valeur estimée pour le modèle saturé de 1156. Cette valeur était proche de celle de la moyenne pondérée (1123), calculée à partir de toutes les estimations des différents modèles.

Tableau 21 : Analyse log-linéaire et modèles, estimation du nombre de cas de tuberculose bactériologiquement confirmés selon les modèles utilisés, région Rhône-Alpes, 1999-2000

Modèles	G2	p	ddl	AIC	BIC	Nobs	X	Nest	IC (95 %)
DC, CL, DL	0	1	0	0,0	0,0	947	209	1156	1053 – 1258
DL, CL	23,63	0,00	1	21,6	16,3	947	160	1107	1049 – 1166
DC, DL	0,02	0,9	1	-2,0	-7,3	947	213	1160	1084 – 1236
CL, DC	2,33	0,13	1	0,3	-5,0	947	80	1027	999 – 1055
DL, C	33,03	0,00	2	29,0	18,4	947	103	1050	1019 – 1081
DC, L	2,83	0,23	2	-1,2	-11,8	947	175	1122	1077 – 1167
CL, D	25,28	0,00	2	21,3	10,7	947	88	1035	1008 – 1062
D, C, L	33,44	0,00	3	27,4	11,5	947	107	1054	967 – 1141

G^2 : test du rapport de vraisemblance, p : adéquation du modèle, ddl : degré de liberté, Nest : nombre de cas, IC : intervalle de confiance du nombre estimé à 95 %.

4.5.2 Introduction des variables d'hétérogénéité pour les cas bactériologiquement confirmés déclarés par les 3 sources

Deux variables d'hétérogénéité ont été tour à tour introduites dans le modèle : la zone géographique telle que définie précédemment et la localisation de la maladie. Le tableau 22 présente les modèles retenus et les statistiques obtenues pour ces modèles, lorsque la probabilité d'être capturé par chacune des sources variait selon chacune des variables d'hétérogénéité.

4.5.2.1 Localisation de la maladie

L'introduction de cette variable, comme pour l'ensemble des cas déclarés faisait apparaître des interactions entre la localisation et les deux sources DO et CAM. Le meilleur modèle était celui prenant en compte ces deux interactions ainsi que l'interaction DO et CAM. Le nombre de cas estimé restait de 1122.

4.5.2.2 Zone géographique

Le même découpage de la région en 4 zones a été conservé pour cette analyse. L'introduction de cette variable faisait apparaître des interactions entre la zone et chacune des sources mais également une interaction entre DO et LABM, en plus de celle entre DO et CAM. Ce modèle n'avait pas une bonne adéquation du fait des interactions d'ordre 3, ne pouvant être prises en compte. Le meilleur modèle était celui ne prenant en compte que ces interactions et le nombre total de cas estimé était de 1144.

4.5.2.3 La zone géographique et la localisation de la maladie

L'introduction simultanée des deux variables précédentes faisait apparaître des interactions entre la zone et chacune des sources, entre la localisation et les sources DO et CAM et entre les sources DO et CAM et DO et LABM. Il n'y avait pas d'interaction entre la localisation et la zone. Les modèles n'avaient pas une bonne adéquation mais le modèle prenant en compte ces interactions avaient des valeurs d'AIC et de BIC négatives. Le nombre de cas estimés était de 1144.

Tableau 22 : estimation du nombre total de cas de tuberculose et des statistiques obtenues en introduisant des variables d'hétérogénéité (zone géographique et localisation), cas bactériologiquement confirmés, Rhône-Alpes, 1999-2000

MODELE	Nest	IC (95 %)	ddl	G ²	p	AIC	BIC
Modèle initial (DL, C)	1122	1067 - 1167	2	2,8	0,24	-1,2	-11,8
Localisation de la maladie							
DC, CL, DL, PD, PC, PL	1156	1082 -1230	3	2,1	0,54	-3,9	-18,9
L,DC,PD,PC	1122	1059 - 1185	6	5,0	0,54	-7,0	-37,1
Zone géographique							
DC, DL, CL, RC, RL, RD	1162	1086 - 1238	9	26,4	0,002	8,4	-36,8
DC, DL, RC, RD, RL	1144	1073 - 1215	10	26,7	0,003	6,7	-43,5
Localisation et zone géographique							
DC, DL, CL, RD, RC, RL, PD, PL, PC, RP	1162	1086 -1238	30	65,9	0,000	5,9	-144,8
DC, DL, PD, PC, RD, RC, RL	1144	1094 - 1194	35	69,1	0,001	-0,91	-176,7

4.5.3 Modèles retenus et estimations finales du nombre total de cas et de l'exhaustivité des sources pour les cas bactériologiquement confirmés, déclarés par les 3 sources

Pour calculer l'exhaustivité globale, le modèle initial a été utilisé. Pour calculer les niveaux d'exhaustivité par strate, les modèles retenus pour chacune des variables d'hétérogénéité et pour les deux variables introduites conjointement ont été utilisés.

Dans le modèle initial, il y avait une différence d'exhaustivité significative entre les sources ($p<0,001$), l'exhaustivité de la source LABM était plus faible que celles des sources DO et CAM.

L'exhaustivité de la source DO est meilleure pour les localisations strictement pulmonaires. Il en est de même pour la source CAM alors qu'aucune différence d'exhaustivité selon la localisation n'est observée pour la source LABM.

En stratifiant sur les zones géographiques, l'exhaustivité de la DO est meilleure dans les départements du Rhône (73 %) et de l'Isère (60 %). Elle est très faible dans la partie sud de la région (Ardèche, Drôme et Loire). L'exhaustivité de la source LABM est très faible dans le département du Rhône et dans la zone nord de la région (Ain, Savoie, Haute-Savoie).

Tableau 23 : estimation de l'exhaustivité des trois sources pour le modèle initial et stratifié selon la zone géographique et la localisation de la maladie

Modèle	N _{est}	IC (95 %)	Exhaustivité des sources		
			DO	LABM	CAM
Initial (DC,L)	1122	1077-1167	54 [52-56]	39 [37-40]	56 [54-58]
Stratifié par localisation (L, DC, PD, PC)					
Pulmonaire	931	877 - 985	56 [53-60]	39 [37-41]	58 [55-62]
Extra-pulmonaire	191	157 - 225	42 [36-51]	38 [32-46]	46 [39-55]
Total	1122	1059 - 1185	54 [51-57]	39 [37-41]	56 [53-59]
Stratifié par zone (DC,DL,RC,RD,RL)					
Nord	303	248 - 357	47 [40-57]	26 [22-31]	51 [43-62]
Sud	311	272 - 351	33 [29-38]	50 [44-57]	43 [38-49]
Isère	208	191 - 225	60 [55-66]	61 [56-67]	59 [54-64]
Rhône	322	286 - 357	73 [66-82]	23 [20-25]	67 [61-76]
Total	1144	1073 - 1215	53 [50-56]	38 [36-40]	55 [52-59]

Tableau 24 : estimation de l'exhaustivité des trois sources selon la zone géographique et la localisation de la maladie (modèle DC,DL,PC,PL,PD,RC,RD,RL)

Modèle	N _{est}	IC (95 %)	Exhaustivité des sources		
			DO	LABM	CAM
Zone Nord					
pulmonaire	253	225 – 281	47 [43-53]	26 [23-29]	54 [49-61]
Extra-pulmonaire	50	28 – 72	44 [31-78]	26 [18-46]	36 [25-64]
Zone Sud					
pulmonaire	244	216 – 272	36 [32-40]	48 [43-54]	45 [40-51]
Extra-pulmonaire	67	47 – 87	24 [18-34]	57 [44-80]	36[28-51]
Isère					
pulmonaire	175	163 – 187	63 [59-68]	67 [63-72]	59 [56-64]
Extra-pulmonaire	33	26 – 40	42 [35-54]	27 [22-35]	55 [45-70]
Rhône					
pulmonaire	276	261 – 291	75 [71-79]	22 [21-23]	69 [65-73]
Extra-pulmonaire	46	35 – 57	61 [49-79]	26 [21-34]	59 [48-76]
Total	1144	1094 - 1194	53 [51-55]	38 [36-40]	55 [53-57]

4.6 Analyse de l'hétérogénéité d'exhaustivité pour les sources DO et CAM

L'hétérogénéité d'exhaustivité pour les sources a été évaluée pour les sources DO et CAM en stratifiant selon les résultats bactériologiques (tableau 25) : l'exhaustivité est similaire dans les deux sources pour les cas bactériologiquement confirmés et pour les cas dont les résultats bactériologiques demeuraient inconnus. Elle diffère pour les cas bactériologiquement négatifs (exhaustivité très faible pour la source DO). Pour la source CAM, l'exhaustivité ne varie pas en fonction des résultats bactériologiques.

L'étude du coefficient de corrélation des taux d'exhaustivité après stratification sur les résultats bactériologiques ne met pas en évidence de dépendance entre les sources DO et CAM ($r=0,22$, $p=0,85$).

Tableau 25 : estimation du nombre total de cas pour les sources DO et CAM, exhaustivité des deux sources et intervalles de confiance, selon les résultats bactériologiques, région Rhône-Alpes, 1999-2000

Résultats bactériologiques	X ₁₁	X ₁₂	X ₂₁	N	IC (95 %)	Exh DO	IC (95 %)	Exh CAM	IC (95 %)
Positifs	605	628	396	959	925- 993	63,1	60,9 - 65,4	65,5	63,3 - 67,9
Négatifs	4	117	2	196	15 - 376	2,0	1,3 - 5,3	59,7	37,0 - 154,6
Inconnus	177	154	91	299	272 - 326	59,2	54,3 - 65,1	51,5	47,2 - 56,6

5 DISCUSSION

5.1 Synthèse des principaux résultats

La méthode capture-recapture a été appliquée aux données de tuberculose pour estimer l'exhaustivité du système de déclaration obligatoire dans la région Rhône-Alpes. Les données de ce système de surveillance national ont été comparées aux résultats des laboratoires d'analyses de biologie médicale et aux demandes d'exonération du ticket modérateur des caisses d'assurance maladie. La participation des laboratoires et des caisses d'assurance maladie était très satisfaisante, bien que variable selon les départements pour les laboratoires d'analyses de biologie médicale. Dans le Rhône, les chiffres observés pour la source LABM laissaient supposer une forte sous-déclaration au sein même des laboratoires participants. Certaines informations étaient mal renseignées

notamment celles concernant les résultats des demandes de mise en culture des prélèvements dans les sources DO et CAM.

Le nombre total de cas observés à partir des données des trois sources était de 1306 cas, dont 947 cas bactériologiquement confirmés.

Le nombre de cas de tuberculose diagnostiquée a été estimé à 1767 cas pour l'ensemble de la région et pour la période 1999-2000 avec un intervalle de confiance allant de 1619 à 1915. L'incidence moyenne annuelle pouvait être estimée à 15,7 /100 000 habitants. L'analyse montrait qu'il existait une dépendance entre les sources DO et CAM et entre les sources DO et LABM. Au total, les trois systèmes ne permettaient d'identifier que 74 % des cas de tuberculose.

L'exhaustivité des données était légèrement meilleure pour les caisses d'assurance maladie (51 %) que pour la déclaration obligatoire (44 %).

L'exhaustivité était meilleure dans les cas de localisation pulmonaire, pour les trois sources, et pour les départements les plus importants de la région (Rhône et Isère), hormis pour la source LABM dans le Rhône.

En effectuant l'analyse sur les cas bactériologiquement confirmés, le nombre de cas était estimé à 1122 pour l'ensemble de la région et pour la période 1999-2000 avec un intervalle de confiance de 1077 à 1167. L'analyse montrait qu'il existait une dépendance entre les sources DO et CAM. L'exhaustivité de la source LABM (39%) restait inférieure à celle des sources DO (56%) et CAM (58%). Les trois systèmes permettaient d'identifier 84% des cas de tuberculose bactériologiquement confirmés.

Pour la source LABM, l'exhaustivité ne dépendait pas de la localisation alors que pour les sources DO et CAM, elle restait meilleure dans les cas de localisation pulmonaire, hormis dans les départements du nord de la région. L'exhaustivité de la DO était meilleure dans les deux départements les plus importants de la région. Elle atteignait 73% pour le département du Rhône.

5.2 Conditions d'application de la méthode capture-recapture

L'estimation du nombre de cas de tuberculose en Rhône-Alpes en 1999-2000 a été obtenue en utilisant la méthode capture-recapture et les modèles log-linéaires. L'utilisation de cette méthode est soumise à des conditions qui, si elles ne sont pas respectées, entachent la validité des résultats.

- La définition de cas devait être identique dans les différentes sources : la définition de cas pour l'ensemble des cas de tuberculose diagnostiqués dans les sources DO et CAM

était identique et correspondait à la notion de diagnostic de tuberculose maladie chez un patient quels que soient les résultats des prélèvements biologiques. Par contre, elle différait dans la source LABM puisque seules les personnes présentant un résultat biologique positif dans un prélèvement quelconque pouvaient être repérées dans cette source. Cette variation de la définition de cas entre une source et les deux autres conduisait très probablement à une surestimation du nombre total de cas dans les sources DO et CAM et du nombre global de cas et à une sous-estimation de l'exhaustivité de la source LABM. Cette condition n'étant pas respectée pour l'ensemble des cas de tuberculose diagnostiquée et pouvant remettre en cause la validité des résultats, a conduit à une analyse restreinte aux cas bactériologiquement confirmés. Ces derniers représentaient 73% de l'ensemble des cas. Par ailleurs, pour l'ensemble des cas, le nombre de cas inclus par la source CAM était le plus élevé (901) alors que la proportion de cas bactériologiquement confirmés était la plus faible (69,7%), ce qui pouvait faire craindre une inclusion de faux cas, par exemple des cas de primo-infection, dans cette source.

- Les cas devaient appartenir à la même zone géographique : du fait de la restriction de l'étude à la région Rhône-Alpes, certains cas, domiciliés en limite de la région pouvaient avoir été hospitalisés en dehors de la zone ou avoir effectué leur diagnostic biologique dans un laboratoire d'un département voisin alors qu'ils auraient effectué leur demande d'exonération du ticket modérateur dans leur département de domicile. Pourtant, compte tenu de l'importance de l'offre de soins au niveau régional, ce phénomène était probablement marginal. Pour la source LABM, le recours à des laboratoires secondaires pour réaliser le diagnostic a pu conduire à sous-estimer le nombre de cas régionaux du fait des difficultés à connaître le code postal des patients positifs dans ces laboratoires de référence. Cependant, dans ces situations, un retour au prescripteur a permis la plupart du temps d'identifier le domicile du patient.
- Les cas devaient appartenir à la même période : les dates d'inclusion pouvaient différer selon les sources : pour la DO, il s'agissait de la date de la mise en route du traitement qui peut s'effectuer avant la confirmation biologique. Pour les LABM, il s'agissait de la date de confirmation biologique et pour les CAM de la date de demande d'exonération du ticket modérateur qui pouvait être retardée par rapport au diagnostic et à la mise sous traitement. De ce fait, par exemple, une personne pouvait être mise sous traitement en décembre 2000 et n'avoir la confirmation bactériologique et la demande d'ALD début 2001. Lors de la première étude menée en 1992-1993, une estimation de la proportion de cas communs non identifiés du fait des différences de date d'identification d'un même cas entre les sources avait été réalisée. Les différences entre DO et LABM n'étaient pas

applicables car pour la DO, la date de déclaration avait été prise en compte, souvent retardée par rapport à la date de mise sous traitement. Entre DO et CAM, la proportion de cas non identifiés était estimée à 4 %, avec des variations selon les départements. Pour cette étude, les cas déclarés dans les 6 premiers mois de l'année 2001 ont été identifiés pour repérer les cas mis sous traitement en 2000 et déclarés tardivement.

- Tous les cas communs et seulement ceux-ci devaient être identifiés : l'absence d'identifiant commun aux trois sources rendait l'identification des cas communs complexe et difficile. Cependant la recherche croisée par combinaisons successives de plusieurs critères permettait d'espérer une identification quasi exhaustive des cas communs.
- Les sources devaient être indépendantes : l'analyse des dépendances entre les sources pour l'ensemble des cas diagnostiqués, selon la méthode de Witten et selon les modèles log-linéaires, montrait qu'il existait des dépendances entre DO et LABM et entre DO et CAM. La dépendance forte entre DO et CAM pouvait s'expliquer par le fait que certains médecins des caisses peuvent être eux-mêmes déclarants dans le système de maladie à déclaration obligatoire, selon les départements. La dépendance entre DO et LABM était plus difficile à interpréter. Ces deux dépendances ont été prises en compte dans les modèles log-linéaires retenus pour l'ensemble des cas. Pour les cas bactériologiquement confirmés, seule la dépendance entre DO et CAM persistait. L'étude de l'exhaustivité en comparant ces deux sources, stratifiée sur les résultats bactériologiques montrait qu'il existait une hétérogénéité d'exhaustivité selon ces résultats.
- La capture des cas devait être homogène : l'introduction de certaines variables dans les modèles log-linéaires permettait de prendre en compte l'hétérogénéité de capture, en particulier la localisation de la maladie pour laquelle il apparaissait une hétérogénéité de capture selon les sources. Cependant, les modèles prenant en compte les variables d'hétérogénéité comme le département n'avaient pas une bonne adéquation ($p>0,05$) bien que les valeurs du BIC soient négatives, du fait d'interactions d'ordre 3 qui ne pouvaient être prises en compte.

La prise en compte nécessaire des conditions de validité de la méthode pour les données de cette étude permet de souligner que les résultats de l'analyse sur les cas bactériologiquement confirmés rend compte d'estimations plus fiables de l'exhaustivité que de celle sur l'ensemble des cas diagnostiqués. Mais cette analyse ne permet pas d'estimer le nombre total de cas de tuberculose diagnostiqués dans la région, sachant que pour une part non négligeable des cas déclarés par les sources DO et CAM, les résultats des cultures ne sont pas connus.

5.3 Analyse des résultats

Les résultats obtenus montraient que globalement l'exhaustivité du système de déclaration obligatoire était médiocre, sans grande évolution depuis la dernière étude nationale réalisée sur la période 1992-1993 et estimant à environ 50% l'exhaustivité de la DO.

Pour l'estimation sur l'ensemble des cas, le meilleur modèle montrait une dépendance entre la source DO et la source CAM et entre la source DO et la source LABM. Ce modèle avait une bonne adéquation statistique ($p>0,05$, AIC et BIC négatifs). Mais la différence de définition de cas pouvait être responsable d'une mauvaise estimation globale du nombre total de cas et sous estimer l'exhaustivité de la DO obtenue par ce modèle (44 %).

La dépendance observée entre la zone géographique et la DO suggérait que l'intensité de la déclaration obligatoire variait selon des caractéristiques liés à la zone (urbanisation, type de l'offre de soins...). Cette variation avait déjà été mise en évidence lors de la première étude. Les dépendances observées entre la zone géographique et les sources LABM et CAM étaient plus difficilement interprétables, pouvant être liées à des biais d'information. Ceci était vérifié pour les données LABM du département du Rhône du fait d'une très mauvaise exhaustivité de réponse des laboratoires. Pour l'estimation du nombre de cas par département et de l'exhaustivité, le modèle prenant en compte ces interactions a été utilisé. Les résultats par zone géographique, même s'ils sous-estimaient globalement l'exhaustivité de la DO étaient importants à prendre en compte pour mieux cibler l'incitation à la déclaration. L'hypothèse d'une meilleure déclaration dans les zones de plus forte incidence, pour lesquelles les médecins sont plus souvent en situation de déclaration pouvait être soulevée. De même, les résultats selon la localisation confirmaient l'idée de meilleure déclaration dans les cas de localisation pulmonaire du fait du plus grand risque de diffusion de la maladie.

L'analyse sur les cas bactériologiquement confirmés donnait une meilleure estimation de l'exhaustivité pour ces cas. L'exhaustivité demeurait faible, ne concernant qu'à peine plus de la moitié des cas pour la DO (54%). L'introduction des variables d'hétérogénéité confirmait les variations observées sur l'ensemble des cas, qu'il s'agisse de la localisation ou de la zone géographique.

La comparaison des résultats de cette étude avec ceux de la première étude ne pouvait être réalisée que pour le département du Rhône, en croisant les sources DO et CAM deux à deux, les données des LABM n'étant pas disponibles pour la première étude. L'exhaustivité dans cette étude était estimée à 73 % [70,4 -75,9] alors qu'elle n'était estimée qu'à 43 % lors de la première étude.

6 CONCLUSION

La surveillance de la tuberculose est justifiée par la gravité de la maladie et ses caractéristiques évolutives au cours du temps, le risque de recrudescence persistant actuellement du fait de la situation mondiale, en particulier de l'importance des formes résistantes.

L'utilisation de la méthode capture-recapture est une alternative à des études longues et coûteuses permettant une évaluation quantitative d'un système de surveillance.

En 1999 et 2000, près de la moitié des tuberculoses diagnostiquées n'a pas été déclarée par les médecins. Ces résultats sont proches de ceux obtenus lors de l'étude menée en 1992-1993.

Ces résultats justifient la nouvelle réglementation sur la déclaration obligatoire incluant les responsables de laboratoires d'analyses de biologie médicale dans le système de déclaration et la sensibilisation des déclarants à la déclaration. Cependant cette évolution du système de DO ne permettra pas d'améliorer la déclaration des cas non bactériologiquement confirmés mais répondant à la définition de mise sous traitement pour lesquels des sensibilisations spécifiques des médecins doivent être engagées.

REFERENCES

- Cormack R.M. Problems with Using Capture-Recapture in Epidemiology: An Example of a Measles Epidemic. *J Clin Epidemiol* 1999, 52 (10): 909-914.
- Gallay A., Nardone A., Vaillant V. et al. La méthode capture-recapture appliquée à l'épidémiologie : principes, limites et applications. *Rev Epidemiol Sante Publique* 2002, 50 (2) : 219-32.
- Gallay A., Vaillant V., Bouvet P. et al. How Many Foodborne Outbreaks of *Salmonella* Infection Occurred in France in 1995? Application of the Capture-Recapture Method to Three Surveillance Systems. *Am J Epidemiol* 2000, 152 (2) : 171-177.
- Guillem P., Lavergne C., Cans C. Comment calculer l'exhaustivité d'un registre de morbidité ? L'exemple du registre des handicaps de l'enfant et observatoire périnatal de l'Isère. *Rev Epidemiol Sante Publique* 2000, 48 (1) : 41-51.
- Hook E.B., Regal R.R. Capture-Recapture Methods in Epidemiology: Methods and Limitations. *Epidemiol Rev* 1995, 17 (2) : 243-264.
- Hook E.B., Regal R.R. The Value of Capture-Recapture Methods Even for Apparent Exhaustive Surveys. The Need for Adjustment for Source of Ascertainment Intersection in Attempted Complete Prevalence Studies. *Am J Epidemiol* 1992, 135 (9): 1060-1067.
- Nardone A., Decladt B. Evaluation épidémiologique du système de surveillance de la légionellose en France en 1998. Institut de veille sanitaire, Saint-Maurice. Décembre 2000, 45 p.
- Perrocheau A. Evaluation de la surveillance des infections à méningocoques en France en 1996 par la méthode capture-recapture. Institut de veille sanitaire, Saint-Maurice. Mai 2001, 41 p.
- Wittes J.T., Colton T., Sidel V.W. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chron Dis* 1974, 27: 25-36.

ANNEXE 1

Fiche de DO de la tuberculose

ANNEXE 2
Fiche de recueil de données des Laboratoires d'Analyses de Biologie Médicale



MINISTÈRE DE L'EMPLOI
ET DE LA SOLIDARITÉ

Étude Tuberculose
Rhône-Alpes 1999-2000



NSTITUT DE VEILLE SANITAIRE

Questionnaire à retourner sous pli confidentiel à :
Cellule inter-régionale d'épidémiologie d'intervention

DRASS Rhône-Alpes
107, rue Servient
69418 Lyon Cedex 03

DIAGNOSTIC BACTERIOLOGIQUE DE TUBERCULOSE

Laboratoire : _____

Adresse : _____

Personne à contacter : _____ Téléphone : _____ Fax : _____

Médecin prescripteur :

Nom : _____

Adresse : _____

Téléphone : _____

Caractéristiques du malade :

Initiale du nom : _____ Prénom : _____
Sexe : M F Date de naissance : |_____|_____|_____|
Code postal du domicile : |_____|_____|_____|_____|

Bactériologie

Recherche dans l'expectoration, les prélèvements bronchiques ou le tubage :

- de bacilles acido-alcoolo-résistants à l'examen direct positive
 négative
date des résultats |_____|_____|_____|

- de bacilles tuberculeux à la culture positive
 négative
 non faite
date des résultats |_____|_____|_____|

Présence d'un bacille tuberculeux dans des prélèvements d'autre origine Oui Non
date des résultats |_____|_____|_____|

Mycobactérie identifiée : _____

Dans quel laboratoire a été réalisée la culture : _____

ANNEXE 3
Fiche de recueil de données des Caisses d'Assurance Maladie



MINISTÈRE DE L'EMPLOI
ET DE LA SOLIDARITÉ

Étude Tuberculose

Rhône-Alpes 1999-2000



NSTITUT DE VEILLE
SANITAIRE

Questionnaire à retourner sous pli confidentiel à :

Cellule inter-régionale d'épidémiologie d'intervention

DRASS Rhône-Alpes

107, rue Servient

69418 Lyon Cedex 03

DEMANDE D'EXONERATION DU TICKET MODERATEUR

Caisse d'Assurance Maladie de : _____

Personne à contacter : _____ Téléphone : _____

Date de la demande : |_____|_____|____|

Médecin demandeur :

Nom : _____

Adresse : _____

Téléphone : _____

Caractéristiques du malade :

Initiale du nom : _____ Prénom : _____

Sexe : M F Date de naissance : |_____|_____|____|

Code postal du domicile : |_____|_____|_____|_____|____|

Date de mise en route du traitement anti-tuberculeux : |_____|_____|____|

Confirmation bactériologique

Recherche dans l'expectoration, les prélèvements bronchiques ou le tubage :

- de bacilles acido-alcoolo-résistants à l'examen direct positive
 négative

- de bacilles tuberculeux à la culture positive
 négative
 non faite

Présence d'un bacille tuberculeux dans des prélèvements d'autre origine Oui

Localisation(s)

Pulmonaire
 Pleurale
 Urinaire
 Ganglionnaire hilare
 Autres *Préciser :* _____

Méningée
 Ostéo-articulaire
 Génitale
 Ganglionnaire autre