

---

Détection précoce automatisée des épidémies  
de gastro-entérites d'origine hydrique  
à partir des données de vente  
ou de remboursement des médicaments

---

Étude de faisabilité :  
choix des données sanitaires et des départements pilotes

<b>1</b>	<b>Introduction</b>	<b>p. 2</b>
	1.1   Contexte	p. 2
	1.2   Démarche	p. 3
<b>2</b>	<b>Objectifs de l'étude</b>	<b>p. 4</b>
<b>3</b>	<b>Matériel et méthodes</b>	<b>p. 5</b>
	3.1   Définition des sujets exposés et non exposés	p. 5
	3.2   Bases de données utilisées	p. 6
	3.3   Critères de choix des départements pilotes	p. 7
<b>4</b>	<b>Résultats</b>	<b>p. 9</b>
	4.1   Données sanitaires	p. 9
	4.2   Taille des unités de distribution : choix de la catégorie cible	p. 12
	4.3   Classement des départements	p. 14
<b>5</b>	<b>Discussion</b>	<b>p. 21</b>
	5.1   Les épidémies passées au regard du système de détection automatisé	p. 21
	5.2   Limites des bases de données de l'Assurance maladie pour l'alerte	p. 24
	5.3   Retour sur la définition de l'événement critique	p. 26
<b>6</b>	<b>Perspectives opérationnelles</b>	<b>p. 29</b>
	6.1   Échéances	p. 29
	6.2   Implication potentielle du projet pour les Ddass	p. 29
	6.3   Ajuster le seuil d'alerte sur le bénéfice de santé publique escompté	p. 30
<b>7</b>	<b>Références bibliographiques</b>	<b>p. 31</b>
<b>8</b>	<b>Annexes</b>	<b>p. 32</b>

# Détection précoce automatisée des épidémies de gastro-entérites d'origine hydrique à partir des données de vente ou de remboursement des médicaments

## Étude de faisabilité : choix des données sanitaires et des départements pilotes

**Rédaction :** Pascal Beaudeau, Damien Mouly et Delphine Lauzeille.

### Remerciements

Nous remercions les membres du groupe de travail "Amélioration des investigations d'épidémies d'origine hydrique" pour leurs conseils avisés et particulièrement M. Daniel Rivière (Cire Centre-Ouest), M. Christian Mannschott (Ddass 54) et Mme Anne Novelli (bureau de l'eau, Direction générale de la santé) qui ont directement contribué à ce travail. Nous remercions également les services de la Direction générale de la santé, notamment le bureau de l'eau et M. Henry Davezac qui a réalisé l'extraction de données de la base SISE-eaux.

Les informations concernant les évolutions récentes des bases de données de la CnamTS ont été fournies par M. Yvon Merlière.

La rédaction de ce rapport a bénéficié des remarques de Mmes Henriette de Valk, Véronique Vaillant, MM. Gilles Brücker, Christian Mannschott, Jean-Nicolas Ormsby et Claude Tillier.

Composition du groupe de travail :

<b>Prénom</b>	<b>Nom</b>	<b>Fonction</b>	<b>Organisme</b>
Marie	Baville	Ingénieur du génie sanitaire	Ministère chargé de la Santé - DGS
Pascal	Beaudeau	Épidémiologiste, coordinateur eau et santé	InVS-DSE
Delphine	Caamano	Responsable de l'Unité d'évaluation des risques liés à l'eau	Afssa
Estelle	Checlair	Technicienne épidémiologiste	InVS-DSE
Catherine	Chubilleau	Épidémiologiste	Cire Centre-Ouest, Drass du Centre
Henriette	de Valk	Épidémiologiste	InVS-DMI
Loïc	Favenec	Professeur, parasitologue	CHU de Rouen
Eugénia	Gomes	Épidémiologiste	InVS-DSE
Martine	Ledrans	Responsable du Département santé environnement	InVS-DSE
Jean	Lesne	Professeur, microbiologiste	ENSP
Christian	Mannschott	Ingénieur du génie sanitaire	Ddass de Meurthe-et-Moselle
Damien	Mouly	Épidémiologiste	InVS-DSE
Anne	Novelli	Chargée d'étude	Ministère chargé de la Santé - DGS
Jeanine	Stoll	Épidémiologiste	Cire Centre-Est, Drass de Bourgogne
Sylvie	Pérelle	Virologue	Afssa
Véronique	Vaillant	Épidémiologiste	InVS-DMI

### Membres anciens

Delphine	Lauzeille	Stagiaire PROFET	InVS-DSE
Eric	Dubois	Virologue	Afssa
Pierre	Le Cann	Virologue	Ifremer
François	Mansotte	Ingénieur du génie sanitaire	Ddass de Guyane
Daniel	Rivière	Ingénieur du génie sanitaire	Drass du Centre

## 1.1 | Contexte

Une épidémie se définit comme l'apparition d'un nombre de cas de maladie supérieur au nombre de cas attendus pendant une période donnée et au sein d'une population donnée. Face à un tel événement, une investigation cherche (i) à identifier et à juguler une source de contamination commune ; (ii) à améliorer la connaissance des déterminants et du contexte ayant conduit à l'épidémie. Dans les épidémies de gastro-entérites d'origine hydrique, l'investigation doit non seulement chercher à établir le rôle de l'eau mais aussi identifier les indicateurs pertinents, sanitaires ou environnementaux, utilisables dans un système d'alerte.

L'étape de détection des épidémies d'origine hydrique est particulièrement délicate, de par la difficulté à identifier les signaux précurseurs ou précoces (environnementaux ou sanitaires) et de par la perte d'information au moment de la remontée du signal aux autorités en charge de la gestion des situations épidémiques (Ddass). Or, une détection précoce des épidémies d'origine hydrique se traduit directement en nombre de cas de maladies évités.

Hormis le cas particulier des maladies faisant l'objet de systèmes de surveillance ad hoc, notamment maladies à déclaration obligatoire, les principales sources

d'information utilisées à l'heure actuelle pour détecter les épidémies d'origine hydrique se situent :

- au niveau des médecins, généralistes ou urgentistes, et des particuliers pour la détection des cas groupés de gastro-entérites ;
- au niveau des Ddass pour les analyses d'eau révélant une contamination microbiologique du réseau.

La capacité de détection des signaux constitue la principale limite de ces deux systèmes de surveillance. Ceci s'explique pour les médecins par une faible sensibilité car un médecin généraliste ne voit en moyenne que moins d'un cas de gastro-entérite par jour de consultation, à laquelle s'ajoute une perte d'information lors de la remontée du signal (démarche non systématique, à l'appréciation du médecin). Concernant le système de surveillance basé sur les résultats du contrôle sanitaire des eaux distribuées, la fréquence des analyses est directement conditionnée par la taille des installations (tableau 1). L'efficacité de ce système est donc limitée aux unités de distribution importantes qui sont cependant rarement impactées par les pollutions microbiologiques.

**Tableau 1 - Fréquences d'analyses du contrôle sanitaire des eaux d'alimentation régies par l'article R.1321 - 15 et l'annexe 13-2 du Code de la santé publique pour les paramètres microbiologiques**

Population desservie (en nombre d'habitants)	Fréquences annuelles d'échantillonnage		
	Au point de mise en distribution (sortie usine)	Au robinet du consommateur	Total
0-499	1 à 2	2 à 4	3 à 6
500-1999	2	6	8
2000-4999	3	9	12
5000-14 999	5	12	17
15 000-29 999	6	25	31
30 000-99 999	12	61	73
>100 000	6 à 144	150 à 800	156 à 944

De ce constat se dégage l'intérêt :

- de développer d'autres sources d'alertes actuellement trop négligées, telles que les plaintes des usagers ou les incidents ou accidents d'exploitation. Ce sujet sera développé dans le "guide d'investigation des épidémies d'origine hydrique", à paraître courant 2006 ;

- d'exploiter pour l'alerte les informations sanitaires des systèmes automatisés qui se sont développés au cours de la dernière décennie. Ce rapport vise à établir la faisabilité d'un tel projet.

## 1.2 | Démarche

La question des épidémies liées à l'eau potable a été soulevée et a connu ses premiers développements au sein du groupe de travail du programme "Amélioration des investigations des épidémies d'origine hydrique" de l'InVS. L'arrêt de ce programme en 2006 rend au projet d'alerte automatisée son autonomie.

Le projet de système d'alerte automatisé ne vient donc pas dans un contexte vierge. Un algorithme développé et programmé par l'InVS permet déjà d'identifier à partir des données de remboursement des prescriptions médicamenteuses les cas de gastro-entérites aiguës (GEA) avec une bonne spécificité. Cette méthode est utilisée dans le cadre de l'étude écologique multicentrique "Turbidité et gastro-entérites" et ponctuellement, pour évaluer l'impact d'épidémies.

L'objet de ce rapport se limite donc d'une part, à l'évaluation et au choix des sources de données sanitaires dans une perspective de détection précoce des épidémies d'origine hydrique et, d'autre part, au classement des départements éligibles comme départements pilotes.

La partie scientifique et technique du projet (géocodage des cas, recherche des clusters, test d'un éventuel lien avec le découpage spatial de la distribution de l'eau, définition des seuils d'alerte) ne relève pas de ce rapport mais des étapes ultérieures, s'il y a lieu.

## 2 Objectifs de l'étude

L'InVS se propose d'explorer la faisabilité d'un système d'alerte sanitaire dédié à la surveillance des épidémies de GEA d'origine hydrique, basé sur une approche automatisée à partir de l'analyse des ventes ou des remboursements de médicaments. Cette première phase comprend :

1. l'évaluation comparée des bases de données sanitaires disponibles ;
2. l'élaboration d'une liste de départements éligibles comme départements pilotes ;

3. la mise en place d'un outil automatisé expérimental au niveau départemental ;
4. la définition d'un algorithme de discrimination des cas groupés de GEA liés à l'eau et de seuils d'alerte ;
5. l'évaluation du système et réflexion sur l'opportunité d'étendre le système au niveau national.

Seuls les deux premiers objectifs sont présentés dans ce protocole. L'atteinte des trois derniers objectifs sera conditionnée par la collaboration de l'Assurance maladie (AM) et l'affectation de moyens spécifiques.

### 3.1 | Définition des sujets exposés et non exposés

**Une unité de distribution (UDI)** désigne le réseau ou la partie physique du réseau de distribution qui délivre une eau de qualité homogène<sup>1</sup>, hors pollution accidentelle ou dégradation dans le réseau. La population alimentée par une même UDI est donc semblable du point de vue de l'exposition potentielle aux pathogènes véhiculés par l'eau distribuée. Dans la suite du protocole, l'UDI représentera l'unité géographique élémentaire d'exposition.

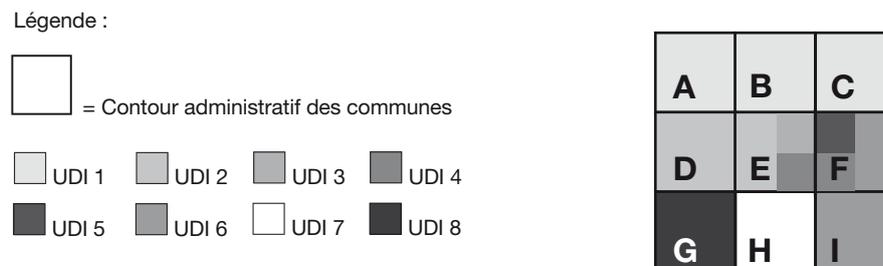
Les UDI sont de tailles très variables : les populations desservies comptent au minimum une dizaine d'habitants en secteur rural et dépassent le million en banlieue parisienne. Cette hétérogénéité a deux conséquences pour la population :

1. non prise en compte des populations alimentées par des petites UDI dans le système d'alerte automatisé (limite liée à la puissance statistique) ;
2. connaissance inégale de l'exposition des populations (limite liée à l'échantillonnage).

Ces problèmes sont développés dans les paragraphes suivants.

Le référencement géographique des cas dans les bases de données sanitaires peut être précis (adresse de résidence) ou imprécis (commune de résidence, code postal...), mais ne se fait en aucun cas par UDI... L'impact de l'adéquation imparfaite entre le découpage administratif des communes et le contour des UDI qui les alimentent (figure 1) sera discuté.

**Figure 1 - Possibilités de recouvrements entre les communes et les UDI qui les alimentent**



Selon les estimations les plus récentes, plus de 99 % de la population française est alimentée par un réseau public d'eau potable [1]. Nous pouvons donc considérer, pour chaque département, que l'ensemble de la population est concerné par l'étude et que le statut exposé/non exposé se définit au niveau de l'unité statistique (UDI) par la présence (exposé) ou non (non exposé) d'une contamination microbiologique de l'eau de distribution (cf. paragraphe 3.2.2).

L'approche écologique<sup>2</sup> de l'étude nécessite cependant de faire plusieurs hypothèses propres à la population :

- consommation homogène d'eau du robinet entre UDI et au sein de chaque UDI ;
- sensibilité homogène vis-à-vis du risque de GEA entre UDI et au sein de chaque UDI.

<sup>1</sup> Définition SISE-eaux : ensemble de tuyaux connexes de distribution dans lesquels la qualité de l'eau est réputée homogène, faisant partie d'une même unité de gestion et d'exploitation, donc géré par un seul exploitant et possédé par un seul et même propriétaire.

<sup>2</sup> Au sens utilisé en épidémiologie, à savoir l'estimation de l'exposition est collective et non individuelle.

## 3.2 | Bases de données utilisées

Deux types de sources de données ont été exploités dans le cadre de cette étude de faisabilité : (i) les systèmes de collecte de ventes ou de prescriptions médicamenteuses

analysés pour leur aptitude à constituer un système d'alerte ; et (ii) les données de la base SISE-eaux qui ont servi de support pour le classement des départements.

### 3.2.1 | Bases de données de vente ou de remboursement de médicaments

L'AM gère la liquidation des actes de santé et dispose de deux bases contenant les données de remboursement de médicaments (bases Erasme et SNIIR-AM). Deux systèmes de collecte nationale des données de ventes médicamenteuses existent par ailleurs, conçus principalement pour répondre à des objectifs commerciaux (marketing, finance, gestion des

stocks...) [2]. Au total, quatre bases de données ont fait l'objet d'une étude comparative :

- l'AM (bases Erasme et SNIIR-AM) ;
- le Groupement pour l'élaboration et la réalisation de statistiques (Gers) ;
- IMS-Health (base XPONENT).

### 3.2.2 | Base de données SISE-eaux

Depuis 1995, la Direction générale de la santé assure le suivi de la qualité des eaux distribuées grâce à la base de données nationale SISE-eaux. Cette base est alimentée au niveau départemental par les Ddass et comporte l'ensemble des résultats analytiques du contrôle sanitaire disponibles depuis plus de 10 ans. L'ampleur de l'historique disponible varie selon les départements car l'implémentation du système dans les Ddass s'est échelonnée de la fin des années 80 à la fin des années 90.

La base de données SISE-eaux est l'outil de choix pour estimer le niveau d'exposition des populations aux toxiques et aux pathogènes véhiculés par l'eau en fonction de l'UDI qui les alimentent. Il n'existe pas d'autre source d'information couvrant la totalité des UDI françaises. La fréquence d'échantillonnage diffère cependant beaucoup selon la taille de l'UDI (tableau 1), cette hétérogénéité se retrouve au niveau de la base SISE-eaux.

Plusieurs paramètres figurant dans la réglementation pour contrôler la qualité microbiologique des eaux d'alimentation sont renseignés de façon continue dans la base SISE-eaux. Parmi les indicateurs bactériens utilisés

en routine, *Escherichia coli* et les entérocoques sont d'origine fécale. Leur présence dans les échantillons analysés signe une probabilité de coexistence dans le réseau d'eau de microorganismes d'origine fécale pathogènes pour l'homme, non mesurés en routine. La méthode analytique normalisée pour ces deux paramètres reposant sur une technique de mise en culture, les résultats sont exprimés en Unités Formant Colonies pour un échantillon de 100mL d'eau (UFC/100mL). Concernant le risque endémique, un lien entre contamination fécale et taux d'incidence des gastro-entérites a été établi [3]. Bien qu'il soit raisonnable de penser que la probabilité d'infection croît avec les concentrations des indicateurs de contamination fécale, la relation dose-réponse n'est pas quantifiée<sup>3</sup>. *A fortiori*, la valeur prédictive d'un niveau de contamination vis-à-vis du risque épidémique n'est pas connue.

Les données 2003 ont été utilisées pour le classement des départements éligibles comme pilotes. Certains départements n'ont pu être classés car ils n'ont pas alimenté (Haute-Corse, Moselle, Lozère, Saint-Pierre-et-Miquelon, Mayotte) ou alimenté partiellement (Var) la base SISE-eaux au moment de la requête.

<sup>3</sup> Il existe des exceptions. Haas et al. (1999) proposent par exemple une correspondance entre la concentration en coliformes fécaux dans l'eau et la concentration en *Salmonella* sp. (1:54 à 1:5400) qui permet l'évaluation du risque, des relations dose/effet existant pour *Salmonella* sp.

## 3.3 | Critères de choix des départements pilotes

### 3.3.1 | Puissance statistique

La puissance statistique se définit comme la capacité d'une étude à mettre en évidence une différence (par exemple de fréquence d'un caractère ou de moyenne d'une variable) entre deux échantillons, quand cette différence existe. Dans le cas présent, cette "différence" est indiquée par le risque relatif (RR) qui est le rapport d'incidence des GEA chez les exposés et les non exposés. Un RR supérieur à 1 traduit une surincidence de la maladie. Le calcul de puissance a pour objectif de déterminer, *a priori*, le nombre de sujets exposés à inclure dans l'étude pour mettre en évidence un risque dont le niveau est choisi. Dans le cadre de cette étude, le calcul sera effectué pour des niveaux de RR allant du risque hyperendémique (RR = 1,4) au risque épidémique très élevé (RR = 20).

Le calcul a été réalisé avec le logiciel Epi-Info 6 en fixant les paramètres suivants :

- test unilatéral au risque de première espèce :  $\alpha = 0,05$  ;
- puissance recherchée :  $1-\beta = 0,8$  ;
- l'incidence hebdomadaire de la maladie chez les non exposés (bruit de fond de la maladie) est calculée à partir de l'incidence annuelle des GEA médicalisées (8 000 cas / 100 000 habitants - source : Réseau Sentinelles) corrigée par des facteurs de non-

exhaustivité du système de l'AM. Ces facteurs sont liés (i) à la structure du système (taux de couverture des régimes pris en compte) et (ii) à la modalité d'enregistrement des ordonnances (télétransmission conditionnée par l'utilisation de la carte vitale *versus* envoi par courrier de la feuille de sécurité sociale). Le taux d'exhaustivité à retenir n'est cependant pas le taux d'exhaustivité *in fine* (atteint au bout d'environ deux mois) mais le taux d'exhaustivité au terme du délai maximal compatible avec la fonction d'alerte, soit une semaine. On verra en pratique qu'il s'agit de prendre en considération la partie télétransmise (taux de télétransmission des ordonnances = 0,9) des données relevant uniquement du régime général (taux de couverture moyen = 0,7). L'incidence hebdomadaire mesurable à une semaine s'élève à :  $(8\ 000/52) \times 0,7 \times 0,9$  soit environ 100 cas / 100 000 habitants (incidence retenue pour le calcul de puissance) ;

- le rapport "taille de la population exposée / taille de la population non exposée" est fixé à 1/10. La taille de la population non exposée n'est pas un facteur limitant car elle peut être représentée par le reste du département. Le résultat du calcul de puissance est en fait peu sensible au niveau choisi pour ce paramètre à l'intérieur de son domaine de vraisemblance (1/5 à 1/50).

### 3.3.2 | Fréquence de l'exposition critique

Les autorités sanitaires interprètent en premier lieu les résultats obtenus pour les deux indicateurs de contamination fécale (*Escherichia coli* et entérocoques) en terme de conformité réglementaire : "une eau est conforme aux limites de qualité microbiologique lorsqu'il y a absence d'*Escherichia coli* et d'entérocoques dans un échantillon de 100 mL". La présence de quelques

bactéries fécales peut cependant relever d'une pollution de l'échantillon lors du prélèvement ou d'un incident local (décrochement de biofilm) et indique mal la possibilité d'une épidémie. Il s'agit donc de prendre en considération un seuil très supérieur à zéro bactérie fécale par 100 mL pour que le dépassement de ce seuil indique un risque non négligeable d'épidémie.

Nous définissons l'**événement critique** comme la survenue d'un prélèvement réalisé au titre du contrôle réglementaire, soit en sortie d'usine, soit sur le réseau de distribution, et révélant une pollution fécale caractérisée par la somme des deux paramètres (*Escherichia coli* + entérocoques) supérieure ou égale à 50 UFC/100mL.

Cette définition repose sur l'hypothèse qu'au-delà du seuil "50", le déclenchement d'une épidémie d'origine hydrique n'est pas improbable. Ce choix reste arbitraire car non documenté au plan scientifique. La distribution des concentrations dans l'eau bue par les malades lors

d'épidémies est rarement connue, faute de prélèvements adéquats. Même si le seuil "50" ne correspond pas à un risque fort de survenue d'une épidémie, on peut augurer que la fréquence de son dépassement constitue une mesure relative acceptable du risque épidémique et

que le classement des départements selon ce critère correspond au classement des départements vis-à-vis du risque épidémique. La valeur de ce seuil pourra

faire l'objet d'une réévaluation en fonction des premiers résultats obtenus après la mise en place du système automatisé.

Par analogie avec la définition de l'événement critique, "**UDI critique**" désigne les installations qui ont présenté au moins un événement critique dans l'année.

Le terme "installation" englobe les UDI et des stations de traitement d'eau potable. Ces dernières représentent 15 % du total des installations ayant connu au moins un événement critique en 2003. Dans la suite du protocole, le terme impropre "UDI" sera utilisé indistinctement pour les stations de traitement et les UDI.

L'approche par UDI complète l'approche par événement dans le sens où elle en atténue deux défauts :

- 1) le biais lié à la différence de pression d'échantillonnage du contrôle réglementaire : l'espérance du nombre d'événements détectés est en effet proportionnelle à la fréquence de prélèvement tandis que la probabilité de détecter au moins un événement critique pour une installation donnée est moins que proportionnelle à cette fréquence (fonction dérivée décroissante). Le classement des départements par nombre d'UDI critiques est donc plus robuste que le classement par nombre d'événements critiques vis-à-vis des variations de fréquence d'échantillonnage ;
- 2) les pollutions chroniques : certaines UDI connaissent des pollutions fortes et chroniques. Il paraît

inconvenant de faire l'hypothèse qu'un événement critique survenant sur ce type d'UDI produise la même probabilité d'épidémie que sur une UDI habituellement exempte de pollution fécale ; d'une part, parce qu'une partie de la population ne s'expose pas (consommation d'eau embouteillée) et d'autre part, parce que les consommateurs résiduels ont acquis une immunité relative vis-à-vis du risque d'infection par certains pathogènes d'origine fécale (principalement les parasites des genres *Giardia* et *Cryptosporidium*, mais aussi les rotavirus et certaines souches d'*Escherichia coli*).

Les départements seront successivement classés par :

- nombre d'événements critiques ;
- nombre d'UDI critiques ;
- population desservie par ces installations.

Il sera finalement proposé un classement synthétique des départements reposant sur le rang moyen entre les deux derniers critères.

### 3.3.3 | Disponibilité des informations

L'étape de mise en place de l'outil automatisé au niveau départemental sera conditionnée par la mise à disposition des données sanitaires et facilitée par le géocodage préexistant des limites des unités de

distribution pour l'ensemble du département, sous un format compatible avec l'utilisation d'un Système d'information géographique (SIG).

# 4 Résultats

## 4.1 | Données sanitaires

### 4.1.1 | Choix de la source de données

Afin d'estimer l'aptitude de chacun des systèmes à répondre à l'objectif d'alerte, trois critères ont été définis et classés par ordre d'importance décroissant (C1 et C2 étant d'égale importance) :

- C1) Résolution spatiale. Dans la mesure où une corrélation sera recherchée entre l'emprise géographique d'un cluster de GEA et le contour d'une unité de distribution d'eau, la capacité du système à localiser chaque cas au plus proche de son lieu d'habitation (communes E et F de la figure 1) est indispensable ;
- C2) Réactivité du système. Elle peut se mesurer par le temps écoulé entre la délivrance des médicaments et la disponibilité de l'ordonnance dans la base de données. Elle conditionne la détection précoce de l'événement épidémique et influe ainsi sur le nombre de cas évités. Les modalités de transmission des informations déterminent cette réactivité ;
- C3) Sensibilité du système de surveillance. Elle dépend directement du taux d'exhaustivité intrinsèque au système de recueil des prescriptions médicamenteuses retenu. Ce taux d'exhaustivité est conditionné à la fois par le taux de couverture global (national) du système mais aussi de sa structure au niveau local. Par exemple, un système couvrant 50 % de la population n'apportera pas la même puissance selon qu'il s'agisse d'une commune sur deux ou de 50 % de la population de toutes les communes. Enfin, la nature de l'information peut différer (ventes totales > ventes prescrites > ventes remboursées), ce qui induit une différence de sensibilité et de puissance du système.

Les résultats de l'étude comparative menée sur les trois systèmes de recueil national des données de prescriptions médicamenteuses (AM, IMS-Health, Gers) sont résumés dans le tableau 2.

Au regard du premier critère, le système Erasme de l'AM permet de localiser les cas de GEA à l'adresse exacte de l'assuré alors que le système SNIIR-AM ne conserve de l'adresse que la commune de résidence. Les autres sources de données identifient dans le meilleur des cas l'officine qui a délivré l'ordonnance (cas d'IMS-Health). La base du Gers qui ne permet pas de localiser les cas ne peut supporter un système d'alerte.

En ce qui concerne la réactivité du système à mettre en place, les bases d'IMS-Health et de l'AM bénéficient du même protocole de télétransmission des données et ne se distinguent donc pas sur ce critère. En revanche, l'exhaustivité des deux systèmes diffère :

- sur la nature des ventes couvertes : uniquement les médicaments prescrits et remboursés pour l'AM contre l'ensemble des ventes pour IMS-Health, c'est-à-dire les ventes prescrites remboursées, non remboursées et les ventes "conseil"<sup>4</sup> ;
- sur l'unité géographique enquêtée : le système Erasme de l'AM contient l'adresse de résidence de l'assuré, l'unité géographique d'agrégation est donc libre et l'UDI peut être choisie ; pour SNIIR-AM, la commune s'impose. IMS-Health agrège l'information à l'échelle d'un regroupement d'officines qui représente au minimum 25 000 habitants.

En conclusion, l'évaluation de l'aptitude des sources de données des prescriptions médicamenteuses existantes à constituer un système de surveillance dédié à l'alerte nous conduit à identifier le système de l'AM comme l'outil de choix pour répondre à cet objectif. Cette position s'explique principalement par sa capacité à localiser les cas à leur domicile et, dans une moindre mesure, à offrir un taux de couverture satisfaisant sur l'ensemble du territoire.

<sup>4</sup> Ventes conseil : ventes directes hors prescriptions médicales.

Tableau 2 - Description succincte des différentes bases de données des prescriptions médicamenteuses

Source	Objectifs	Modalités	Information recueillie	Localisation des cas	Bilan critères	Limites
<b>Producteur :</b> IMS-Health <b>Nom de la base :</b> XPONENT	Toute prestation commerciale basée sur le suivi de la délivrance des médicaments	<b>Données :</b> tickets de ventes dans les pharmacies. <b>Émetteur :</b> pharmaciens. <b>Taux de couverture :</b> réseau d'environ 10 400 officines (45 % du nombre total d'offices, soit 1 officine "IMS-Health" / 6 000 habitants). <b>Fréquence de transmission :</b> en moyenne 1,5 fois par semaine, coordonnée avec la télétransmission des données de l'AM (1/3 quotidien à hebdomadaire, 1/3 hebdomadaire à mensuel, 1/3 mensuel ou plus).	Code pharmacie Âge du bénéficiaire Régime de sécurité sociale Code CIP Nombre d'unités dispensées Date d'achat Prescripteur	Au niveau de la pharmacie	<b>C1 : +</b> <b>C2 : +++</b> <b>C3 : +</b>	<b>Qualité des données :</b> adresse du patient et date de prescription non renseignées.  <b>Unité géographique :</b> regroupement de 4 officines "IMS-Health", soit environ 25 000 habitants.  <b>Tarifs :</b> environ 1 000 euros par département pour un an, 88 000 euros France entière (estimation 2004).
<b>Producteur :</b> Gers Norm de la base : Gers	Collecte les données de ventes directes et indirectes auprès des laboratoires pharmaceutiques et de grossistes répartiteurs adhérents du Gers, afin de diffuser aux adhérents des études statistiques	<b>Données :</b> données de livraison avec prise en compte du retour marchandises. <b>Émetteur :</b> grossistes-répartiteurs. <b>Taux de couverture :</b> 99,9 % des données des grossistes répartiteurs en France. <b>Fréquence de transmission :</b> quotidienne ou mensuelle.	Code CIP	NON	<b>C1 : -</b> <b>C2 : +</b> <b>C3 : +++</b>	<b>Qualité des données :</b> aucun renseignement sur les patients.  <b>Unité géographique :</b> grossistes répartiteurs (3 à 10 officines), soit environ 7 500 à 25 000 habitants.  <b>Tarifs :</b> ?
<b>Producteur :</b> Assurance Maladie <b>Nom de la base :</b> Erasmé	Liquidation des actes de soins remboursés	<b>Données :</b> prescriptions médicales faisant l'objet d'un remboursement couvert par le régime général. <b>Émetteur :</b> pharmaciens. <b>Taux de couverture : Régime général</b> De 50 % (zone rurale) à 95 % (zone urbaine). <b>Fréquence de télétransmission par les pharmaciens :</b> en moyenne 1,5 fois par semaine (1/3 quotidien à hebdomadaire) en 2003 ; fin 2005, 90 % de l'info serait colligée à 2 jours ouvrés.	Données nominatives de l'assuré + âge, sexe, bénéficiaire Code CIP Nombre d'unités	<b>Adresse complète de l'assuré</b> plus de 80 % des cas (objectif 100 % courant 2006) : bénéficiaire, adresse prescripteur et pharmacie	<b>C1 : +++</b> <b>C2 : +++</b> <b>C3 : ++</b>	<b>Qualité des données :</b> nécessité d'avoir au moins un médicament remboursé. Uniquement Régime général : taux de couverture 50 % à 90 %  <b>Accessibilité :</b> ?
<b>Producteur :</b> Assurance Maladie <b>Nom de la base :</b> SNIIR-AM	Copie anonymisée d'Erasmé à vocation statistique	<b>Données :</b> Idem Erasmé. <b>Émetteur :</b> Erasmé <b>Taux de couverture : tous les régimes</b> Intégration en cours (couverture nombre d'unités fin 2005 : 98 %)	Données anonymisées assuré + âge, sexe, bénéficiaire Code CIP Nombre d'unités	<b>Code commune</b> Insee, code postal : bénéficiaire <sup>5</sup> , prescripteur et pharmacie	<b>C1 : ++</b> <b>C2 : +++</b> <b>C : ++</b>	<b>Qualité des données :</b> uniquement régime général disponible pour l'alerte. Résolution spatiale limitée à la commune.  <b>Accessibilité :</b> +

<sup>5</sup> Information non disponible au 01/04/06

### 4.1.2 | Bases de données de l'AM

L'AM dispose de deux bases de données susceptibles d'être utilisées :

- Erasme (Extraction recherches analyses pour un suivi médico-économique) qui couvre le régime général, sert à la liquidation des actes de santé dispensés en médecine de ville et remboursés ;
- SNIIR-AM (Système national d'information inter-régimes de l'AM) qui couvre l'ensemble des régimes de l'AM a une vocation statistique.

Du fait de l'audience de la Mutualité sociale agricole (MSA) en secteur rural, Erasme ne couvre que 50 % des régimes en campagne, contre 95 % en zone urbaine.

Le taux de télétransmission des ordonnances vers Erasme, c'est-à-dire la proportion des règlements qui se font avec la carte vitale, est estimée par l'AM à plus de 90 %. Les 10 % restant, qui correspondent aux feuilles de soins transmises par courrier, ne peuvent être pris en compte dans le système d'alerte du fait d'un temps de traitement de l'ordre d'un à deux mois.

La fréquence de télétransmission des officines vers Erasme varie en fonction des officines et en fonction de la base considérée. Ainsi, au niveau national en 2003, un tiers des officines télétransmettaient "tous les jours à une fois par semaine", un tiers "une fois par semaine à une fois par mois" et un tiers "une fois par mois". Cette fréquence a considérablement évolué en 2004 et 2005 : des tests récents ont montré que 90 % des données du Régime général sont disponibles à 48 heures ouvrables (Merlière, communication personnelle).

La base SNIIR-AM est mise à jour quotidiennement à partir d'Erasme. Cela ne concerne que les données du Régime général. L'intégration des données provenant des autres régimes est mensuelle.

Par rapport à l'objectif d'alerte, les bases SNIIR-AM et Erasme présentent donc la même sensibilité<sup>6</sup> et une réactivité proche. La supériorité d'Erasme sur SNIIR-AM se situe au niveau de la localisation de l'assuré (adresse vs. code postal / commune de résidence<sup>5</sup>).

L'existence d'épidémies de GEA d'origine hydrique, affectant principalement les populations de passage et épargnant les résidents, montre qu'il existe des différences sensibles de susceptibilité à certains pathogènes, comme les parasites. Une immunité pourrait être acquise chez les résidents par la consommation régulière d'eau du robinet contaminée.

Cela renvoie à la question de la participation des populations de passage au système d'alerte. Il est possible de repérer la population en transit en conditionnant l'extraction des données de la base SNIIR-AM sur la localisation du médecin qui a prescrit ou du pharmacien qui a délivré les médicaments, plutôt que sur celle de l'assuré, et de récupérer ainsi les cas non résidents. Cependant, le recouvrement des frais d'achat des médicaments par les personnes de passage se fait généralement *a posteriori* par l'envoi de la feuille de soins. Ces données ne sont pas saisies rapidement dans la base et les populations transitoires ne peuvent donc en pratique participer à l'alerte<sup>7</sup>.

Un logiciel a été développé par l'InVS et est utilisé depuis 2003 pour exploiter les données de l'AM à des fins de surveillance épidémiologique des GEA (Beaudeau *et al.*, soumis à la Revue médicale de l'AM). Il implémente notamment une fonction de discrimination des cas de GEA médicalisés basée sur l'analyse des données de remboursement des médicaments. L'indicateur d'incidence des GEA médicalisées qui en résulte présente une sensibilité de 88 % et une spécificité de 86 %.

<sup>5</sup> Information non disponible au 01/04/06.

<sup>6</sup> Compte tenu de la contrainte de réactivité, l'exploitation de SNIIR-AM se limiterait au flux de données provenant du Régime général.

<sup>7</sup> Une utilisation *a posteriori* de ces données permettrait d'estimer l'impact de l'épidémie parmi la population de passage. L'absence d'information démographique sur cette population au niveau communal empêche cependant la comparaison des taux d'incidence entre les deux populations.

## 4.2 | Taille des unités de distribution : choix de la catégorie cible

Ce paragraphe définit les limites de taille des UDI à prendre en considération dans le choix des départements.

Toutes les UDI sont *a priori* susceptibles de voir leurs eaux polluées et de provoquer une épidémie. Cependant, en deçà d'une certaine taille de population desservie, le système d'alerte pressenti ne possède plus la puissance statistique suffisante pour la détection d'une épidémie touchant cette population.

Le scénario épidémique retenu est celui d'une exposition touchant la totalité du réseau de distribution et d'une épidémie durant une semaine. Les résultats du calcul de puissance sont présentés dans la figure 2 pour des RR de 1,4 à 3 correspondant à une situation hyperendémique et dans la figure 3 pour des RR de 3 à 20 (détail des valeurs en annexe 1).

Sur chaque graphique, deux courbes sont tracées :

- "n1 Epi-info" qui correspond à la taille de la population exposée estimée à partir d'Epi-info. L'exposition comprend les expositions occultes (contact, ingestion fortuite pendant le brossage des dents ou *via* les aliments lavés...) à prendre en considération en cas de pollution massive de l'eau, obtenues par exemple par des retours d'eau usée dans le réseau d'eau potable ;
- "n1 Epi-info\_corrigeé\_%NC" qui correspond à "n1 Epi-info" corrigée du pourcentage de non-consommateurs d'eau du robinet, estimé à 30 % en France (valeur moyenne) [4]. Le scénario d'exposition se réduit alors à la consommation d'eau non chauffée pour la boisson et s'adresse à des contaminations plus ordinaires. C'est celui-ci qui est pris en considération par la suite.

Figure 2 - Taille de la population exposée en fonction du risque relatif (1,4-3)

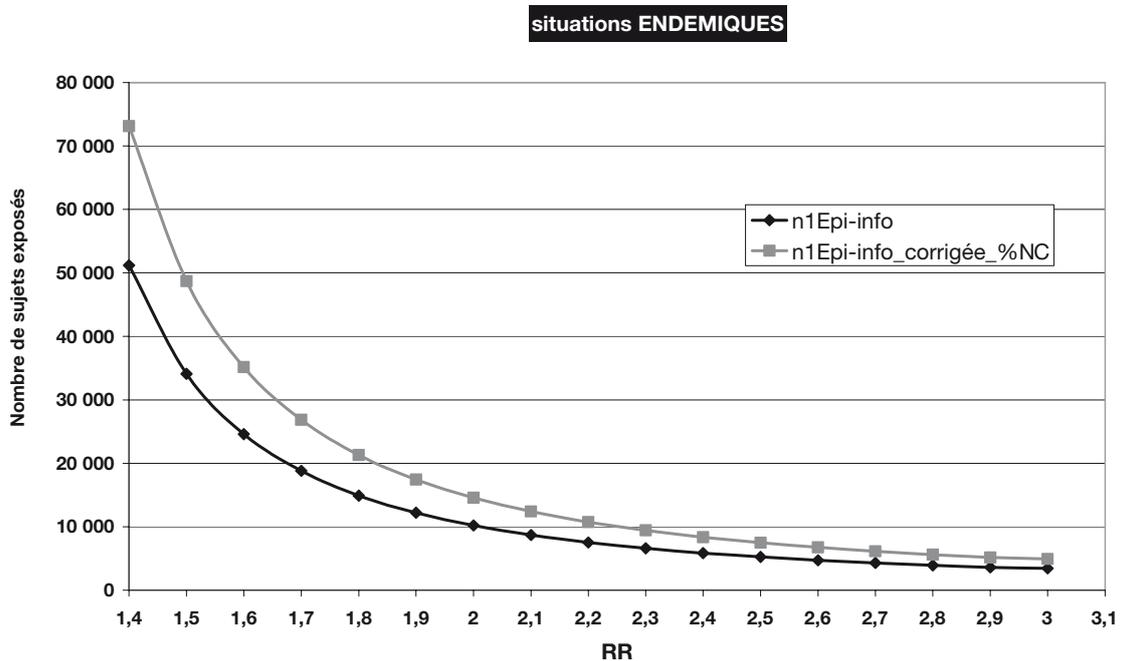
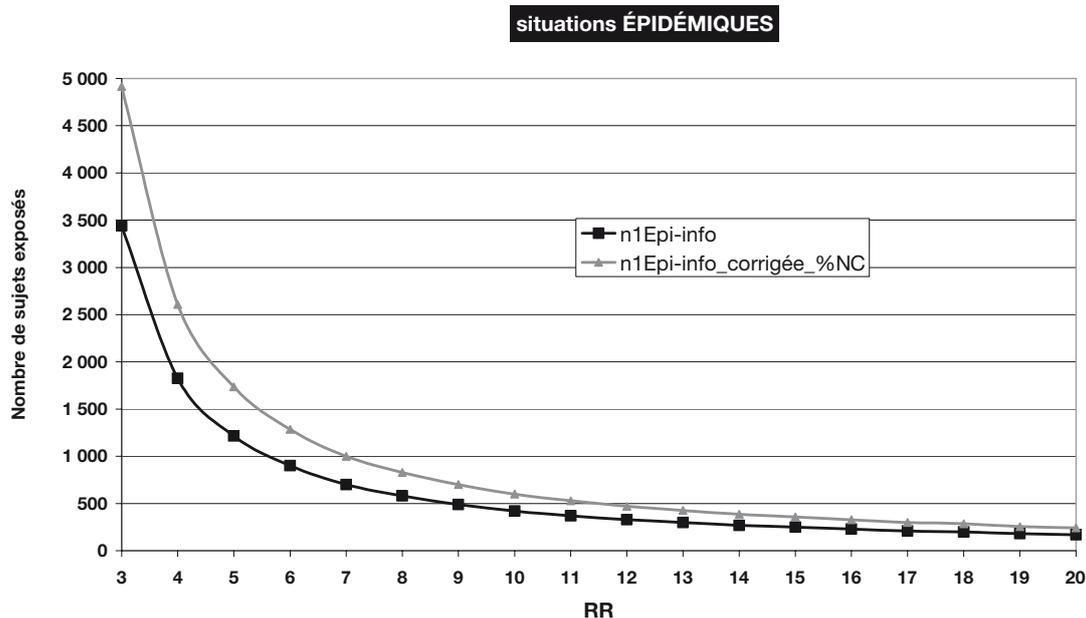


Figure 3 - Taille de la population exposée en fonction du risque relatif (3-20)



Pour mettre en évidence un RR compris entre 2 et 10 avec une probabilité supérieure à  $p = 0,8$  (puissance recherchée), les tailles des populations exposées correspondent respectivement à environ 10 000 et 500 personnes. Les épidémies les plus sévères (RR = 20) peuvent être détectées sur des UDI de très faible taille (200 personnes desservies).

Toutes les UDI ne sont pas égales en probabilité vis-à-vis de la survenue d'une épidémie. Les épidémies provoquées par la pollution de réseau desservant plus de 10 000 habitants sont rares, ainsi que les pollutions microbiologiques importantes (cf. paragraphe 4.3.1). Ces événements exceptionnels, non reproductibles d'une année sur l'autre et non représentatifs de la

vulnérabilité de l'approvisionnement en eau du département, ne doivent pas intervenir dans le choix des départements pilotes. De plus, ils sont facilement repérables par de nombreux intervenants (professionnels de santé, enseignants...) et portés à la connaissance des autorités sanitaires avec une bonne probabilité.

Pour le classement des départements, la catégorie cible pour la taille de l'UDI est donc fixée à [500 ; 10 000]. Le système d'alerte pressenti serait cependant apte à détecter des clusters concernant des réseaux de grande taille. Le domaine d'utilisation du système automatisé couvrirait au total seulement 39 % des UDI mais 96 % de la population française (tableau 3).

Tableau 3 - Distribution du nombre d'UDI et de la population selon la taille - situation 2004 [1]

Classes d'UDI (en nombre d'habitants)	Nombre d'UDI	% UDI	Population desservie	% pop
0-499	16 348	61 %	2 198 000	4 %
500-9999	9 510	35 %	22 363 000	37 %
>10 000	987	4 %	36 144 000	60 %
Total	26 845	100 %	60 705 000	100 %

### 4.3 | Classement des départements

#### Avertissement

Compte tenu des limites liées à la disponibilité des données et à la puissance statistique, sont exclus du classement :

- les départements pour lesquels l'information était absente ou incomplète dans la base SISE-eaux au moment de l'extraction faite sur l'année 2003

(Haute-Corse, Moselle, Lozère, Saint-Pierre-et-Miquelon, Mayotte, Var) ;

- les départements ne présentant pas d'installation (station de traitement ou UDI) dans la classe de taille [500-10 000[ (cas des Hauts-de-Seine).

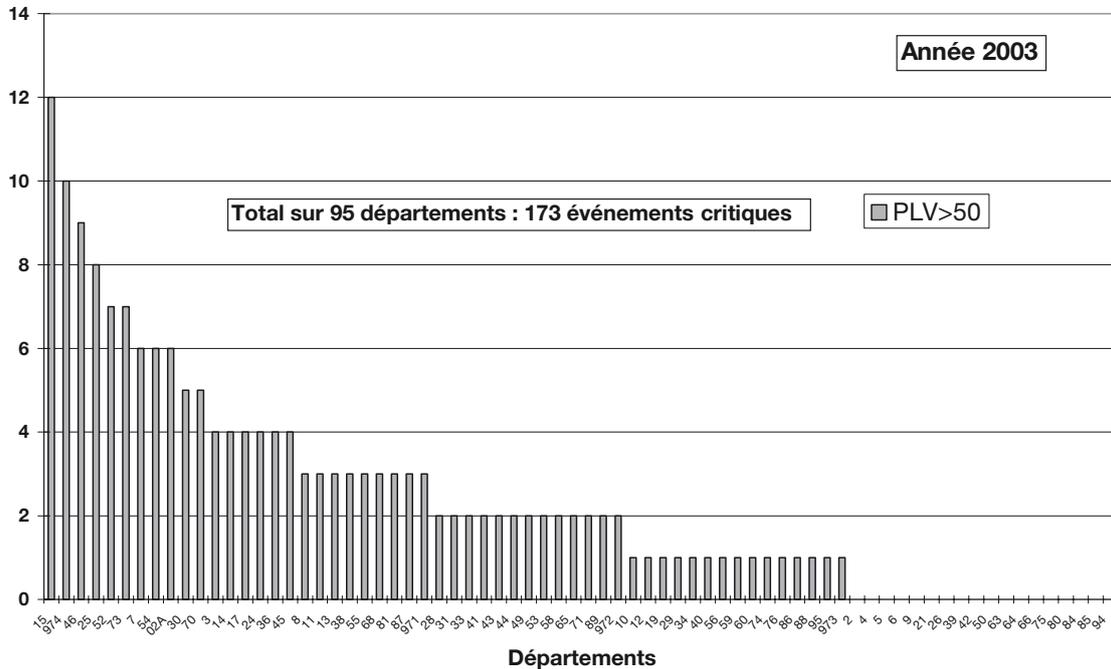
Le classement des départements concerne finalement 95 départements.

#### 4.3.1 | Classement par nombre d'événements critiques

En 2003, 173 événements critiques ont été détectés dans les départements de France métropolitaine et d'outre-mer pour la classe [500-9 999] habitants (figure 4), soit en moyenne près de deux événements critiques par département et par an. Cinquante-quatre départements

ont enregistré au moins un événement critique en 2003. Deux départements (le Cantal et la Réunion) ont subi au moins 10 dépassements du seuil critique, neuf autres départements ont connu entre 5 et 9 dépassements.

Figure 4 - Distribution par département du nombre d'événements critiques (nombre de prélèvements > 50 UFC/100mL en *Escherichia coli* + entérocoques) pour les UDI de 500 à 9 999 habitants



La distribution des événements critiques par type d'installation (UDI : 67 % ; sortie usine : 33 %) ne correspond pas à la répartition de l'échantillonnage défini par le contrôle sanitaire (tableau 1) (UDI : 78 % ; sortie usine : 22 %). La différence pourrait s'expliquer

par la combinaison d'une désinfection incomplète en sortie d'usine (temps de contact réduit) et plus achevée en distribution, avec une éventuelle rechloration sur réseau.

### 4.3.2 | Classement par nombre d'unités de distribution critiques

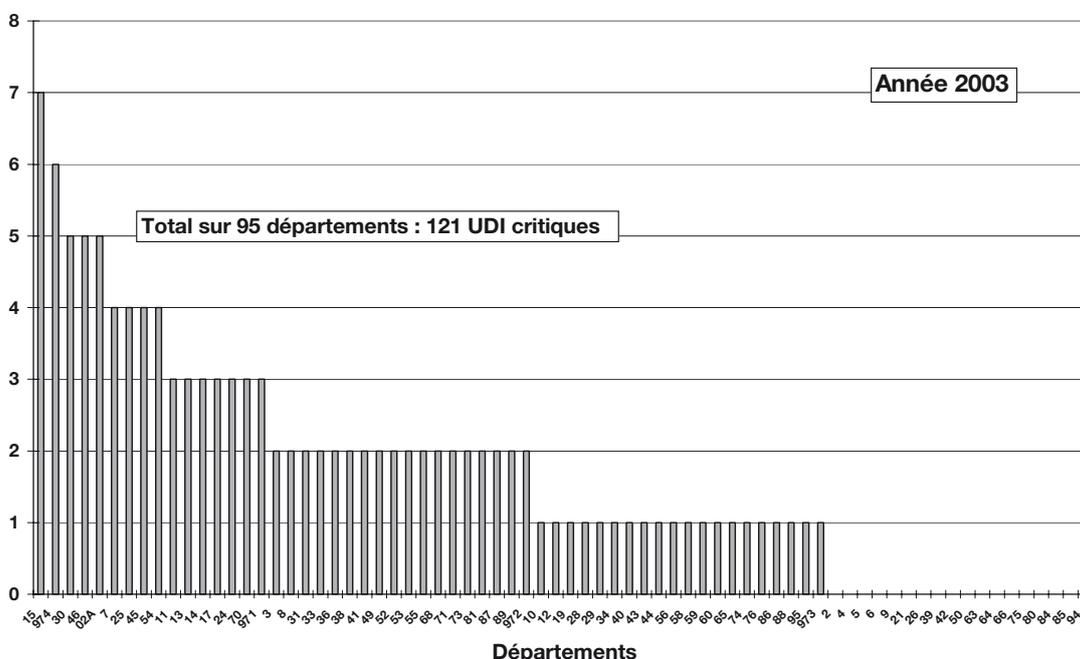
La figure 5 présente le classement des départements par ordre décroissant du nombre d'UDI critiques.

Dans cette approche, chaque UDI ayant présenté au moins un événement critique dans l'année est comptabilisée comme une UDI critique. Pour l'ensemble des départements, les résultats d'analyses identifient 121 UDI critiques pour la classe [500-9 999] habitants (1,3 % des UDI de cette taille), desservant 262 317

personnes exposées (1,2 % de la population desservie par les UDI de cette taille).

La prise en compte des UDI critiques desservant moins de 500 habitants (non envisageable par manque de puissance statistique) élèverait de 510 le nombre d'UDI critiques (3,1 % des UDI de cette taille) et de 54 192 le nombre de personnes exposées (2,4 % de la population desservie par les UDI de cette taille).

**Figure 5 - Distribution par département du nombre d'UDI critiques (nombre d'installations ayant eu au moins 1 prélèvement > 50 UFC/100mL en *Escherichia coli* + entérocoques) pour les UDI de 500 à 9 999 habitants**



Les résultats de ces deux approches (par événement ou par UDI) conduisent sensiblement au même ordre des départements pour la tête du classement (80 % des dix premiers départements sont inchangés : Cantal, Réunion, Lot, Doubs, Ardèche, Meurthe-et-Moselle,

Corse-du-Sud, Gard). Les différences qui peuvent apparaître au sein des départements s'expliquent par la présence de pollutions répétées, voire chroniques, de certains réseaux d'eau.

### 4.3.3 | Classement par taille de la population exposée

La figure 6 décrit la distribution des populations desservies par les UDI critiques par classes de taille d'UDI. Cette répartition montre notamment que près de 80 % des UDI critiques sont inférieures à 500 habitants et totalisent 3,5 % de la population exposée ; à l'inverse, 2,3 % des UDI critiques sont supérieures à 10 000 habitants et totalisent 80 % de la population exposée.

L'ensemble de la population à risque d'après les critères définis plus haut s'élèverait à près de 1,5 millions de personnes chaque année (tableau 4). Il est à noter que seuls trois événements critiques ont été détectés sur des UDI desservant plus de 100 000 personnes, dont un sur une UDI supérieure à 500 000 habitants.

Figure 6 - Répartition des UDI critiques et de la population exposée en fonction de la taille des UDI

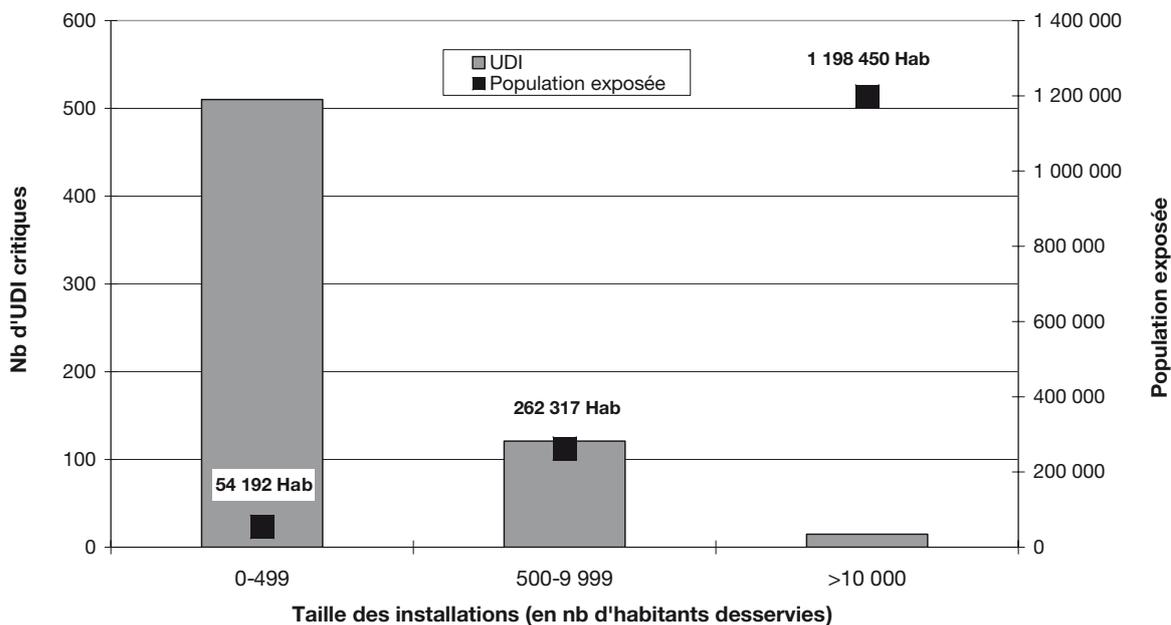


Tableau 4 - Nombre d'UDI critiques et de la population exposée en fonction de la taille des UDI

Classes d'UDI (en nombre d'habitants)	<100-499	500-9 999	>10 000	Total
Nb de PLV>1 UFC/100mL	1 938	1 227	2 673	5 838
Nb d'événements critiques	783	173	17	973
Nb d'UDI critiques	510	121	15	646
Population exposée	54 192	262 317	1 198 450	1 514 959

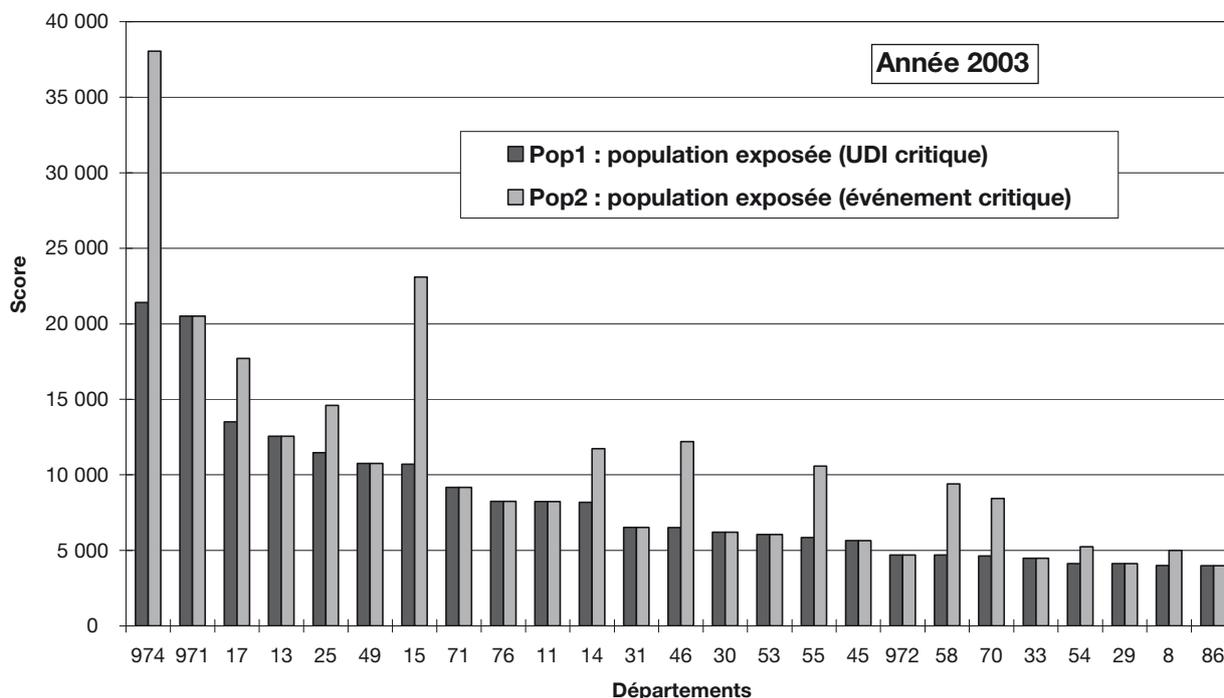
Dans la logique des approches par événement et par UDI critique, deux critères de population exposée sont utilisés pour classer les départements :

- **Pop1** : population desservie par des UDI critiques de 500 à 9 999 habitants ;
- **Pop2** : population exposée à un événement critique donné touchant les UDI de 500 à 9 999 habitants.

Dans ce cas, la population d'une UDI critique sera comptabilisée autant de fois qu'il y a eu d'événements critiques sur cette UDI.

Les départements sont ensuite triés selon Pop1 par ordre décroissant et les résultats sont présentés sur la figure 7 pour les 25 premiers départements (détails des valeurs pour l'ensemble des départements en annexe 2).

Figure 7 - Classement des 25 départements présentant le plus fort risque épidémique en considérant les installations desservant 500 à 9 999 habitants



En tête du classement selon Pop1, se positionnent les départements qui totalisent le plus grand nombre de personnes exposées et desservies par des UDI critiques de 500 à 9 999 habitants. L'ordre des départements s'explique soit par un nombre important de petites UDI critiques (2 500 habitants desservis) : cas du Cantal (15), du Doubs (25), de l'Aude (11), du Calvados (14); soit par un nombre moins important de grosses UDI (7 000 habitants desservis) : cas de la Seine-Maritime (76), du Maine-et-Loire (49) et de la Guadeloupe (971).

La différence des classements obtenus selon Pop1 et Pop2 traduit la répétition d'événements critiques sur certaines UDI au cours de l'année 2003 : c'est le cas du Cantal (15), du Calvados (14), du Lot (46), de la Meuse (55).

Les classements obtenus, d'une part sur la base du nombre d'UDI critiques (figure 5), d'autre part sur la base de la population exposée (pop1 de la figure 7), diffèrent peu. Parmi les dissonances observées, certains départements se positionnent en tête sur le nombre d'UDI critiques, alors qu'ils figurent bien plus loin dans le classement lorsque le critère "population exposée" est pris en compte : c'est le cas de la Corse-du-Sud (02A) qui passe de la 5<sup>ème</sup> à la 26<sup>ème</sup> place, et inversement : la Seine-Maritime (76) passe de la 50<sup>ème</sup> à la 9<sup>ème</sup> place. Ces dissonances s'expliquent par des différences importantes dans la distribution par taille des UDI critiques : dans le cas (02A), les UDI critiques sont toutes inférieures à 1 000 habitants alors que dans le deuxième cas (76), les UDI critiques sont toutes supérieures à 5 000 habitants.

### 4.3.4 | Classement synthétique

Considérer le nombre d'UDI critiques conduit à un classement des départements moins biaisé par l'hétérogénéité de l'effort d'échantillonnage que celui obtenu sur le critère du nombre d'événements critiques. Étant donné la disparité des tailles des UDI impliquées, il apparaît pertinent en terme de santé publique d'ajouter au critère du nombre d'UDI critiques celui de la taille de la

population exposée (Pop1). Un classement synthétique des départements a donc été établi en fonction du nombre d'UDI critiques et de la population qui y est rattachée (rang moyen). Le tableau 5 propose une liste pouvant servir de base pour le choix des départements pilotes.

**Tableau 5 - Classement par ordre d'intérêt décroissant des départements pilotes potentiels**

Département	Classement par nb. d'UDI critiques (Fig. 5)	Classement par pop. exposée (Pop1, Fig. 7)	Classement synthétique	Population (Insee, 1999)	Pop. exposée/pop. Insee
Réunion, 974	2	1	<b>1</b>	706 300	3,03 %
Cantal, 15	1	7	<b>2</b>	150 778	7,10 %
Doubs, 25	6	5	<b>3</b>	499 062	2,30 %
Guadeloupe, 971	10	2	<b>4</b>	422 496	4,85 %
Charente-Maritime, 17	10	3	<b>5</b>	557 024	2,43 %
Bouches-du-Rhône, 13	10	4	<b>6</b>	1 835 719	0,68 %
Lot, 46	3	13	<b>7</b>	160 197	4,06 %
Gard, 30	3	14	<b>8</b>	623 125	0,99 %
Aude, 11	10	10	<b>9</b>	309 770	2,66 %
Calvados, 14	10	11	<b>10</b>	648 385	1,26 %

### 4.3.5 | Existe-t-il un effet “département” ?

Le classement des départements par nombre d'événements critiques dans la classe d'UDI [500-10 000] pourrait simplement refléter l'effort de surveillance fait pour cette classe de taille en 2003. Au-delà de cet effet de taille, existe-t-il un effet “département” qui engloberait la vulnérabilité des ressources, la qualité des traitements et des réseaux et, sur un autre plan, l'effort d'investissement et d'exploitation, notamment la veille technique mise en place en vue de prévenir le risque de pollution ?

Si et seulement si le phénomène est purement aléatoire, le nombre d'événements critiques (on sous-entendra désormais parmi les UDI de taille [500-10 000]) dans le département  $i$  suit une loi de Poisson  $n_i \approx P(N_i, p)$  où  $N_i$  est le nombre d'analyses réalisées dans le département  $i$  sur la classe d'UDI d'intérêt et  $p$  est la probabilité qu'un prélèvement soit critique.

$p$  est estimé par la fréquence des événements critiques au niveau national :

$$\hat{p} = \frac{n}{N} (= 173/93854 = 0,0018).$$

Pour les départements en tête de classement, on a calculé la probabilité d'égaliser ou de dépasser le nombre d'événements critiques observés ( $k$ ). On a ensuite comparé cette probabilité avec la position du département dans le classement par nombre d'événements critiques au niveau national (tableau 6), classement prenant en compte les départements qui ont alimenté la base SISE-eaux en 2003 et pour lesquels au moins un prélèvement a été réalisé sur la classe d'UDI [500-10 000], soit 95 départements.

**Tableau 6 - Comparaison de la probabilité d'observer au moins k événements critiques au niveau départemental et au niveau national**

Départements	Classement synthétique (tableau 5)	Classement par nb. d'événements critiques	k	a p(ni≥k)	b	b/a
<b>Réunion, 974</b>	<b>1</b>	<b>2</b>	<b>10</b>	<b>1,6E-05</b>	<b>0,02</b>	<b>1,3E+03</b>
<b>Cantal, 15</b>	<b>2</b>	<b>1</b>	<b>12</b>	<b>3,5E-10</b>	<b>0,01</b>	<b>3,0E+07</b>
<b>Doubs, 25</b>	<b>3</b>	<b>4</b>	<b>8</b>	<b>1,3E-05</b>	<b>0,04</b>	<b>3,2E+03</b>
Guadeloupe, 971	4	18	3	1,4E-01	0,19	1,4E+00
Charente-Maritime, 17	5	12	4	3,8E-01	0,13	3,3E-01
Bouches-du-Rhône, 13	6	18	3	3,9E-01	0,19	4,8E-01
<b>Lot, 46</b>	<b>7</b>	<b>3</b>	<b>9</b>	<b>3,3E-06</b>	<b>0,03</b>	<b>9,5E+03</b>
Gard, 30	8	10	5	2,9E-01	0,10	3,6E-01
Aude, 11	9	18	3	3,4E-01	0,19	5,5E-01
Calvados, 14	10	12	4	6,3E-01	0,13	2,0E-01

b : Proportion observée de départements de rang égal ou inférieur dans le classement par nombre d'événements critiques (cf. paragraphe 4.3.1).

Pour 4 départements, la probabilité d'obtenir le nombre d'événements critiques observé est très faible, même en tenant compte de la sélection des départements opérée sur la base du rang dans le classement. Le Cantal, le Lot, le Doubs et la Réunion se détachent de l'ensemble des autres départements, dont le classement est en revanche compatible avec l'hypothèse d'équiprobabilité d'occurrence d'événements critiques.

L'effet "département", très sélectif, correspondrait soit à un artefact, s'il découle d'un effort d'échantillonnage

concentré sur certains secteurs ou situations à risque, soit à une probabilité plus forte qu'ailleurs d'occurrence de pollutions susceptibles d'engendrer l'éclosion de cas groupés observables, au sens des hypothèses de travail et après contrôle de l'effet de taille.

Dans la deuxième hypothèse (à consolider sur des données SISE-eaux issues d'une période de plusieurs années et avec les partenaires locaux), ces départements présenteraient un intérêt particulier en termes de facteurs de risques accessibles à l'étude épidémiologique.

### 4.3.6 | Critères pour le choix des départements pilotes

Pour des raisons d'organisation administrative, l'unité éligible comme secteur pilote est le département. Plusieurs départements pourraient être retenus. Le critère synthétique de classement prend en considération le nombre d'UDI critiques, préféré au nombre d'événements pour des raisons de robustesse, et la population exposée. Ce critère s'accorde de l'objectif scientifique de l'étude pilote projetée, qui est de maximiser les chances de trouver "matière" pour la mise au point de la chaîne de traitement de l'information.

D'autres critères peuvent ou doivent cependant intervenir.

#### 1. Critères scientifiques ou techniques :

- a) présence probable de facteurs de risque particuliers (tableau 6, paragraphe "Existe-t-il un effet département ?"). C'est un critère important à

terme pour l'évaluation et la gestion des facteurs de risque, mais ce n'est pas l'objet principal de l'étude pilote,

- b) présence d'objectifs scientifiques particuliers (par exemple suivi prospectif de l'impact d'inondations dans le Gard) ou existence d'une expertise locale permettant l'étude de certains facteurs de risque (par exemple, travaux de l'Université de Rouen sur les transferts solides et microbiens dans le karst en Haute-Normandie). Là encore, ces objectifs restent secondaires dans la phase pilote,
- c) proportion de buveurs d'eau du robinet. Ce facteur conditionne l'exposition autant que la contamination de l'eau (exposition = contamination x consommation) mais n'a pas été évoqué jusque-là. La proportion de non-consommateurs est connue au niveau

régional et seulement en métropole [3]. Les départements de la moitié sud de la France sont *a priori* mieux placés (19 % de non-consommateurs) que les départements de la moitié nord (33 % ou plus de non-consommateurs). Il n'y a pas de données disponibles pour les départements d'outre-mer. Il faut cependant se méfier d'effets locaux de désaffection liés à la survenue de pollutions, de consignes de restriction temporaire de l'usage, voire d'épidémies notoires ;

## 2. Critères pratiques :

- a) importance de la taille de la population du département (tableau 5) ou de la fraction de la population exposée (les deux critères sont très liés). La taille du département conditionne le volume de données à extraire des bases de données de la CnamTS, puis à manipuler,
- b) préexistence d'un Système d'information géographique sur l'eau distribuée et comprenant notamment le contour des UDI. L'argument n'est

pas déterminant car les moyens à engager pour la constitution du fond de carte et son codage restent modérés,

- c) la proximité. L'articulation entre les travaux statistiques et les investigations de terrains serait plus difficile pour les départements d'outre-mer et nécessiterait une grande autonomie des services locaux,
- d) il peut exister outre-mer des spécificités locales ou tropicales de prescription des médicaments qui nécessiteraient une adaptation de l'algorithme dédié à l'identification des cas de GEA médicalisés. Un surcroît de travail serait à prévoir dans le traitement de l'information,
- e) intérêt et engagement des partenaires locaux (Ddass, Cire) sur le sujet. Ce critère est déterminant.

La liste proposée par ordre de priorité à l'AM se fera donc en concertation avec les partenaires locaux. Des indications sur l'effort à produire localement sont données au paragraphe "Perspectives opérationnelles".

# 5 Discussion

## 5.1 | Les épidémies passées au regard du système de détection automatisé

Six épidémies de GEA ont été investiguées par les Cire ou l'InVS entre 2000 et 2005 (tableau 7). Nous confrontons ici les caractéristiques de ces épidémies

(impact, délai de signalement) avec les potentialités du système automatisé de détection précoce esquissé en partie "Résultat".

**Tableau 7 - Bilan des épidémies infectieuses d'origine hydrique investiguées par l'InVS et les Cire, période 2000-2005 (rapports d'investigation disponibles sur le site InVS [www.invs.sante.fr](http://www.invs.sante.fr) ou auprès de l'InVS)**

Épidémie	Date	Point d'intro. pollution	Circonstances / causes	Taille pop. expos.	Investig. épidémiol.	Nb cas identifiés (consult. MG)	Nb cas cliniques estimé	TA	Déf. RR	RR*	Date pic épid.**	Date alerte**
Strasbourg (67)	mai 2000	réseau	travaux sur le réseau ; erreur procédure remise en eau	60 000	cas témoin + ED	53	?	faible	pendant / avant	2	+0 jour	+0 jour
Gourdon (46)	août 2000	forage	infiltration au forage d'eau de rivière polluée + dysfonction. chloration	4 900 + 2 200 touristes	cohorte rétrospect + ED	1 037	2 600	37 %	buv / non buv pendant / avant	(3,2) 10	+10 jours	+11 jours
Dracy (21)	sept 2001	réseau	interconnexion entre STEP et réseau AEP + retour d'eau	1 100	cohorte rétrospect + ED	364	560	51 %	buv / non buv pendant / avant	(2,0) 24	+5 jours	+5 jours
Isère (38)	nov 2002	captage	Précipitations => débordement STEP amont captage + inond. captage	5 600	cohorte rétrospect + ED	98	2 000	41 %	sect+ / sect- : buv / non buv :	20 (4,6)	+1 jour	+2 jours
Divonne (01)	août 2003	réseau	interconnexion entre STEP et réseau AEP + retour d'eau	2 500	ED	387	1 040	42 %	pendant / avant	4	+7 jours	+9 jours
Pont de Roide (25)	fév 2004	Réseau industriel alimenté par rivière	interconnexion entre réseau industriel et réseau AEP + retour d'eau	4 800	ED	444	1 000	20 %	obs / attendu (rés. sentinelle)	36	+1 jour	+1 jour

\* entre parenthèses : RR incomparable avec celui employé ici pour le calcul de puissance.

\*\* datation relative par rapport au début de l'épidémie (j0).

STEP : station d'épuration urbaine.

TA : taux d'attaque.

RR : risque relatif.

ED : enquête descriptive.

MG : médecin généraliste.

### 5.1.1 | Probabilité de détection

#### Note méthodologique : comparaison des TA et RR

Le maniement des RR et des TA proposés par les auteurs d'investigation pose problème du fait de l'hétérogénéité des modalités de calculs.

**Délimitation spatio-temporelle de l'épidémie.** Les valeurs des TA et des RR dépendent en premier lieu de la délimitation spatio-temporelle de l'épidémie. La délimitation géographique des épidémies n'est triviale que si la totalité du réseau est contaminée, ce qui correspond à une pollution de la ressource ou à une défaillance du traitement (2 cas sur 6 épidémies). Dans le cas d'une pollution introduite au niveau du réseau de distribution (4/6), le secteur exposé n'est pas connu précisément, notamment si la circulation de l'eau dans le réseau varie au cours d'une même journée (cas des réseaux dits en "distribution-refoulement"). Les limites temporelles de l'épidémie se définissent par comparaison du taux d'incidence pendant l'épidémie et du niveau du bruit de fond en situation non épidémique. On peut par exemple considérer arbitrairement que le seuil épidémique est 3 fois supérieur au niveau endémique. En l'absence d'une règle conventionnelle, il existe une marge d'appréciation sur les limites temporelles de l'épidémie ; elle est généralement réduite en ce qui concerne le début (augmentation abrupte), mais elle est plus importante en ce qui concerne la fin de l'épidémie (effet de traîne dû aux cas tardifs ou secondaires...).

**Dénominateur du RR.** Une fois le lieu et l'espace de l'épidémie établis, l'estimation du RR nécessite encore de définir le dénominateur (risque chez les non exposés). Quatre options co-existent parmi les six investigations réalisées : (i) risque observé chez les personnes déclarant ne pas consommer l'eau du robinet (études de cohorte), (ii) risque observé dans un secteur témoin pendant l'épidémie, (iii) risque observé avant l'épidémie sur le secteur touché ou encore (iiii) risque attendu estimé sur la base de données telles que celles du Réseau Sentinelles des médecins généralistes. Ce choix conduit à des estimations différentes du RR. La définition (i) ne peut en particulier produire un résultat comparable aux autres définitions, la définition de l'exposition n'étant pas la même (consommer l'eau du robinet vs. résider sur le secteur pollué). L'option (i) engendre en outre un biais spécifique. L'oubli du répondant ou l'occurrence d'expositions occultes (par exemple : ingestion d'eau à l'occasion du brossage des dents) produisent des erreurs de classement non différentielles sur l'exposition et une sous-estimation du RR (exemple de Dracy et de Gourdon où l'on a constaté un TA significatif chez les prétendus non exposés).

En ce qui concerne l'épidémie de l'Isère qui a bénéficié d'une investigation classique (cohorte rétrospective) et d'une évaluation d'impact basée sur l'analyse des prescriptions médicamenteuses de la CnamTS, les RR varient de 5 (option (i), délimitation spatio-temporelle extensive de l'épidémie) à 24 (option (ii), délimitation spatio-temporelle stricte de l'épidémie).

Le RR utilisé dans ce travail pour le calcul de puissance est le risque épidémique sur une semaine et sur l'ensemble de l'UDI, rapporté au risque sur un secteur témoin. Il est donc plus pertinent de retenir pour les épidémies historiques les valeurs de risque s'appuyant sur l'observation du risque dans un secteur ou une période témoin et sur une délimitation spatio-temporelle "serrée" de l'épidémie.

Les tailles des populations exposées lors des récentes épidémies s'échelonnent de 1 100 (Dracy) à 60 000 (Strasbourg). Les TA rapportés par les investigateurs se

situent entre 20 % à Pont-de-Roide et 51 % à Dracy et le nombre de cas estimé entre 560 (Dracy) et 2 600 (Gourdon). Les valeurs minimales des RR "comparables" s'établissent à 2 (Strasbourg) et 4 (Divonne).

Parmi les six épidémies investiguées, une a été signalée à la Ddass par le Samu (Strasbourg), une par un particulier (Isère), une par un maire (Pont-de-Roide), une par l'exploitant d'eau potable (Divonne)<sup>8</sup> et seulement deux par des généralistes. Aucune n'a été signalée par un généraliste sur la base d'un excès de cas constaté en population générale. Le signalement par le généraliste

<sup>8</sup> Des signalements avaient été adressés une semaine auparavant à la mairie et à l'exploitant par des généralistes mais n'avaient pas été relayés vers la Ddass.

est en fait conditionné par l'identification d'un excès flagrant de cas de GEA, c'est-à-dire un TA localement fort. Le regroupement d'un grand nombre de cas sur un même lieu (hôtel, institution, etc.) a par exemple provoqué la détection des épidémies de Dracy et de Gourdon. L'épidémie de Strasbourg se distingue par un RR faible. La conjonction du pic épidémique et du week-end a cependant concentré la demande de soins sur le Centre 15 et a permis le repérage du phénomène par les médecins urgentistes. À Pont-de-Roide, c'est l'absentéisme scolaire qui a révélé l'épidémie et provoqué le signalement du maire.

Mis à part Strasbourg où le TA semble avoir été faible et Pont-de-Roide où l'épidémie a pu être concomitante et confondue avec l'épidémie virale hivernale, l'existence des épidémies n'a pu échapper aux généralistes (13 environ) en exercice sur les lieux et au moment des épidémies (plus d'un tiers de la population touchée !).

### 5.1.2 | Délai de détection

Pour cinq des six épidémies investiguées, le signal parvenu à l'autorité sanitaire consistait en un excès de cas de GEA. Seule l'épidémie de Strasbourg a été précédée de 2 jours par le signalement de la pollution du réseau. Ceci a permis aux autorités sanitaires d'émettre un avis de restriction de consommation le jour même de l'éclosion. A l'opposé, les épidémies de Gourdon et de Divonne ont été signalées à l'autorité 9 et 11 jours après leur début, respectivement. En moyenne, 5 jours se sont écoulés entre le début de l'épidémie et le signalement. Dans tous les cas, celui-ci a eu lieu au mieux le jour même du pic épidémique. Cette observation suscite deux interprétations opposées : soit l'alerte a été donnée au moment du pic naturel de l'épidémie, soit les mesures prises consécutivement au signalement ont fait chuter le nombre de cas incidents à 24 heures.

En termes de réactivité, les performances d'un système automatisé se situeraient au niveau du système déclaratif. Ce diagnostic devra être précisé par l'étude pilote car un éventuel avantage pourrait apparaître selon le rythme de transmission des données, entre le pharmacien et l'AM d'une part, l'AM et l'InVS d'autre part. La source de progrès la plus intéressante en matière de réactivité reste l'anticipation que procurerait une meilleure prise en compte des signaux environnementaux. Parmi ceux-ci, les plaintes de la population relatives à la qualité de l'eau distribuée auraient permis le repérage des

On dénombre cependant globalement seulement 2 signalements efficaces, ce qui indique une mobilisation modérée de la profession pour le signalement des cas groupés de GEA.

Quelle que soit l'épidémie investiguée, c'est la concentration de cas en un lieu et à un moment particulier qui a révélé l'épidémie. Il est donc difficile d'établir une comparaison entre les performances du système basé sur les déclarations de témoins et d'un système automatisé fondé sur l'émergence d'un RR à l'échelle de la commune. Compte tenu de l'importance de la commune et des RR constatés, un système d'alerte basé sur l'analyse hebdomadaire des données de l'AM aurait vraisemblablement détecté toutes les épidémies évoquées. La sensibilité d'un tel dispositif dépasserait largement celle du système déclaratif conventionnel et l'affranchirait notamment du facteur humain et de l'occurrence d'opportunités telles que celles qui viennent d'être décrites.

pollutions par introduction directe d'eaux usées ou d'eau naturelle polluée dans le réseau, qui représente le risque le plus sournois (aucun signe précurseurs), le plus fort (fortes concentrations en pathogènes), le plus immédiat (exposition des buveurs sans délai) et le plus fréquent (3/6) ! En effet, ces situations s'accompagnent systématiquement (3/3) de plaintes nombreuses. Il est remarquable à ce niveau de constater que la survenue de plaintes n'est considérée, ni par l'autorité publique (pas d'obligation de signalement), ni par les exploitants, comme le précurseur d'un risque sanitaire. L'argument selon lequel ce signal souffre d'un défaut rédhibitoire de spécificité (abondance des plaintes contre le chlore, le calcaire, la couleur rouille...) possède une portée réduite car, en cas de pénétration d'eau usée dans le réseau, le sujet des doléances est spécifique : odeur d'égout, trouble inhabituel.

En conclusion, l'intérêt du système automatisé par rapport au système déclaratif réside principalement dans sa sensibilité, qui devrait surclasser celle du dispositif en place. En terme de réactivité, la comparaison reste à établir. Il est d'une part difficile de présager des délais de transmission réels dans le système automatisé. Le système déclaratif possède d'autre part l'avantage (seulement potentiel en l'état actuel des choses) de disposer de signaux environnementaux qui anticipent les effets sanitaires.

## 5.2 | Limites des bases de données de l'AM pour l'alerte

### 5.2.1 | Limites de sensibilité

Les bases de l'AM sont les seules ressources autorisant un objectif de détection automatisée et précoce de cas groupés de GEA liés à la distribution d'eau. Ces systèmes d'information ont cependant des limites de puissance et de réactivité.

Certains facteurs entament la sensibilité du système :

- absence des ventes conseils (30 % de perte de puissance par rapport au total des cas vus par le pharmacien) ;
- erreurs sur l'adresse de l'assuré liées aux pratiques de certaines mutuelles (14 %). Ce problème devrait se résorber au cours de l'année 2006 ;
- erreurs sur l'adresse du bénéficiaire, donnée indisponible et remplacée par celle de l'assuré (< 10 %) ;
- remboursements par envoi de la feuille de soin (< 10 %).

Ces facteurs défavorables restent cependant secondaires en regard des deux facteurs suivants :

- Réactivité. Un délai d'une semaine entre la délivrance des médicaments et la disponibilité des données paraît être le délai limite pour conférer au système un rôle d'alerte, une majorité d'épidémies de GEA d'origine hydrique ayant une dynamique explosive. Or en 2003, seulement un tiers des pharmacies télétransmettaient les données de prescription au moins une fois par semaine à l'AM. Ce problème se serait traduit par une perte de puissance moyenne du système d'un facteur 3, mais surtout par l'occultation des phénomènes épidémiques survenant dans les secteurs où l'ensemble des usagers de l'UDI contaminée sont clients d'une pharmacie à basse fréquence de connexion. La fréquence des connexions s'est cependant

considérablement accrue et des tests récents (ventes de TAMIFLU™ du dernier trimestre 2005) indiqueraient que plus de 90 % des prescriptions réalisées au titre du Régime général sont accessibles à 2 (Erasme) ou 3 (SNIIR-AM) jours ouvrables à compter de la délivrance des médicaments. L'effet limitant de ce facteur semble donc maintenant en grande partie résorbé ;

- Taux de couverture (TC) du Régime général. Le TC d'Erasme est limité au Régime général. SNIIR-AM couvre l'ensemble des régimes mais seules les données du Régime général bénéficient d'une mise à jour quotidienne. Les données des autres régimes sont introduites mensuellement. Vis-à-vis de l'objectif d'alerte, les deux bases sont donc en pratique limitées au Régime général. Cette contrainte pénalise les secteurs ruraux, où le TC du Régime général descend jusqu'à 50 % de la population, qui correspondent malheureusement aux secteurs les plus à risque.

Globalement, seulement 30 % (secteurs ruraux) à 60 % (secteurs urbains) des cas incidents de GEA médicalisés seraient actuellement exploitables pour l'alerte. L'option du repérage automatisé des clusters de GEA est très intéressante au regard du système déclaratif actuel dont le taux d'exhaustivité n'est pas connu, mais probablement faible du fait du manque de sensibilité intrinsèque des signaux sanitaires disponibles (sauf TA élevé) et d'un signalement peu fréquent des événements repérés par les professionnels de santé. Quant aux systèmes d'information disponibles pour la surveillance de l'exposition, ils se limitent à SISE-eaux. Étant donnée la fréquence des analyses (tableau 1), la sensibilité de cette source ne deviendrait correcte dans un objectif d'alerte que pour les UDI desservant plus de 100 000 habitants. Elle serait en revanche inacceptable en deçà de 15 000 usagers.

### 5.2.2 | Contraintes sur l'unité géographique élémentaire

L'UDI s'impose par définition comme unité géographique pour l'étude des problèmes imputables à la qualité de l'eau distribuée. Il existe cependant des pollutions dont l'extension spatiale ne correspond pas à celle de l'UDI. De plus, les échelons administratifs disponibles pour l'agrégation des données sanitaires ne correspondent pas à l'UDI. Ces deux problèmes sont détaillés ici. Une conduite à tenir est proposée.

Une étude menée par l'InVS en région Auvergne a montré que la fréquence des communes desservies par plusieurs UDI concernait environ 20 % de la population. Ce problème prend une importance très variable selon le type de paysage. La taille des UDI dépend en effet d'une part, des ressources disponibles et d'autre part, de la géographie de l'habitat. La dispersion de l'habitat et le relief montagneux constituent des facteurs

d'atomisation de la distribution, y compris au sein d'une même commune. On peut ainsi opposer les secteurs urbains ou les régions rurales d'habitat groupé, pour lesquels l'adéquation entre limites communales et limites d'UDI est globalement satisfaisante, aux secteurs ruraux d'habitat dispersé, et *a fortiori* montagnards, pour lesquels cette adéquation est moins bonne.

Pour les communes de ces derniers secteurs, la puissance statistique du système de détection se trouve diminuée. Par exemple, si la commune est desservie par deux UDI d'égale importance, le RR minimal détectable est deux fois plus élevé que ceux indiqués en figures 2 et 3.

En résumé, l'inadéquation entre les limites communales et les limites d'UDI n'est pas la règle et, quand elle survient, cela n'anéantit pas la possibilité de repérage des épidémies.

La pertinence du choix de l'UDI comme unité d'exposition trouve, d'autre part, ses limites dans des situations exceptionnelles. Une première catégorie regroupe les situations où le secteur d'exposition est plus petit que l'UDI :

- retour d'eau dans le réseau ; dans ce cas, l'exposition se limite à la portion polluée du réseau ;
- des goûts ou des odeurs désagréables engendrées par la stagnation de l'eau dans des parties distales ou mal irriguées du réseau ; cela peut conduire à des variations de consommation, le cas échéant d'exposition, au sein de l'UDI.

Dans ce cas, comme dans le cas d'une commune englobant plusieurs UDI, le RR serait dilué et la probabilité de détection du cluster diminuée.

Une deuxième catégorie inclut les situations où la zone exposée comprend au contraire plusieurs UDI :

- interconnexion temporaires entre réseaux ; l'ensemble des UDI interconnectées constituent l'unité d'exposition ;
- événements météorologiques de grande ampleur (inondations) entraînant la contamination concomitante de nombreuses UDI d'un même secteur par des eaux de ruissellement. Ces situations ne sont plus exceptionnelles.

Il existe alors un risque de faux négatif consécutif au choix d'un secteur de référence touché pour l'estimation du dénominateur dans le calcul du RR.

Tous ces cas de figure ont en commun de brouiller le rapport entre UDI et exposition d'origine hydrique à risque. Ils devront faire l'objet d'un traitement spécifique dans le test d'adéquation entre les clusters de cas de GEA et le découpage de la distribution de l'eau.

À ces erreurs liées à l'extension spatiale des pollutions s'ajoutent les erreurs sur la délimitation géographique des UDI renseignées dans SISE-eaux. La mise à jour dans SISE-eaux de l'information sur la structure de la distribution est en pratique subordonnée ou portée à connaissance de la Ddass par l'exploitant. L'expérience montre qu'il peut exister dans SISE-eaux des informations périmées. Le taux d'erreur, non précisément estimé, semble cependant limité.

L'adresse de résidence de l'assuré est renseignée dans Erasme. Cette information est intéressante dans le sens où elle permet d'éviter les erreurs d'affectation des cas aux UDI imputables à l'inadéquation spatiale entre communes et UDI. L'obtention et l'exploitation de l'adresse pose cependant des problèmes de trois ordres :

- acceptation de la part de l'AM de greffer, à terme, le système d'alerte sur Erasme. SNIIR-AM a été créée pour libérer Erasme d'une utilisation statistique pouvant perturber la liquidation des actes de soins, qui est maintenant la seule vocation d'Erasme ;
- obtention de l'autorisation de la Commission nationale de l'informatique et des libertés. Cette donnée rompt l'anonymat des cas car les données géocodées et référencées à la journée deviennent indirectement nominatives ;
- nécessité d'incorporer dans la chaîne de traitement de l'information une fonction de géocodage automatique des adresses dont le fonctionnement alourdirait le processus et pourrait altérer sa fiabilité.

Globalement, il semble raisonnable de poursuivre l'étude de faisabilité en se satisfaisant d'un positionnement par commune de résidence. Cette option permet notamment de recourir à SNIIR-AM dès lors que les codes commune et postaux y seront renseignés.

## 5.3 | Retour sur la définition de l'événement critique

Nous avons défini l'événement critique comme le dépassement d'un seuil dans le dénombrement des bactéries fécales obtenu sur un échantillon réglementaire. Nous avons classé les départements sur la fréquence des événements critiques ou sur des indicateurs dérivés. Quelles sont les limites des indicateurs bactériens

utilisés ? Le protocole d'échantillonnage mis en œuvre pour le contrôle réglementaire engendre-t-il des biais dans le classement des départements ? Le classement est-il dépendant de l'année d'étude ? Diffère-t-il enfin du classement fondé sur la fréquence des non-conformités ?

### 5.3.1 | Limite des indicateurs bactériens de contamination fécale

Le choix des bactéries indicatrices de pollution fécale pour approcher le risque d'infection prête à discussion :

- le contrôle réglementaire n'est pas apte à saisir les pollutions transitoires, notamment quand elles se produisent en dehors des heures ouvrables. Les données fournies par les analyseurs en continu montrent par exemple que les défaillances de désinfection de durée horaire se concentrent pendant la nuit ou le week-end [5]. Ce fait entraîne une sous-estimation de la fréquence des pollutions de l'eau distribuée. Le biais n'est cependant pas différentiel et n'altère donc pas le classement proposé des départements à risque épidémique ;

- l'interprétation de la présence des bactéries indicatrices dans un échantillon diffère selon le traitement subi. Si l'eau n'est pas désinfectée (1,5 % des UDI de la classe des UDI d'intérêt [1]), l'absence d'indicateur accrédite l'absence de pathogènes d'origine fécale. Si l'eau est désinfectée (chloration...), l'absence d'indicateur accrédite seulement l'absence des pathogènes dont la résistance au chlore est inférieure ou égale à celle de l'indicateur, c'est-à-dire les bactéries d'origine fécale ; tandis que le risque parasitaire, réfractaire à la chloration, est juste masqué. La prise en compte du traitement dans le critère de jugement est cependant complexe et difficile sur la base de l'information disponible dans SISE-eaux. Ce sujet pourrait être l'objet de développements ultérieurs.

### 5.3.2 | Fréquence de prélèvement et impact sur le choix des départements pilotes

La fréquence d'échantillonnage varie d'un facteur 2 entre la borne inférieure (500 habitants desservis) et la borne supérieure (10 000 habitants desservis) de l'intervalle de taille des UDI d'intérêt. Ce ratio se reporte sur la probabilité de détection d'un événement critique pour une UDI donnée.

L'espérance  $m$  du nombre  $n$  d'événements critiques identifiés au cours d'une année est proportionnelle au nombre  $N$  d'analyses effectuées :  $n$  suit une loi de Poisson de paramètre  $m = N.p$ , où  $p$  est la probabilité de survenue d'un événement critique, considérée comme

constante au cours du temps pour une UDI donnée. La probabilité de détecter au moins un événement, sous-jacente à la définition d'UDI critique, est  $1 - P(0 | m) = 1 - e^{-N.p}$ . Cette dernière probabilité est en revanche moins sensible à  $N$  que  $m$ , ce qui justifie de privilégier le 2<sup>ème</sup> critère dans le choix des départements pilotes.

Il n'en demeure pas moins que  $P(0 | m)$  reste dans l'absolu sensible à  $N$  (tableau 8). Dans le domaine de vraisemblance maximale de  $p$  :  $[10^{-2} ; 10^{-3}]$ , la quantité  $P(0 | p)$  varie presque proportionnellement à  $N$ .

**Tableau 8 - Probabilité de détecter au moins un événement de pollution critique au cours d'une année en fonction de la fréquence des pollutions critiques ( $p$ ) et du nombre ( $N$ ) d'analyses pratiquées annuellement**

$p$	$N = 8$ (a)	$N = 16$ (b)	(b/a)
0,2	0,798	0,959	1,20
0,1	0,551	0,798	1,45
<b>0,01</b>	<b>0,077</b>	<b>0,148</b>	<b>1,92</b>
<b>0,001</b>	<b>0,008</b>	<b>0,016</b>	<b>1,99</b>

En conséquence, il faut considérer que le nombre de petites UDI critiques [500-2 000] est sous-estimé d'un facteur proche de 2 par rapport au nombre de grandes UDI [5 000-10 000] critiques. Ceci engendre un biais dans la sélection des départements dans la mesure où la répartition par taille des UDI critiques varie entre les départements.

Ce biais n'a pas été traité car il ne paraît pas à même de changer les premières positions du classement.

Enfin, il existe une possibilité de biais différentiel dans le classement des départements lié au fait que certaines

Ddass accentuent leur effort d'échantillonnage sur des périodes défavorables ou sur les secteurs à problème, tandis que d'autres adoptent d'autres stratégies. Ce biais ne peut pas être contrôlé simplement mais devrait être limité par les contraintes réglementaires pesant sur l'échantillonnage (le supplément d'échantillonnage imposé au producteur d'eau ne peut légalement excéder 20 % de la prescription de base) et par les contraintes logistiques (organisation des tournées en périodes pluvieuses).

### 5.3.3 | Variations interannuelles de classement

La notion d'événement critique reste discutable parce que l'usage de bactéries indicatrices peut occulter les risques attribuables à d'autres catégories de microbes, parce que l'échantillonnage n'est pas aléatoire et parce que la justification du seuil est plus pratique que scientifique. Il n'existe cependant pas d'alternative réaliste à cet index du risque d'épidémie, sinon la réalisation de l'étude pilote projetée.

Aux biais mentionnés, s'ajoute une variabilité interannuelle de la fréquence des événements critiques, aux niveaux national et départemental. Le classement des départements vis-à-vis du risque épidémique peut donc changer d'une année sur l'autre.

En Haute-Normandie, les variations annuelles des précipitations conduisent à des années avec ou sans épisode de ruissellement. Comme les pollutions observées sur les réseaux dépendent en grande partie des pollutions de la ressource, l'effet "département",

qui modifie la probabilité de l'événement critique, peut être significatif une année et disparaître l'année suivante. On peut ainsi légitimement rapprocher l'absence d'effet "département" significatif en Seine-Maritime de l'absence d'épisode de ruissellement en 2003 (Jean-Paul Dupont, communication personnelle). La valeur moyenne interannuelle de l'effet "département" ne peut cependant être appréciée que sur une longue période de temps, à l'instar des phénomènes paroxysmiques d'origine météorologique. En l'occurrence, la prise en compte de 2004 et 2005, nouvelles années sèches, ne modifierait probablement pas le statut de la Seine-Maritime.

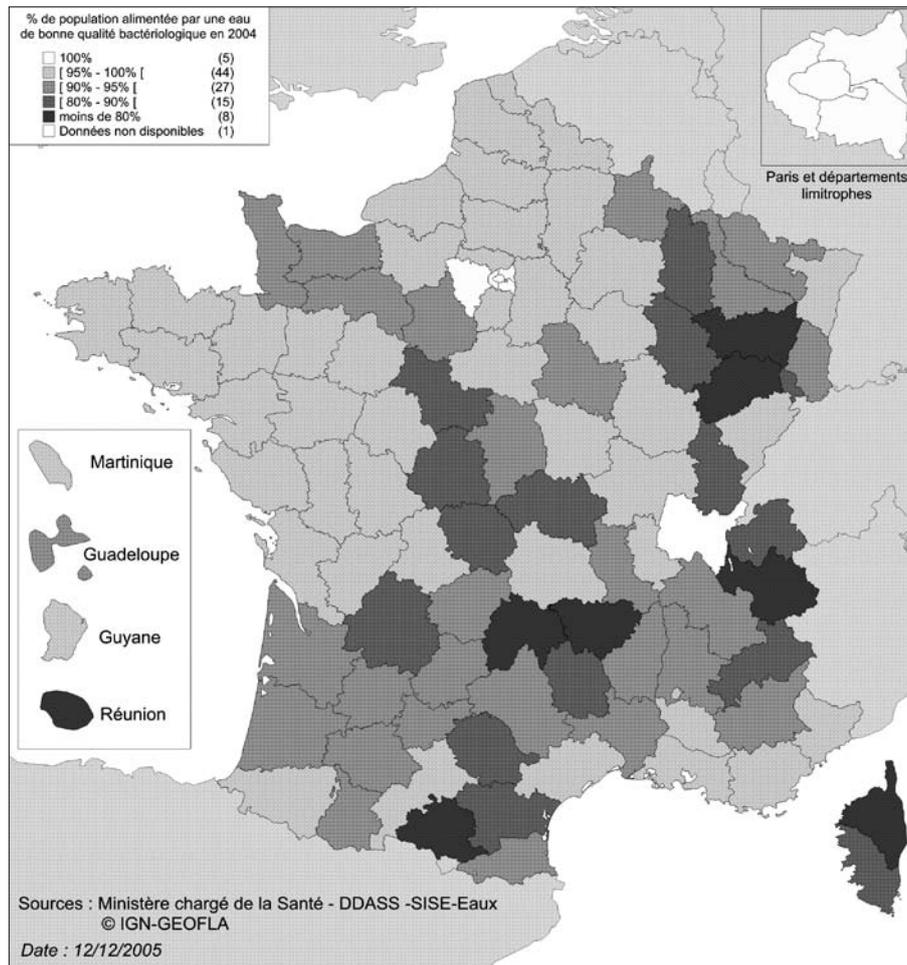
L'exploitation d'une longue période de temps n'est pas actuellement envisagée, les objectifs du travail étant atteints (mettre en évidence quelques départements à risque, avec validation du résultat par les experts locaux).

### 5.3.4 | Choix du seuil : événement critique vs. non-conformité

L'absence de données épidémiologiques permettant d'établir une relation dose-réponse entre les niveaux de contamination des réseaux d'eau, appréciés par les résultats du contrôle sanitaire, et les cas de GEA nous ont conduit à proposer arbitrairement 50 bactéries fécale/100 mL dénombrées par prélèvement réglementaire sur l'eau traitée comme valeur seuil pour le risque d'épidémie d'origine hydrique. L'application de cette règle à l'ensemble du territoire sur l'année 2003 donne un effectif de population exposée de 1,5 millions

de personnes. Pour la même année, 3,5 millions de personnes ont été exposées à une eau non conforme (au moins une bactérie fécale/100 mL identifiée). La catégorisation des départements selon le critère de fréquence des non-conformités (figure 8) diffère de celle obtenue selon le critère Pop1 parce que d'une part, Pop1 ne prend en compte que les UDI de taille 500-10 000 et d'autre part, parce que le rapport entre le nombre de non-conformités et le nombre d'événements critiques pourrait varier d'un département à l'autre.

Figure 8 - Qualité microbiologique des eaux potables - Situation en 2004 (source : ministère chargé de la Santé – Drass – Ddass – SISE-eaux)



Les départements alpins ou pyrénéens, caractérisés par une forte proportion de très petites UDI (par exemple, 875 UDI en Isère) subissant fréquemment des pollutions fécales, ressortent en figure 8 mais n'apparaissent pas en tête du classement restreint aux UDI de taille [500-10 000]. Il serait intéressant d'établir des comparaisons entre les classements des départements par fréquence de "non-conformité" vs. "événement critique", en contrôlant l'effet différentiel de la distribution par taille.

Une fois que le système d'alerte fonctionnera, on pourra envisager de réviser le seuil de 50 UFC/100mL et plus généralement redéfinir l'événement critique (prise en compte du traitement, de la turbidité...) en fonction des observations accumulées. Il s'agira de rechercher la définition de l'événement critique qui maximise la corrélation entre leur fréquence et celle d'occurrence des clusters de GEA, c'est-à-dire le pouvoir prédictif de l'événement critique vis-à-vis du risque épidémique.

Les perspectives offertes par ce projet sont lointaines et incertaines. Il dépend en effet de l'adhésion de la CnamTS au projet, de la dotation en moyens, de la mise au point du traitement et de l'implémentation informatique.

Des échéances sont avancées correspondant au cas favorable. L'impact pour les services déconcentrés du ministère est abordé ainsi que les possibilités d'adaptation aux conditions locales.

## 6.1 | Échéances

Dans le meilleur des cas, des données seraient fournies courant 2006. En se basant sur cette hypothèse, les développements méthodologiques et la constitution de la plateforme logicielle de traitement des données pourraient avoir lieu en 2007-2008. La mise au point et le test de ces outils se baseraient sur la simulation de données rétrospectives et ne déboucheraient donc

pas sur des situations d'alerte durant cette phase. Au terme de cette étape et dans la mesure où elle serait concluante, une phase opérationnelle pilote pourrait démarrer. Les investigations de terrain destinées à valider le caractère hydrique de l'exposition des clusters pourraient commencer en 2009-2010, dans le meilleur des cas.

## 6.2 | Implication potentielle du projet pour les Ddass

En dépit du caractère lointain et incertain de ces perspectives, il semble utile de souligner l'implication potentielle du projet au niveau des partenaires de terrains :

- ce projet ne s'inscrit pas dans le cadre d'une mission nouvelle des services déconcentrés mais au contraire vise à fournir aux partenaires de terrains un outil nouveau dédié à la mission d'investigation des épidémies ;
- ce projet ne se substituera pas à l'approche traditionnelle reposant sur le signalement de cas groupés ou de pollutions accidentelles des eaux ; il la compléterait et la stimulerait ;
- la détection d'un cluster possiblement lié à l'eau distribuée suppose une réponse immédiate de santé publique : a minima, il s'agit d'une enquête environnementale sommaire visant à établir le rôle de l'eau et, le cas échéant, (i) à prendre les mesures pour faire cesser l'exposition, (ii) à identifier les facteurs de risque pour éviter les récurrences ;
- l'exploitation des données de la CnamTS constitue en soi une investigation épidémiologique descriptive, avec localisation spatio-temporelle des cas à court terme (la résolution spatiale du système étant limitée

à la commune) et possibilité d'établir un bilan sanitaire *a posteriori* (à 2 mois et limité aux cas médicalisés). Le système pourrait donc soulager les Ddass à ce niveau ;

- les autres investigations ne présentent pas le même degré de priorité et doivent être menées en fonction des moyens disponibles, des opportunités et de l'importance du problème. Les investigations microbiologiques ont un degré de priorité supérieur aux investigations épidémiologiques "analytiques", qui sont à envisager avec circonspection (plusieurs mois x homme...) et plutôt en partenariat avec la Cire.

La simulation effectuée sur les données de Sise-eaux de l'année 2003 montre qu'en moyenne moins de deux événements critiques de pollution de l'eau distribuée sont identifiés par département et par an. Cette donnée ne permet en aucun cas de présager du nombre de clusters à enquêter annuellement. D'une part, les prélèvements réglementaires ne concernent que 8 à 17 jours de l'année (pour la classe des UDI d'intérêt, soit 500-10 000 habitants desservis). Il y a donc une sous-évaluation notoire du nombre d'événements critiques survenant réellement. D'un autre côté, la valeur prédictive de la pollution critique par rapport à l'émergence d'une

épidémie n'est pas établie et vraisemblablement assez faible (le critère "fréquence des pollutions critiques" est un "proxy" du risque dont l'ambition se limite au classement des départements). Le solde de ces

deux effets antagonistes ne peut être évalué. Seule l'étude pilote permettra d'estimer le nombre potentiel d'investigations et de préciser le surcroît éventuel de travail demandé aux Ddass.

### 6.3 | Ajuster le seuil d'alerte sur le bénéfice de santé publique escompté

Les acteurs de terrain doivent pouvoir maîtriser leur niveau d'implication par le contrôle du seuil de déclenchement de l'alerte.

L'effort à consacrer à une investigation comporte une part variable, fonction de la taille de la population et du nombre de cas. La part fixe liée à la prise de contact, à l'enquête environnementale, aux calculs et à la rédaction du rapport est cependant toujours importante. Il est dès lors souhaitable de diriger les moyens disponibles vers les investigations les plus rentables en terme de santé publique, c'est-à-dire celles qui maximisent l'impact évité par unité d'effort investi. Dans la mesure où il peut exister des différences dans les dotations en moyens selon les départements, il apparaît souhaitable que le réglage du seuil soit laissé à l'appréciation de la Ddass.

Les scénarios limites utilisés pour le calcul de puissance peuvent être traduits en nombre de cas pathologiques (détail en annexe 1). Le tableau 9 montre que le nombre de cas impliqués dans les scénarios d'épidémie à la limite de l'observable (cf. calcul de puissance) est très faible, notamment dans le cas d'UDI desservant moins de 2 000 habitants. Selon ces scénarios, les N cas pathologiques sont survenus dans la semaine précédant l'alerte et ont permis son déclenchement. Dans l'hypothèse où l'épidémie perdurerait avec une intensité égale au-delà de cette première semaine, le nombre hebdomadaire de nouveaux cas serait pratiquement égal à N dans un premier temps, puis diminuerait jusqu'à épuisement de la population susceptible. N peut donc aussi être vu comme le nombre hebdomadaire de cas évités si des mesures adéquates sont prises.

**Tableau 9 - Scénarios d'épidémie à la limite de la possibilité de détection.  
Hypothèses sous-jacentes décrites au paragraphe "Puissance statistique"**

Population exposée	Risque relatif	Taux d'attaque	Nombre de cas (N)
500	10,0	1,54 %	8
1 000	5,0	0,77 %	8
2 000	4,0	0,62 %	12
5 000	2,5	0,38 %	19
10 000	2,0	0,31 %	31

- [1] Ministère chargé de la santé. L'eau potable en France, situation en 2003. 1-18. 2005. Paris.
- [2] Conseil national de l'information statistique. Connaissance statistique du médicament. 2005.
- [3] Zmirou D, Ferley JP, Collin JF, Charrel M, Berlin J. *A follow-up study of gastro-intestinal diseases related to bacteriologically substandard drinking water*. Am.J.Public Health 1987;77(5):582-84.
- [4] Beaudéau P, Zeghnoun A, Ledrans M, Volatier JL. Consommation d'eau du robinet pour la boisson en France métropolitaine : résultats tirés de l'enquête INCA1. Environnement, risques & santé 2003;2(3):147-58.
- [5] Beaudéau P, Payment P, Bourderont D *et al.* *A time series study of anti-diarrheal drug sales and tap-water quality*. Int J Environ Health Res 1999;9:293-311.

# 8 Annexes

## Annexe 1 | Résultats du calcul de puissance sous Epi-info

PARAMÈTRES	SYMBOLES	VALEURS
Risque de première espèce	$\alpha$	0,05
Puissance recherchée	$1-\beta$	0,8
Rapport NE/E	n0/n1	1/10
Incidence hebdomadaire estimée à partir des données de l'AM	P0	0,001

Risque Relatif	n1Epi-info = population exposée	n0Epi-info = population non exposée	Nombre de sujets nécessaires compte tenu de la proportion de non consommateurs (30 %)	Incidence hebdomadaire chez les exposés	Nb de cas hebdomadaires chez les exposés (N)
1,4	51 200	512 000	73 143	0,14 %	102
1,5	34 100	341 000	48 714	0,15 %	73
1,6	24 600	246 000	35 143	0,16 %	56
1,7	18 800	188 000	26 857	0,17 %	46
1,8	14 900	149 000	21 286	0,18 %	38
1,9	12 200	122 000	17 429	0,19 %	33
2	10 200	102 000	14 571	0,20 %	29
2,1	8 690	86 900	12 414	0,21 %	26
2,2	7 520	75 200	10 743	0,22 %	24
2,3	6 600	66 000	9 429	0,23 %	22
2,4	5 850	58 500	8 357	0,24 %	20
2,5	5 240	52 400	7 486	0,25 %	19
2,6	4 730	47 300	6 757	0,26 %	18
2,7	4 300	43 000	6 143	0,27 %	17
2,8	3 930	39 300	5 614	0,28 %	16
2,9	3 610	36 100	5 157	0,29 %	15
3	3 440	34 400	4 914	0,30 %	15
4	1 825	18 250	2 607	0,40 %	10
5	1 215	12 150	1 736	0,50 %	9
6	900	9 000	1 286	0,60 %	8
7	700	7 000	1 000	0,70 %	7
8	580	5 800	829	0,80 %	7
9	490	4 900	700	0,90 %	6
10	420	4 200	600	1,00 %	6
11	370	3 700	529	1,10 %	6
12	330	3 300	471	1,20 %	6
13	300	3 000	429	1,30 %	6
14	270	2 700	386	1,40 %	5
15	250	2 500	357	1,50 %	5
16	230	2 300	329	1,60 %	5
17	210	2 100	300	1,70 %	5
18	200	2 000	286	1,80 %	5
19	180	1 800	257	1,90 %	5
20	170	1 700	243	2,00 %	5

## Annexe 2 | Description des données microbiologiques des départements ayant présenté au moins un dépassement du seuil 50 UFC/100mL pour la somme *Escherichia coli* + entérocoques au cours de l'année 2003

Départements	<100-499				500-4 999				5 000-9 999				>10 000				Total					
	PLV >=50	Nb PLV	nb INS	Population concernée par NC	PLV >=50	Nb PLV	nb INS	Population concernée par NC	PLV >=50	Nb PLV	nb INS	Population concernée par NC	PLV >=50	Nb PLV	nb INS	Population concernée par NC	PLV >=50	Nb PLV	nb INS	Population exposée	Pop totale département (INSEE)	POP exposée/ POP totale
2 (Aisne)	3	8	2	233	0	0	0	0	0	0	0	0	0	0	0	0	3	8	2	233	535 489	0,04%
3 (Allier)	5	14	2	250	4	10	2	2 685	0	0	0	0	0	0	0	0	9	24	4	2 935	344 721	0,85%
4 (Alpes-de-Haute-Provence)	10	27	9	818	0	0	0	0	0	0	0	0	0	0	0	0	10	27	9	818	139 561	0,59%
5 (Hautes-Alpes)	15	39	13	1 162	0	0	0	0	0	0	0	0	0	0	0	0	15	39	13	1 162	121 419	0,96%
6 (Alpes-Maritimes)	17	41	10	860	0	0	0	0	0	0	0	0	0	0	0	0	17	41	10	860	1 011 326	0,09%
7 (Ardèche)	51	83	31	2 333	6	22	4	3 228	0	0	0	0	0	0	0	0	57	105	35	5 561	286 023	1,94%
8 (Ardennes)	6	7	4	324	3	2	2	4 000	0	0	0	0	0	0	0	0	9	9	6	4 324	290 130	1,49%
9 (Ariège)	53	115	34	864	0	0	0	0	0	0	0	0	0	0	0	0	53	115	34	864	137 205	0,63%
10 (Aube)	0	0	0	0	1	8	1	2 870	0	0	0	0	0	0	0	0	1	8	1	2 870	292 131	0,98%
11 (Aude)	33	76	22	1 659	2	14	2	3 000	1	11	1	5 241	0	0	0	0	36	101	25	9 900	309 770	3,20%
12 (Aveyron)	20	27	13	1 275	1	2	1	900	0	0	0	0	0	0	0	0	21	29	14	2 175	263 808	0,82%
13 (Bouches-du-Rhône)	1	7	1	350	2	16	2	4 650	1	22	1	7 900	0	0	0	0	4	45	4	12 900	1 835 719	0,70%
14 (Catalvas)	0	0	0	0	4	52	3	8 175	0	0	0	0	0	0	0	0	5	101	4	27 673	648 385	4,27%
15 (Cantal)	50	105	29	3 170	12	87	7	10 700	0	0	0	0	0	0	0	0	62	192	36	13 870	150 778	9,20%
17 (Charente-Maritime)	0	0	0	0	3	44	2	7 906	1	22	1	5 610	0	0	0	0	4	66	3	13 516	557 024	2,43%
19 (Corrèze)	4	6	3	109	1	11	1	3 872	0	0	0	0	0	0	0	0	5	17	4	3 481	232 576	1,50%
21 (Côte-d'Or)	8	35	6	510	0	0	0	0	0	0	0	0	0	0	0	0	8	35	6	510	506 755	0,10%
24 (Dordogne)	1	5	1	109	4	17	3	2 749	0	0	0	0	0	0	0	0	1	5	1	10 250	388 293	3,38%
25 (Doubs)	37	90	23	2 758	7	29	3	2 946	1	24	1	8 518	0	0	0	0	45	143	27	14 222	499 062	2,85%
26 (Drôme)	17	22	12	200	0	0	0	0	0	0	0	0	0	0	0	0	17	22	12	200	437 778	0,05%
28 (Eure-et-Loir)	4	11	3	835	2	9	1	906	0	0	0	0	0	0	0	0	6	20	4	1 741	407 665	0,43%
29 (Finistère)	2	11	2	396	1	5	1	4 120	0	0	0	0	0	0	0	0	3	16	3	4 516	852 418	0,53%
30 (Gard)	21	37	16	2 145	5	18	5	6 200	0	0	0	0	0	0	0	0	26	55	21	8 345	623 125	1,34%
31 (Haute-Garonne)	5	13	4	64	1	4	1	1 025	1	7	1	5 500	3	181	2	98 871	10	205	8	103 460	1 046 338	9,89%
33 (Gironde)	0	0	0	0	2	19	2	4 484	0	0	0	0	0	0	0	0	2	19	2	4 484	1 287 334	0,35%
34 (Hérault)	24	106	21	2 050	1	6	1	911	0	0	0	0	0	0	0	0	25	112	22	2 961	896 441	0,33%
36 (Indre)	3	21	3	1 136	4	12	2	2 585	0	0	0	0	0	0	0	0	7	33	5	3 721	231 139	1,61%
38 (Isère)	6	13	4	294	3	31	2	3 127	0	0	0	0	0	0	0	0	9	44	6	3 421	1 084 006	0,31%
39 (Jura)	30	79	15	1 790	0	0	0	0	0	0	0	0	0	0	0	0	30	79	15	1 790	250 857	0,71%
40 (Landes)	0	0	0	0	1	7	1	984	0	0	0	0	0	0	0	0	1	7	1	984	327 334	0,30%
41 (Loir-et-Cher)	0	0	0	0	2	13	2	2 840	0	0	0	0	0	0	0	0	2	13	2	2 840	314 968	0,90%
42 (Loire)	3	4	1	114	0	0	0	0	0	0	0	0	0	0	0	0	3	4	1	114	728 524	0,02%
43 (Haute-Loire)	34	95	21	1 757	2	8	1	1 134	0	0	0	0	0	0	0	0	36	103	22	2 891	209 113	1,38%
44 (Loire-Atlantique)	0	0	0	0	2	12	1	1 080	0	0	0	0	0	0	0	0	2	12	1	1 080	1 134 266	0,10%
45 (Loiret)	0	0	0	0	4	22	4	5 646	0	0	0	0	0	0	0	0	5	27	5	16 696	618 126	2,70%
46 (Lot)	10	26	8	1 991	9	31	5	6 501	0	0	0	0	0	0	0	0	19	57	13	8 492	160 197	5,30%
49 (Maine-et-Loire)	0	0	0	0	1	9	1	3 425	1	7	1	7 335	0	0	0	0	2	16	2	10 760	732 942	1,47%

Départements	<100-499			500-4 999			5 000-9 999			>10 000			Total		
	PLV >=50	Nb PLV	population concernée par	PLV >=50	Nb PLV	population concernée par	PLV >=50	Nb PLV	population concernée par	PLV >=50	Nb PLV	population concernée par	nb population exposée	pop totale département (INSEE)	POP exposée/POP totale
50 (Marne)	2	4	10	0	0	0	0	0	0	0	0	0	10	481 471	0,00%
52 (Haute-Marne)	29	59	2 138	7	30	2 050	0	0	0	0	0	0	4 188	194 873	2,15%
53 (Mayenne)	1	1	300	1	5	550	1	6	1 500	0	0	0	6 350	285 338	2,23%
54 (Meurthe-et-Moselle)	0	0	0	6	27	4 121	0	0	0	1	17	16 295	20 416	713 779	2,86%
55 (Meuse)	21	60	1 624	3	222	5 848	0	0	0	0	0	0	7 472	192 198	3,89%
56 (Morbihan)	0	0	0	1	13	724	0	0	0	0	0	0	724	643 873	0,11%
58 (Nièvre)	2	1	39	2	12	4 697	0	0	0	0	0	0	4 736	225 198	2,10%
59 (Nord)	0	0	0	1	3	1 430	0	0	0	0	0	0	1 430	2 555 020	0,06%
60 (Oise)	0	0	0	1	8	537	0	0	0	0	0	0	537	766 441	0,07%
63 (Puy-de-Dôme)	14	37	909	0	0	0	0	0	0	0	0	0	909	604 266	0,15%
64 (Pyrénées-Atlantiques)	11	47	942	0	0	0	0	0	0	0	0	0	942	600 018	0,16%
65 (Hautes-Pyrénées)	15	46	1 303	2	7	2 180	0	0	0	0	0	0	2 483	222 368	1,12%
66 (Pyrénées-Orientales)	5	20	463	0	0	0	0	0	0	0	0	0	463	392 803	0,12%
68 (Haut-Rhin)	0	0	0	3	22	2 708	0	0	0	0	0	0	2 708	708 025	0,38%
70 (Haute-Saône)	35	123	3 621	5	30	4 636	0	0	0	0	0	0	8 257	229 732	3,59%
71 (Saône-et-Loire)	1	4	352	2	16	9 172	0	0	0	0	0	0	9 524	544 893	1,75%
73 (Savoie)	6	15	956	7	9	1 680	0	0	0	0	0	0	2 636	373 258	0,71%
74 (Haute-Savoie)	19	39	1 918	1	3	1 150	0	0	0	0	0	0	3 068	631 679	0,49%
75 (Paris)	0	0	0	0	0	0	0	0	0	3	913	3 289 121	2 125 246	13,60%	13,60%
76 (Seine-Maritime)	0	0	0	0	0	0	1	5	1 825	0	0	0	8 255	1 239 138	0,67%
80 (Somme)	1	5	295	0	0	0	0	0	0	0	0	0	295	555 551	0,05%
81 (Tarn)	20	46	1 088	3	10	1 521	0	0	0	0	0	0	2 609	343 402	0,76%
84 (Vaucluse)	1	3	30	0	0	0	0	0	0	0	0	0	30	499 685	0,01%
85 (Vendée)	1	5	20	0	0	0	0	0	0	0	0	0	20	539 664	0,00%
86 (Vienne)	0	0	0	1	3	3 980	0	0	0	1	10	10 555	14 535	399 024	3,64%
87 (Haute-Vienne)	3	9	546	3	9	1 650	0	0	0	0	0	0	2 196	353 893	0,62%
88 (Vosges)	1	2	450	1	4	1 575	0	0	0	0	0	0	2 025	380 952	0,53%
89 (Yonne)	9	14	788	2	9	2 490	0	0	0	0	0	0	3 258	333 221	0,98%
94 (Val-de-Marne)	0	0	0	0	0	0	0	0	0	1	1 234	1 522 260	1 227 250	42,56%	42,56%
95 (Val-d'Oise)	0	0	0	1	10	2 476	0	0	0	0	0	0	2 476	1 105 464	0,22%
971 (Guadeloupe)	0	0	0	1	6	1 500	2	19	2 19 000	0	0	0	20 500	422 496	4,85%
972 (Martinique)	0	0	0	2	17	4 700	0	0	0	2	75	138 000	142 700	381 427	37,41%
973 (Guyane)	3	4	550	1	4	605	0	0	0	1	181	74 250	75 405	157 213	47,96%
974 (La Réunion)	5	10	356	8	35	13 412	2	16	1 8 000	2	3	10 300	32 068	706 300	4,54%
02A (Corse-du-Sud)	105	261	6 958	6	54	3 917	0	0	0	0	0	0	10 875	118 593	9,17%
<b>Total</b>	<b>783</b>	<b>1 938</b>	<b>54 192</b>	<b>161</b>	<b>1 088</b>	<b>181 458</b>	<b>12</b>	<b>139</b>	<b>11 80 859</b>	<b>17</b>	<b>2 673</b>	<b>15 1 198 450</b>	<b>5 154 959</b>	<b>41 156 498</b>	<b>3,68%</b>







Les bases de données de l'Assurance maladie sont les seules ressources disponibles en France qui puissent prétendre à être exploitées pour la détection en ligne des clusters de gastro-entérites aiguës. Il est possible de concevoir en aval une procédure automatisée pour examiner la vraisemblance de l'étiologie hydrique pour les clusters identifiés, par comparaison des emprises géographiques du cluster et des réseaux de distribution d'eau.

Trente à 60 % des cas de gastro-entérites ayant fait l'objet d'une consultation en France seraient disponibles dans un délai compatible avec l'objectif d'alerte (3 jours ouvrables). Les consultations faites en dehors du lieu de résidence de l'assuré ne seraient en revanche pas utilisables. L'adresse de l'assuré est disponible, mais une localisation par commune simplifierait la procédure d'exploitation et altérerait *a priori* peu les performances du système.

Ces premières conclusions doivent être mises à l'épreuve du terrain.

Pour cela, une liste de départements éligibles au rôle de pilote a été dressée. La catégorie de taille des unités de distribution cible a été fixée entre 500 habitants desservis, limite de puissance statistique du système d'alerte, et 10 000, limite au-dessus de laquelle les pollutions microbiologiques sont trop rares pour établir des conclusions statistiques robustes. A été retenu comme indice du risque épidémique la survenue "d'événements critiques" définis comme le dépassement de 50 colonies de bactéries d'origine fécale (*Escherichia coli* et entérocoques) dénombrées lors de l'analyse d'un prélèvement collecté pour le contrôle réglementaire. Les critères de classement des départements ont été (i) le nombre d'unités de distribution de la classe de population desservie [500-10 000] frappées par au moins un événement critique au cours de l'année 2003 et (ii) la taille de la population concernée. Les départements d'outre-mer (Réunion, Guadeloupe) et les départements ruraux ou karstiques (Cantal, Charente-Maritime, Doubs, Lot...) arrivent en tête de classement, sur la base des données 2003. Les départements de haute montagne n'apparaissent pas car la distribution de l'eau y est atomisée ce qui rend difficile l'approche épidémiologique. Indépendamment de l'importance de la classe des distributions d'intérêt et de l'effort d'échantillonnage, il semble se dégager un effet "département" prédisposant le Cantal, le Lot, la Réunion et le Doubs à un risque de pollution microbiologique susceptible d'engendrer des clusters "observables" plus élevés qu'ailleurs.

Les Ddass éligibles et l'Assurance maladie vont être sollicitées pour le choix des départements pilotes, l'obtention des données et la constitution du partenariat.

*The data bases of the French Health Insurance are the only available resources in France which can be run for the on-line detection of the clusters of acute gastro-enteritis. An automated process could be set up to test the spotted clusters for their relationship to the distributed water, by comparing the geographical fitting of clusters and water distribution networks. In France, 30 to 60% of the cases of gastro-enteritis detected by medical care would be available within a delay which could meet the alert objective (ie. 3 working days). The consultations made out of the insurant residence county ("département") would be excluded. The full address of the insurant is available, but a location by municipality would keep a good resolving power and simplify the operation.*

*These first conclusions have to be challenged by real experiment. In that aim, a list of departments eligible as pilots was drawn up.*

*The size category of the targeted water distribution systems ranged from 500 inhabitants, limit of statistical power, to 10 000, limit above which the microbiological pollutions are rare and no robust conclusion can be worked out. A "critical event" was defined as the occurrence of a water sample for which the number of faecal bacteria colonies (summing up *Escherichia coli* and *Enterococci*) exceeded 50, as numbered during the analysis of a sample within the frame of statutory control. The frequency of critical events was held as an index of the outbreak occurrence risk. The county classification criteria finally were (i) the number of distribution units of the size class [500-10 000] struck by at least one critical event during the year 2003 and (ii) the size of the concerned population. French overseas counties (Reunion, Guadeloupe) and rural or karstic counties (Cantal, Charente-Maritime, Doubs, Lot) came at the head of classification whereas the high mountain counties did not appear, since the water distribution is there divided there in many small systems, which makes difficult the epidemiological approach. Once controlled the importance of the targeted distribution system size class and the effort of sampling, a "county" effect raised, pointing out the Cantal, the Lot, the Reunion and the Doubs for a specific risk of microbiological pollution, possibly resulting in observable outbreaks.*

*The local partners and the French Health Insurance, will be consulted on the choice of the pilot counties, data provision and the constitution of collaborative networks.*



INSTITUT  
DE VEILLE SANITAIRE

Département santé environnement

12, rue du Val d'Osne - 94415 Saint-Maurice cedex  
Tél. : 33(0) 1 41 79 67 00 - Fax : 33(0) 1 41 79 67 67  
<http://www.invs.sante.fr>

ISBN : 2-11-096288-7  
Tirage : 600 exemplaires  
Dépôt légal : Septembre 2006  
Réalisation : Labrador