

SURVEILLANCE
SYNDROMIQUE

AVRIL 2017

MÉTHODES

LES DONNÉES DE L'ENTREPÔT DE CAS
DE GASTRO-ENTÉRITE MÉDICALISÉS
ISSUES DU SNIIRAM :
DESCRIPTION, QUALITÉ ET UTILISATION

Résumé

Les données du Système d'information inter-régimes de l'Assurance maladie (SNIIRAM) représentent une ressource unique pour développer en France l'épidémiologie des maladies d'origine infectieuses d'origine fécales véhiculées par l'eau distribuée. Les cas médicalisés de gastro-entérite aiguë (GEA), soit 32% des cas de GEA en moyenne, peuvent ainsi être repérés à partir du contenu des ordonnances médicales et étiquetés au jour et à la commune. Les cas de GEA contractée hors du lieu de résidence (touristes...) peuvent être isolés à partir de la distance entre la commune de résidence du patient et celle de prescription des soins.

Les données manquantes touchent principalement la commune de résidence du patient (taux de données manquantes ou erronées de 3,3% en 2014) et celle du cabinet du médecin (7,3%), à laquelle on peut toutefois substituer celle de la pharmacie où les médicaments ont été délivrés (0,3%). Les données sont quasi-exhaustives à quatre mois, mais leur complétude est suffisante à deux mois (99%) pour évaluer l'impact des épidémies. Ce rapport fournit à l'utilisateur une description des données et des processus impliqués dans leur constitution et leur validation, quelques repères statistiques sur les indicateurs d'incidence qui en résultent, ainsi que les informations détaillées sur leurs potentiel et limites d'utilisation. La publication de ce document a été couplée avec la mise en libre-service des données auprès des Cellules d'intervention en région de Santé publique France.

MOTS CLÉS : SURVEILLANCE SYNDROMIQUE, EAU DU ROBINET, FRANCE, BASE DE DONNÉES DE L'ASSURANCE MALADIE (SNIIRAM), PRESCRIPTIONS MÉDICALES, MÉDICAMENT

Citation suggérée : Beaudeau P., Bentayeb M., Corso M., Rambaud L., Galey C. *Les données de l'entrepôt de cas de gastro-entérite médicalisés issues du SNIIRAM : description, qualité et utilisation*. Saint-Maurice : Santé publique France, 2017. 42 p. Disponible à partir de l'URL : www.santepubliquefrance.fr

ISSN : EN COURS — ISBN-NET : 979-10-289-0335-0 — RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ PUBLIQUE FRANCE — DÉPÔT LÉGAL : AVRIL 2017

Abstract

Data from the Inter-Scheme Information System of Health Insurance (SNIIRAM) represent a unique resource in France for developing the epidemiology of fecal-borne infectious diseases transmitted through distributed water. Medical cases of acute gastroenteritis (AGE), representing on average 32% of all AEG cases, can thus be identified from the contents of medical prescriptions, and matched with the day and commune. AGE cases contracted out of the place of residence (tourists ...) can be isolated from the distance between the patient's commune of residence and that of care prescription.

The missing data mainly concern the patient's commune of residence (rate of missing or incorrect data of 3.3% in 2014) and that of the physician's office (7.3%), which can be replaced by that of the pharmacy where the drugs were delivered (0.3%). The data are almost complete at 4 months, but their completeness is sufficient at 2 months (99%) to assess the impact of epidemics.

This report provides the user with a description of the data and processes involved in their creation and validation, some statistical benchmarks on the resulting impact indicators, as well as detailed information on their potential and limits of use. The publication of this document was coupled with the provision of data to the French Regional Health Units.

KEY WORDS : SYNDROMIC SURVEILLANCE, GASTROENTERITIS, DRINKING WATER, FRANCE, HEALTH INSURANCE DATABASE, MEDICAL PRESCRIPTIONS, DRUG

Ont participé à ce rapport

Auteurs : Pascal Beaudeau, Malek Bentayeb, Magali Corso, Loïc Rambaud et Catherine Galey

Relecteur : Damien Mouly

Validation : Jean-Marc Yvon

Équipe projet : Pascal Beaudeau, Malek Bentayeb, Bénédicte Bérat, Marjorie Boussac, Magali Corso, Catherine Galey, Grégoire Falq, Javier Nicolau et Loïc Rambaud

Remerciements

Nous remercions la Cnamts pour la mise à disposition des données du SNIIRAM et pour le soutien qu'elle apporte au développement de ce projet.

Sommaire

Abréviations	5
Introduction.....	6
1. L'ALGORITHME DE DISCRIMINATION DES CAS DE GEAm.....	10
1.1 Évaluation de l'algorithme : Valeur prédictive positive (VPP) et sensibilité	12
1.2 Validation croisée de l'algorithme	13
2. L'ENTREPÔT DES RESUMÉS DE CAS DE GEAm	153
2.1 Contenu*	153
2.2 Contrôle qualité des données de l'entrepôt	18
3. QUALITÉ DES DONNÉES : LIMITES LIÉES À LA SÉLECTION DES CAS DE GEA PAR LE RECOURS AU SOIN*	197
4. QUALITÉ DES DONNÉES : LIMITES LIÉES À L'ALGORITHME ET À SON IMPLÉMENTATION*	19
4.1 Manque de spécificité de l'algorithme pour certaines classes d'âge	19
4.2 Consultations multiples	21
4.3 Renouvellements d'ordonnance	19
5. QUALITÉ DES DONNÉES : LIMITES LIÉES À LA SOURCE DE DONNÉES	23
5.1 Les données manquantes ou erronées et leur traitement : généralités*	23
5.2 Cas de GEAm absents du SNIIRAM*	23
5.3 Dossiers présentant des champs manquants	24
5.3.1 DM empêchant la catégorisation en cas/non cas de GEAm.....	24
5.3.2 DM empêchant la localisation des cas de GEAm*	25
5.3.3 Cas de GEAm dont la géolocalisation est erronée*	27
6. DÉLAIS DE CONSOLIDATION ET DE MISE À DISPOSITION DES DONNÉES*	28
7. PRÉCONISATION D'UTILISATION DES DONNÉES**	30
7.1 Généralités**	30
7.2 Cas « résident présent » et cas « touriste »**	32
8. DE L'ENTREPÔT DES CAS AUX DONNÉES « UTILISATEURS »	34
8.1 Paramétrage standard en libre-service	34
8.2 Paramétrage de données commandées	36
8.3 Fichiers de sortie.....	37
Références bibliographiques	38
Annexe 1. Historique du projet	40
Annexe 2. Ressources alternatives	42

* Section importante pour une utilisation experte des données

** Section à lire avant toute utilisation des données

Abréviations

Base de données, assurance maladie

AM Assurance maladie

BD Base de données

Cnamts Caisse nationale de l'assurance maladie des travailleurs salariés

CPAM Caisse primaire d'assurance maladie (sauf exception, il y a une CPAM par département)

Datamart Ensemble cohérent de données thématiques (ici fichiers) créé à partir de la base de données princeps (ici le SNIIRAM) fourni pour une utilisation spécifique

DCIR Datamart de consommation inter-régime du SNIIRAM. Il est accessible par Santé publique France. Il est mis à jour chaque mois

Géofla GEOFLA® est une BD de l'Institut national géographique (IGN). Elle contient la description de l'ensemble des unités administratives du territoire national. Les données existent en 2 versions: GEOFLA® Communes, et GEOFLA® Départements

Insee Institut national de la statistique et des études économiques

MSA Mutualité sociale agricole

RSI Régime social des indépendants

SLM Section locale mutualiste encore appelée mutuelles décompteuses. Il s'agit de mutuelles qui prennent en charge à la fois le remboursement de la part « sécurité sociale » qui relève du régime général (les SLM sont donc rattachées à ce régime dans le SNIIRAM) et celui de leur propre part dans la prise en charge des soins. Les SLM récupèrent ensuite la part sécu auprès de l'AM. Ces mutuelles négligeaient de communiquer les adresses de leurs sociétaires à l'Assurance maladie dans le passé.

SNIIRAM Système d'information inter-régimes de l'Assurance maladie

Médicaments

ATC (Classification Anatomical Therapeutical Chemical) : Classification OMS des médicaments à 5 niveaux. Dans sa déclinaison la plus détaillée, l'ATC correspond à un produit actif ou à un mélange de produits actifs, et à une **dénomination commune internationale (DCI)**

CIP (Code identifiant de présentation) : Code commercial des spécialités propre à la France. La correspondance CIP/ATC est actualisée tous les ans, ce qui implique que les CIP apparaissant entre 2 mises à jour ont un classement ATC provisoire jusqu'à la mise à jour suivante. Le CIP est actuellement codé sur 13 caractères

Spécialité Présentation commerciale d'un médicament. Une spécialité correspond à un code CIP

SRO Soluté de réhydratation orale

Terminologie spécifique à la méthode de discrimination des cas de Gastro-entérite aiguë médicalisée

Médicaments (spécialités/classe thérapeutique) cibles : Médicaments utilisés pour l'extraction des ordonnances. Par extension, sont inclus les solutés de réhydratation orale (SRO) qui ne sont pas considérés comme des médicaments à proprement parler mais comme des « produits de santé ». Sont exclus les antispasmodiques, jugés trop peu spécifiques

Médicaments GEA : Médicaments utilisés pour le traitement des GEA (médicaments cibles + antispasmodiques)

Médicaments excluants : Médicaments dont la présence sur l'ordonnance conduit à l'exclusion de l'ordonnance

Médicaments utiles : Ensemble des médicaments utilisés par l'algorithme. Ils comprennent les médicaments GEA et les médicaments excluants

Autres

AEC Unité air, eau, climat de l'InVS (devenue l'Unité Rariec de Santé publique France, unité qui s'intéresse aux risques associés aux contaminations de l'air et de l'eau, aux rayonnements ionisants, ainsi qu'au climat)

Algorithme GEAm Algorithme de discrimination des ordonnances prescrites pour traiter des gastro-entérites aiguës des autres ordonnances. Les ordonnances sélectionnées correspondent aux cas de GEAm

ARS Agence régionale de santé

Cire Cellule d'intervention en région de Santé publique France

Code commune (sur 5 caractères) : Code résultant de la fusion du code département (2 ou 3 premiers caractères) et du code Insee de la commune (sur 3 ou 2 caractères)

DM Données manquantes

DES Direction santé environnement de Santé publique France (jusqu'en 2016, département santé environnement de l'InVS)

Ehpad Établissement d'hébergement pour personnes âgées dépendantes

GEA Gastro-entérite aiguë. Un épisode de GEA se définit par l'émission de 3 selles liquides ou plus au cours d'une période de 24 heures ou d'au moins un vomissement. Les cas d'exclusion sont les gastro-entérites chroniques, l'indigestion, l'ingestion excessive d'alcool, les symptômes liés à la grossesse, aux règles ou à certains traitements médicamenteux. Un intervalle de 7 jours sans symptômes est requis pour considérer deux épisodes indépendants [1]

GEAm Gastro-entérite aiguë médicalisée (le malade a consulté un médecin et s'est fait délivrer les médicaments prescrits)

InVS Institut de veille sanitaire. Regroupé avec d'autres agences en 2016 pour former Santé publique France

Rariec Unité des risques liés à l'air, aux rayonnements ionisants, à l'eau et au climat au sein de la DES

TI Taux d'incidence

VPP Valeur prédictive positive (proportion de vrais positifs parmi les positifs supposés via l'application de l'algorithme de discrimination)

Introduction

Contexte institutionnel

Santé publique France assure la surveillance épidémiologique des gastro-entérites aiguës (GEA) adaptée à l'étude du risque d'origine hydrique. Cette surveillance nourrit différentes applications : évaluation du risque dans les agglomérations urbaines (projet fini), surveillance des épidémies (depuis 2010, avec des développements en cours) et surveillance du risque non épidémique (projet). Ces applications ont en commun d'être tournées vers la prévention car elles visent en premier lieu à identifier les facteurs de risque, qu'ils soient naturels, techniques ou humains. Dans l'optique d'une efficacité optimale de santé publique, les autorités sanitaires (ministère chargé de la santé et agences régionales de santé - ARS) sont associées au développement des applications.

La surveillance des GEA s'appuie sur les données de remboursement des prescriptions médicales visant à traiter une GEA, colligées par l'Assurance maladie dans la base de données SNIIRAM de la Cnamts. En 2014, la livraison semestrielle des données par la Cnamts a été remplacée par un accès direct aux données du DCIR, partie du SNIIRAM accessible à Santé publique France. L'extraction directe permet désormais de maîtriser la planification des extractions et leur fréquence (tous les deux mois dans la routine prévue). Ces extractions alimentent l'entrepôt des résumés de cas des GEA médicalisée (GEAm). En 2013-2014, le département santé-environnement de l'Invs a mis en œuvre une démarche d'assurance qualité du processus de constitution et des données de l'entrepôt des cas des GEAm (procédure DSE/AEC/005) et en 2015 la mise en œuvre d'un accès direct des Cire à l'entrepôt.

Ce guide s'adresse à l'utilisateur de ces données. Après un rappel des modalités de discrimination des cas des GEAm à partir des ordonnances colligées dans le SNIIRAM, il décrit le contenu de l'entrepôt de cas de GEAm, les modalités de mise à jour de l'entrepôt, la qualité des données originelles et des indicateurs d'incidence des GEAm qui en résultent, ainsi que les limites d'utilisation*.

Ce guide complète une procédure [2] qui s'adresse aux opérateurs du système de surveillance. Cette procédure décrit de façon détaillée les processus de constitution de l'entrepôt de données, et du contrôle de la qualité des données mises en œuvre à chaque mise à jour de l'entrepôt.

Contexte scientifique

Les données du SNIIRAM constituent la principale source de données pour la surveillance épidémiologique des GEA en France et la seule qui permette l'étude du rôle de l'eau distribuée [3].

La fréquence de la GEAm (0,11 par personne.an) et le niveau de détail des données du SNIIRAM (données individuelles étiquetées au jour et à la commune) sont adaptées à l'étude du risque infectieux d'origine hydrique. L'intérêt du système repose sur son exhaustivité géographique et le caractère pérenne de la collecte qui permettent de documenter rétrospectivement les épidémies de GEA hydrique avérées ou suspectées par les intervenants de terrain (collectivités locales, exploitants privés bénéficiant d'une délégation de service public, ARS).

Une exploration systématique des données par une méthode de détection automatisée des agrégats de cas devrait permettre de repérer rétrospectivement les épisodes épidémiques sur l'ensemble du territoire. La méthode est en cours de développement, la connexion avec

les investigations de terrain doit faire l'objet d'une étude pilote en 2016 dans le cadre d'un groupe de travail rassemblant l'unité Rariec de la DSE, les Cire et les ARS.

1. L'ALGORITHME DE DISCRIMINATION DES CAS DE GEAm

Les données de remboursement des médicaments prescrits permettent un comptage quasi exhaustif des cas de GEAm, c'est-à-dire ayant consulté un médecin et s'étant fait délivrer les médicaments prescrits. Les données du SNIIRAM sont extraites du DCIR (partie du SNIIRAM accessible par Santé publique France). Elles comportent des informations relatives aux prescriptions médicales (les ordonnances), aux prescripteurs et aux bénéficiaires. L'extraction d'une ordonnance a lieu lorsqu'au moins un des médicaments prescrits est un médicament cible utilisé dans le traitement des GEA et pris en charge par la sécurité sociale soit plus de 90 spécialités en 2014. Le tableau 1 donne les principaux médicaments utilisés dans le traitement des GEA, ceux qui sont disponibles dans le SNIIRAM (médicaments utiles) et ceux qui sont utilisés pour l'extraction des ordonnances (médicaments cibles). Les données couvrent l'ensemble du territoire national (métropole et DOM), quasiment tous les régimes de l'Assurance maladie.

I TABLEAU 1 I

Médicaments utilisés dans le soin des gastro entérites aiguës et utilisés dans l'algorithme GEAm

Classe thérapeutique	Classe ATC (1)	Principales spécialités	Principe actif (DCI)
Antispasmodiques intestinaux (2)	(A03A) (A03A) (A03A)	SPASFON DUSPATALIN DEBRIDAT	Phloroglucinol Mébévérine Trimébutine
Anti-émétiques	A04A A03F A03F A03F	VOGALENE MOTILIUM PERIDYS PRIMPERAN	Métopimazine Dompéridone Dompéridone Métoclopramide
Anti-diarrhéiques :			
- Ralentisseurs du transit intestinal	A07D A07D	IMODIUM ARESTAL	Lopéramide Lopéramide
- Anti-sécrétoires	A07X	TIORFAN	Racécadotril
- Adsorbants intestinaux	A07B A07BB	CARBOSYLANE SMECTA	Charbon Diosmectite
Solutés de réhydratation orale	-		-
Anti-infectieux intestinaux (3)	A07A	ERCEFURYL	Nifuroxazide
Produits microbiens antidiarrhéiques (3)	A07F - A07B	ULTRALEVURE LACTEOL CARBOLEVURE	<i>Saccharomyces boulardii</i> Microorg. product. d'acide lactique <i>S. cerevisiae</i>

(1) Entre parenthèse : classe incomplètement incluse

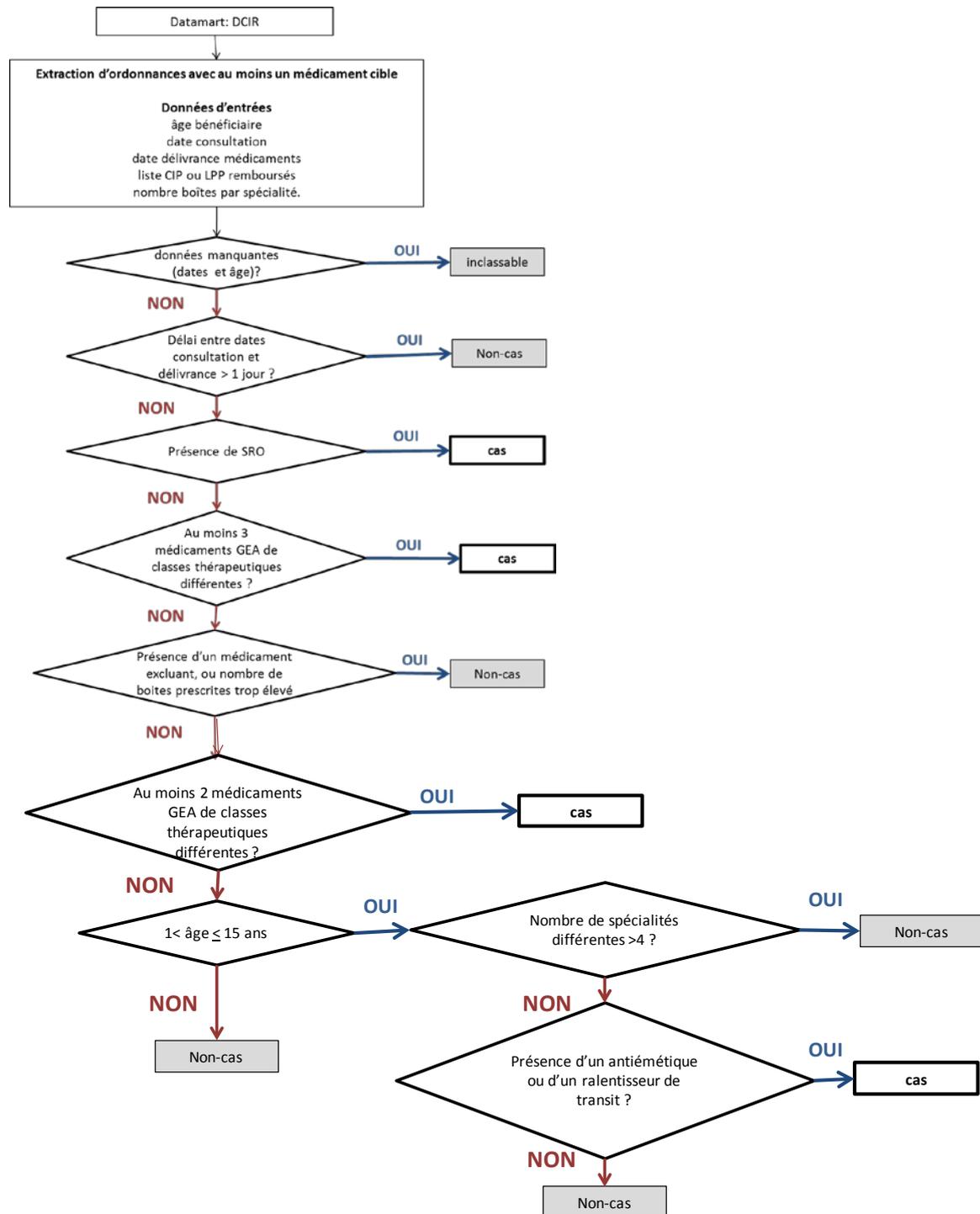
(2) Les antispasmodiques ne sont pas utilisés pour l'extraction mais ils sont utilisés dans l'algorithme

(3) Déremboursés à partir de 2008 et donc absents du SNIIRAM

Un algorithme « GEAm » a été développé [4] afin de classer les ordonnances en cas ou non cas de GEAm à partir de la composition de l'ordonnance (figure 1). Les médicaments « excluants », c'est-à-dire dont la présence sur l'ordonnance aboutit au rejet de l'ordonnance (sauf si 3 médicaments utilisés pour le traitement des GEA ou plus sont présents), sont les antibiotiques, les médicaments anticancéreux, les médicaments du reflux gastro-oesophagiens, de l'acidité gastrique, des rectocolites hémorragiques et maladie de Crohn ainsi que certaines présentations (antiémétiques injectables). Les détails de l'implémentation se trouvent dans la « Procédure » [2]. La construction et la mise à jour de l'algorithme GEAm sont fondées sur l'exploitation d'enquêtes répétées en pharmacie.

I FIGURE 1 I

Algorithme simplifié utilisé pour la discrimination des cas de GEAm



L'algorithme distingue différents types de cas de GEAm (donnée DCIR 2014, France entière) :

- (1) Cas de GEAm dont l'ordonnance comprend un antiémétique et un ralentisseur du transit (33%) ;
- (2) Cas de GEAm dont l'ordonnance comprend un antiémétique mais pas de ralentisseur du transit (27%) ;
- (3) Cas de GEAm dont l'ordonnance comprend un ralentisseur du transit mais pas d'antiémétique (24%) ;
- (4) Cas de GEAm SRO. Il s'agit exclusivement de cas entre 0 et 5 ans car les SRO ne sont pas remboursés au-delà. 11% des cas sont des cas SRO mais seuls les cas de 1 à 5 ans (5%) sont pris en considération (voir section suivante) ;
- (5) Cas de GEAm divers (6%).

1.1 Évaluation de l'algorithme : Valeur prédictive positive (VPP) et sensibilité

La validité de l'algorithme de discrimination des cas de GEAm [4] (figure 1) est examinée annuellement par la faculté de pharmacie de l'Université de Rouen, afin de pouvoir procéder, le cas échéant à sa mise à jour. L'évaluation s'appuie sur la surveillance des prescriptions en pharmacie, opérée par des étudiants lors de leur période de stage en officine. Le recrutement des cas reproduit le processus d'extraction des ordonnances du DCIR (présence d'un médicament cible). En revanche, le diagnostic médical est renseigné par interrogation du client. La définition de référence de la GEA est issue d'une conférence de consensus [1] (encart ci-dessous). Une centaine d'ordonnances avec diagnostic sont ainsi collectées chaque année depuis l'an 2000 qui permettent d'adapter l'algorithme de discrimination des cas et d'évaluer ses performances. Les changements thérapeutiques, commerciaux ou de prise en charge sont suivis, ainsi que leur effet sur les performances de l'indicateur. L'algorithme GEAm a été ainsi modifié en 2008, et pourra encore l'être si nécessaire.

Les déremboursements de médicaments sont les facteurs les plus importants de changement des prescriptions de médicament. Les listes de médicaments nécessaires au déroulement de l'algorithme sont mises à jour annuellement (ajout de spécialités avec possibilité de reconstitution rétroactive de l'entrepôt). La sensibilité de l'algorithme actuel a été évaluée à 0,90 (IC95=[0,87;0,92]) et la VPP à 0,83 (IC95=[0,80;0,85]) sur les données 2012-2014 (N=1208) [5]. Cette étude montre en outre qu'il n'y pas de différences significatives de sensibilité ou de VPP entre les 0-14 ans (sensibilité = 0,93 et VPP = 0,85, N = 394) et les ≥ 15 ans (sensibilité = 0,87 et VPP = 0,81, N = 814). Ces estimations sont fondées sur les données qui ont servi à construire l'algorithme GEAm et donc probablement légèrement surévaluées. La collecte des données ne couvre d'autre part que le premier semestre de l'année (calendrier des stages en officine) ce qui peut induire un biais limité sur les résultats (la période épidémique hivernale et la période non épidémique étant toutes les 2 représentées dans l'échantillon, avec toutefois une légère surreprésentation de l'épidémie hivernale).

Gastro-entérite aiguë (GEA) : L'« *International Collaboration on Enteric Disease Burden of Illness* » définit un cas de GEA par l'émission de 3 selles liquides ou plus au cours d'une période de 24 heures ou d'au moins un vomissement. Les cas d'exclusion sont les gastro-entérites chroniques, l'indigestion, l'ingestion excessive d'alcool, les symptômes liés à la grossesse, aux règles ou à certains traitements médicamenteux. Un intervalle de 7 jours sans symptômes est requis pour considérer deux épisodes indépendants [1].

Des définitions plus anciennes (non retenues ici) font intervenir la durée de l'épisode de diarrhée (au minimum deux [6] ou trois jours [7]) mais la prise précoce d'anti-diarrhéiques par certains malades engendrait autant de faux négatifs. En l'absence d'autres mentions, cette définition symptomatique est utilisée par la suite pour la GEA.

Diarrhée : l'émission d'une selle liquide ou plus au cours d'une période de 24 heures.

Diarrhée aiguë : l'émission de 3 selles liquides ou plus au cours d'une période de 24 heures.

Il existe cependant des populations particulières pour lesquelles les performances de l'algorithme GEAm sont moindres.

La VPP de l'indicateur pour les **enfants de moins d'un an** (cas SRO) est altérée par le fait que le praticien peut prescrire de façon systématique des SRO au sortir de la maternité ou lors de la première consultation en pédiatrie de ville.

La sensibilité de l'algorithme est inférieure pour les **personnes âgées** par rapport à la population générale. Les ordonnances destinées aux personnes âgées visent souvent aux traitements de plusieurs pathologies chroniques et comportent plus de médicaments prescrits que celles destinées au reste de la population. Le choix a été fait de privilégier la spécificité (approchée ici par la VPP) au détriment de la sensibilité, par exemple en rejetant les ordonnances où figurent plus de 4 médicaments.

1.2 Validation croisée de l'algorithme

Il n'existe pas de « gold standard » pour évaluer les estimations d'incidence des GEAm fournies par l'application de l'algorithme aux données du SNIIRAM. Plusieurs sources externes ont été utilisées pour la comparaison :

- 1) Le Réseau sentinelles : on n'observe aucune différence entre la dynamique d'incidence des 2 indicateurs au niveau national [3]. L'incidence moyenne des cas de GEAm était de 0,106 par personne-année en 2009-2010 en utilisant les données fournies par la Cnamts, contre 0,094 en utilisant les données du Réseau sentinelles. Une définition de cas plus restrictive de la part du Réseau Sentinelles pourrait expliquer cette différence (diarrhée aiguë vs. GEA).

L'enquête Ira-GEA sur l'incidence des infections respiratoires aiguës et des GEA et leur prise en charge conclut au même taux d'incidence des GEAm qu'avec les données fournies par la Cnamts (0,11 épisodes par année et par personne) [3]. Il existe cependant une différence d'incidence par classe d'âge, avec une sur-représentation apparente des enfants (1-15 ans : $T_{I_{CnamGastro}}/T_{I_{Ira-GEA}} = 1,9$) et une

- 2) sous-représentation des adultes (16-64 ans : $TI_{\text{CnamGastro}}/TI_{\text{Ira-GEA}} = 0,41$) avec les données de la Cnam. Le ratio est de 1 pour les plus de 65 ans. Les différences n'ont pas été éclaircies.
- 3) Des rapports d'investigation d'épidémies. Une seule comparaison a été publiée [8] montrant l'aptitude des données de la Cnamts à reproduire la courbe épidémique des GEAm mise en évidence par une étude cohorte rétrospective. D'autres expériences de comparaison avec les résultats de cohorte sont en cours (2 épidémies, Mouly, publication en cours) avec des résultats mitigés. La comparaison quantitative est potentiellement contrariée par un biais de recrutement en faveur de la surreprésentation des cas pour les questionnaires auto-administrés. Les GEAm virales, souvent associées à des symptômes bénins et de courte durée, sont moins médicalisées et moins traçables dans les données de la Cnamts.

2. L'ENTREPÔT DES RESUMÉS DE CAS DE GEAm

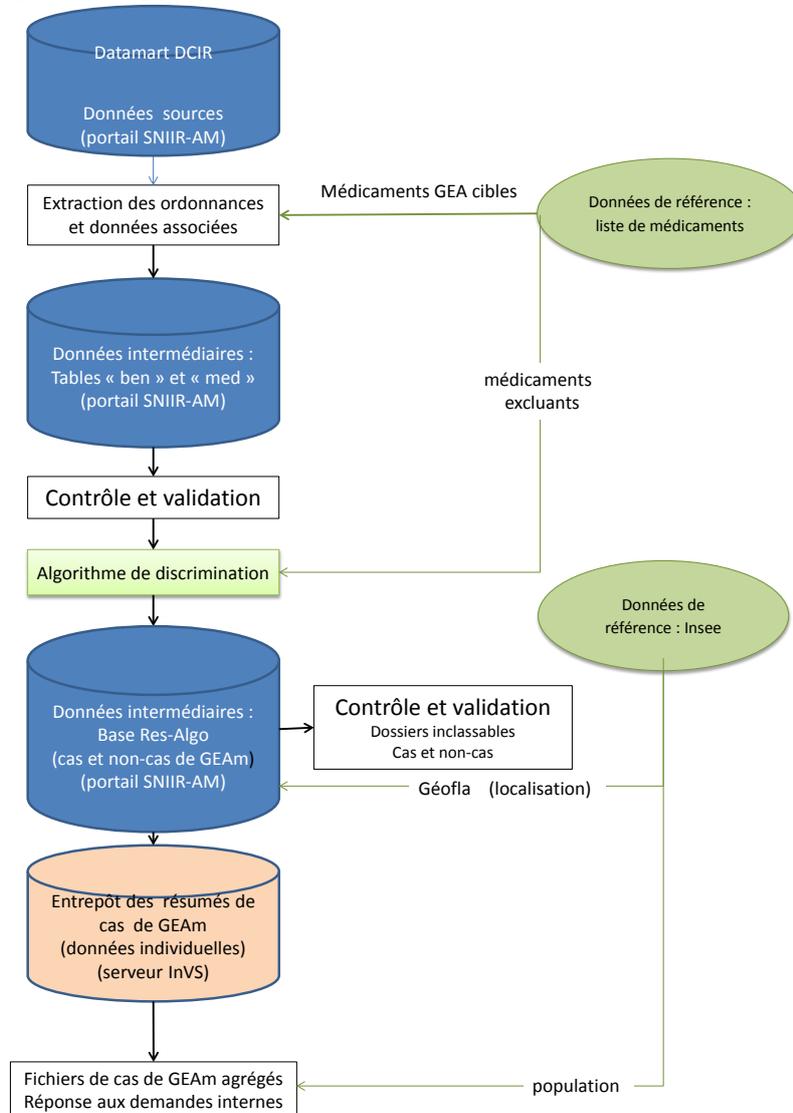
2.1 Contenu*

Une fois l'algorithme appliqué aux dossiers de remboursement des médicaments du SNIIRAM, les résumés des cas sont entreposés dans « l'entrepôt des cas individuels » (figure 2).

I FIGURE 2 I

Schéma général de l'exploitation du SNIIRAM (DCIR) pour la constitution de l'entrepôt des cas de GEAm

Géofla : base de données sur les communes incluant les coordonnées (x,y).



L'entrepôt des cas de GEAm (tableau 2) permet de générer les fichiers de données utilisateurs. Il contient des informations quantitatives (âge, distance entre les communes de résidence du cas de GEAm et d'exercice du praticien) qui permettent de stratifier les comptes de cas en fonction de catégories définies à partir de ces variables.

Il contient aussi des informations réservées à des usages particuliers (prescription d'une coproculture). Le « type de cas » (détail en section « L'algorithme ») et le « régime d'assurance maladie » sont des informations destinées au contrôle de l'homogénéité de la définition de cas dans le long terme. La conservation de ces informations dans l'entrepôt des cas individuels permet en effet d'opérer l'exclusion rétrospective des catégories dont le sens a varié au cours de la période d'étude.

I TABLEAU 2 I

Contenu de l'entrepôt des cas de GEAm

Variable	Libellé	Modalités	Origine
ID	Numéro d'ordre (non identifiant)		Calcul : date de flux (année-mois) + compteur
Date	Date de prescription de l'ordonnance	Date	DCIR
Âge	Âge en année au moment de la rescription	Réel	Calcul à partir de l'année et du mois de naissance et de la date de prescription, arrondi à une décimale pour les 0-24 mois, à l'entier sinon
Sexe	Sexe du bénéficiaire	1 = Homme 2 = Femme	DCIR
B_Rg1	Régime d'assurance maladie	1 = Régime général' 2 = Régime agricole' 3 = Régime social des indépendants' 8 = Régime des militaires de carrière 10 = Régime clerks/employés de notaires	Recodage
Dept_B	Département de résidence du bénéficiaire	3 car.	Recodage
NC_b	Code commune de résidence du bénéficiaire	3 car.	Recodage
Insee_dr_ph	Code Insee département et commune du médecin ou, à défaut, de la pharmacie	5 car.	Recodage
Qua_dr_ph	Type praticien associé au code commune praticien	1 = Médecin 2 = Pharmacien	Calcul
Cas_type	Type de cas de GEAm en fonction des médicaments présents sur l'ordonnance	1 = Antiémétique' 2 = Antiémétique et Ralentisseur de transit' 3 = Antiseptique' 4 = Ralentisseur de transit' 5 = SRO'	Calcul
B_Cop	Prescription d'une coproculture	1 = Oui 2 = Non	Recodage du code Biologie 5207
Distance	Distance « commune bénéficiaire-commune praticien ». Le praticien est le médecin, ou bien le pharmacien si le code commune du médecin est manquant.	Entier (km)	
Résident		1 : Résident : distance en bénéficiaire et médecin/pharmacien < 50 km 2 : Touriste : distance en bénéficiaire et médecin/pharmacien ≥ 50 km 3 : Uniquement la localisation du bénéficiaire manquante 4 : Localisation du médecin/pharmacien manquant 5 : Localisation du bénéficiaire ET du médecin/pharmacien manquant	Construite

2.2 Contrôle qualité des données de l'entrepôt

Les données de l'entrepôt sont validées selon une routine mise en œuvre à chaque extraction de données du DCIR pour une mise à jour de l'entrepôt. La validation est décrite dans la procédure de constitution de l'entrepôt [2]. Elle vise d'une part à identifier et à quantifier les données manquantes et d'autre part à détecter des changements anormaux d'incidence des différentes catégories de cas et de non cas, qu'on peut suspecter d'être en lien avec des erreurs de manipulation des données ou des changements de pratiques. Les anomalies repérées sur cette base sont investiguées et si possible corrigées. Elle font l'objet d'un rapport d'anomalie.

3. QUALITÉ DES DONNÉES : LIMITES LIÉES À LA SÉLECTION DES CAS DE GEA PAR LE RECOURS AU SOIN*

Van Cauteren *et al.* [9] estiment dans une étude en population générale qu'en France 33% des cas GEA consultent un médecin et que 32% des cas GEA consultent un médecin et se font délivrer les médicaments prescrits par le médecin. **La sensibilité moyenne de l'algorithme par rapport aux GEA cliniques s'élève donc à (32% X sensibilité de l'algorithme par rapport aux GEAm), soit >25%**. Seulement un pourcent des cas de GEA (au sens clinique) consultent donc un médecin et ne se voient délivrer aucun médicament utilisé pour le traitement des GEA (situation rarissime en France, selon les médecins) ou bien ne vont pas retirer les médicaments (plus probablement).

Le ratio de médicalisation de la GEA est cependant très variable selon l'**agent pathogène** responsable et l'âge du patient. Une enquête anglaise Wheeler [10] fournit un point de comparaison pour jauger la distorsion de la partition des cas de GEA par agent pathogène causal lié au fait de consulter ou non. Certaines bactéries apparaissent 3 fois plus souvent parmi les étiologies quand il y a consultation (*Salmonella non typhi*, *Shigella*), d'autres agents 2 fois plus (les bactéries *Campylobacter* sp., *Clostridium perfringens*, et les parasites (*Cryptosporidium* sp. et *Giardia* sp.), tandis que *Yersinia* était 3 fois moins représentée, et *Clostridium difficile*, *Aeromonas* sp. et les norovirus 2 fois moins. Pour les autres virus (rotavirus, astrovirus et adenovirus groupe F) et certaines bactéries (souche non O157 entéroinvasives, entérotoxigènes ou entéropathogènes) les auteurs observaient une fréquence similaires d'apparition parmi les cas de GEA et de GEAm¹.

Ces observations sont cohérentes avec les taux de consultation observés durant les épidémies investiguées en France. On a ainsi observé des taux de 80% lors d'une épidémie de cryptosporidiose parmi la population des quartiers les plus exposés de Dracy-le-Fort (71) [11]. A l'opposé, lors d'une épidémie probablement virale [12] survenue parmi les clients d'un camping, seuls 10% des cas avaient consulté. **Appliquer le ratio moyen de 32% pour estimer le nombre de cas clinique à partir du nombre de GEAm n'a pas de sens dans le cas d'épidémies**, dans la mesure où les épidémies sont le plus souvent causées par un agent pathogène unique, associé à une probabilité de consultation spécifique [11]. Les inférences (nombre de cas de GEA = nombres de cas de GEAm/0,32) sont en revanche acceptables pour des périodes d'une année ou d'un nombre entier d'années pour tenir compte des importantes variations saisonnières affectant la circulation des agents pathogènes.

La sélection se fait en faveur des agents causaux provoquant les symptômes les plus graves et durables. Ainsi une enquête française montre qu'une durée de plus de 3 jours favorise significativement la consultation du médecin (OR = 4), ainsi que la présence de vomissement, de maux de tête ou de fièvre (OR d'environ 2 pour chacun des symptômes). Ainsi, si le choix de la surveillance syndromique impose comme objet les GEAm plutôt que les GEA, l'effet de sélection de doit pas être seulement vu comme une perte de sensibilité mais aussi comme un focus sur les cas les plus sérieux, ce qui représente en soi un intérêt de santé publique.

Le taux de médicalisation peut aussi baisser dans des circonstances où l'**accès au médecin** est difficile. C'est l'interprétation qui prévaut de l'apparent effet protecteur de la tempête Klaus vis-à-vis des GEA [13], alors que la distribution de l'eau était défaillante et les routes

¹ Tableau de synthèse en p. 131 de <https://tel.archives-ouvertes.fr/tel-00795215/document>

coupées par les chutes d'arbres. Des recherches sont prévues pour élaborer un indicateur. Une étiologie à faible taux de médicalisation et une restriction d'accès peuvent se conjuguer. Il a été ainsi observé à Plaux (2012) [14] une épidémie virale débutant à la veille du week-end du 15 août. Le très faible taux de consultation observé parmi les cas pourrait en partie provenir de leur guérison avant la reprise d'une activité médicale normale.

4. QUALITÉ DES DONNÉES : LIMITES LIÉES À L'ALGORITHME ET À SON IMPLÉMENTATION*

4.1 Manque de spécificité de l'algorithme pour certaines classes d'âge

La spécificité (approchée par la VPP) et la sensibilité de l'algorithme de discrimination des cas de GEAm sont décrites en § « Évaluation de l'algorithme : VPP et sensibilité ».

Compte tenu de la mauvaise spécificité possible de l'algorithme pour **les enfants de moins d'un an**, nous préconisons de **ne pas les inclure dans les études**. Les données sont présentes dans l'entrepôt mais exclues des analyses « standard ». Concernant les études centrées sur le rôle de l'eau, d'autres arguments s'ajoutent pour renoncer à l'inclusion des enfants de moins d'un an : (i) ceux-ci bénéficient d'une immunité forte héritée de la mère et (ii) ne boivent généralement pas d'eau du robinet.

On sera enfin prudent sur l'interprétation d'une sur-incidence des prescriptions pédiatriques pendant une canicule. Le remboursement des prescriptions de SRO (cas SRO), spécifique aux enfants de moins de 5 ans, peuvent relever du traitement ou de la prévention de la déshydratation pouvant résulter de la chaleur, d'une diarrhée ou de la combinaison des deux causes.

La présence fréquente de nombreux médicaments sur les ordonnances prescrites aux personnes âgées rend incertaine l'imputation de la consultation à une GEA. L'exclusion des ordonnances pléthoriques dans le déroulé de l'algorithme vise à favoriser la spécificité de l'algorithme dans cette classe d'âge. La comparaison avec d'autres sources de données (Réseau Sentinelles, enquête Ira-GEA voir § Validation croisée de l'algorithme) n'indique cependant pas de perte de sensibilité significative.

4.2 Consultations multiples

Le nombre moyen de consultations par épisode de GEAm est 1,1 (médiane = 1), c'est-à-dire que 9% (IC95 = [5% ; 13%]) des cas de GEAm sont l'objet de plusieurs consultations [3]. Dans le cas général, les pratiques habituelles de prise en charge de la GEA comprennent en première intention un traitement à visée symptomatique et en deuxième intention la prescription d'une coproculture, puis éventuellement d'un antibiotique adapté. Sauf trithérapie², la présence d'antibiotique sur l'ordonnance est utilisée dans l'algorithme comme critère d'exclusion, sachant qu'il y a beaucoup moins d'ordonnances correspondant à une antibiothérapie de GEA que d'ordonnances comportant la prescription (i) d'un antibiotique non destiné au traitement d'une GEAm et (ii) d'un antidiarrhéique destiné à traiter d'éventuels effets secondaires de l'antibiothérapie. Les prescriptions de deuxième intention pour un même épisode de GEA sont donc en partie exclues et les cas correspondant ne sont alors comptés qu'une fois, en cas d'exclusion. La part résiduelle des consultations multiples n'a pas été quantifiée.

4.3 Renouvellements d'ordonnance

Les **renouvellements** d'ordonnances constituent une pratique incompatible avec le diagnostic de GEA. Ils sont repérables d'une part par comparaison de la date de prescription

² Trithérapie : présence sur l'ordonnance de 3 classes thérapeutiques utilisées dans le traitement des GEA (tableau 1).

et de délivrance, d'autre part par le contenu de l'ordonnance qui ne correspond pas au traitement d'une pathologie aiguë. La première délivrance n'est cependant pas identifiable sur le premier critère et ne peut être exclue que sur le deuxième. Une étude interne a montré que le délai entre le 1^{er} passage et le 2^e passage en pharmacie pour un renouvellement d'ordonnance est de moins d'un, deux ou trois mois pour respectivement 50, 90 et 97% des ordonnances renouvelées. La procédure de mise à jour de l'entrepôt comprend une étape consacrée à la recherche des délivrances ordonnances renouvelées : 1) repérage et exclusions des délivrances de rang >1 sur le critère des dates, 2) recherche dans l'historique des délivrances correspondant à la même prescription (c'est-à-dire au même numéro de dossier) et exclusion. Pour que le processus soit quasi exhaustif, 4 mois consécutifs de données sont pris en considération.

5. QUALITÉ DES DONNÉES : LIMITES LIÉES À LA SOURCE DE DONNÉES

5.1 Les données manquantes ou erronées et leur traitement : généralités*

L'Assurance maladie estime que le SNIIRAM couvre plus de 99% des résidents en France et le DCIR 98% (des régimes marginaux sont retirés). Les données manquantes (DM) sont peu fréquentes et les calculs d'incidence parmi les résidents peuvent donc utiliser comme dénominateur les données démographiques communales de l'Insee dans les applications courantes.

On distingue les dossiers absents dans leur globalité et les dossiers présents dans le SNIIRAM mais comportant des champs manquants ou erronés. Dans ce dernier cas, les données manquantes ou erronées dans le DCIR peuvent empêcher le classement en cas ou en non cas, ou encore leur géolocalisation à la commune, et donc leur classement en résident ou touriste.

Les DM se partagent aussi entre les DM qui affectent l'ensemble du territoire national (par exemple personnes âgées en Ehpad) et les DM concernant le département (liées par exemple aux pratiques des CPAM³). Les investigations d'épidémies qui procèdent de comparaison entre communes au sein d'un même département sont ainsi peu perturbées : les taux d'incidence et les impacts peuvent être sous-estimés, mais les estimations de risque relatif fondées sur des comparaisons de taux d'incidence entre communes restent non biaisées.

Le redressement des données par département est en revanche à envisager si on s'intéresse au risque endémique et si le secteur d'étude englobe plusieurs départements. Le seul problème identifié comme susceptible d'entraîner des biais non négligeables dans les conclusions est alors celui de la géolocalisation des cas.

Des informations sur les DM sont colligées en routine qui permettent éventuellement de procéder (i) à l'exclusion éventuelle d'une catégorie de données qui concentre les DM (régime d'assurance maladie), ou tel mois ou tel département (problème en lien avec l'activité d'une CPAM) ou (ii) au redressement éventuel des données.

5.2 Cas de GEAm absents du SNIIRAM*

En 2009, 793 000 personnes âgées (9,4% des ≥ 75 ans) vivaient en établissements d'hébergement pour personnes âgées dépendantes (Ehpad), c'est-à-dire maisons de retraite ou établissements à caractère sanitaire ou social de moyen ou de long séjour (tableau 3). La moitié des **personnes en Ehpad** bénéficient d'une prise en charge médicale par l'établissement (400 000 personnes soit 0,6% de la population française et 5% des ≥ 75 ans) tandis que l'autre moitié conserve une relation individuelle directe avec leur généraliste. Seule cette dernière moitié se retrouve dans le décompte des cas de GEAm tiré du DCIR. La couverture des personnes âgées par le SNIIRAM est sensiblement affectée à partir de 80 ans (96%) et atteint 74% à 100 ans.

³ 102 CPAM en France, une CPAM correspondant en général à un département.

La classe des **15-24 ans** est aussi sujette à la vie en communauté mais cette classe d'âge est plus généralement celle dont la domiciliation est la plus méconnue (voir § Cas de GEAm dont la géolocalisation est erronée).

I TABLEAU 3 I

Fraction de la population vivant en communauté par tranche d'âge (2009, Insee, http://www.insee.fr/fr/themes/document.asp?ref_id=ip1434)

Tranches d'âge	Femmes	Hommes	Ensemble
de 0 à 4 ans	0.2%	0.2%	0.2%
de 5 à 9 ans	0.3%	0.4%	0.3%
de 10 à 14 ans	1.2%	2.0%	1.6%
de 15 à 19 ans	8.1%	10.2%	9.2%
de 20 à 24 ans	3.0%	5.7%	4.3%
de 25 à 29 ans	1.0%	3.0%	2.0%
de 30 à 34 ans	0.6%	1.9%	1.2%
de 35 à 39 ans	0.5%	1.5%	1.0%
de 40 à 44 ans	0.5%	1.4%	0.9%
de 45 à 49 ans	0.5%	1.4%	0.9%
de 50 à 54 ans	0.5%	1.3%	0.9%
de 55 à 59 ans	0.6%	1.4%	1.0%
de 60 à 64 ans	0.7%	1.7%	1.2%
de 65 à 69 ans	1.0%	1.9%	1.5%
de 70 à 74 ans	1.7%	2.3%	2.0%
de 75 à 79 ans	3.5%	3.1%	3.3%
de 80 à 84 ans	8.3%	5.3%	7.2%
de 85 à 89 ans	17.9%	10.6%	15.5%
de 90 à 94 ans	32.7%	19.6%	29.3%
de 95 à 99 ans	46.4%	30.5%	43.4%
100 ans et plus	55.1%	32.2%	51.7%
Ensemble	2.4%	2.7%	2.5%

5.2 Dossiers présentant des champs manquants

5.2.1 DM empêchant la catégorisation en cas/non cas de GEAm

Les données manquantes qui empêchent le déroulement de l'algorithme GEAm sont l'année de naissance du bénéficiaire, la date de prescription ou la date de délivrance des médicaments. Leur taux était de 0,2% en 2014 et concernait quasi uniquement l'année de naissance du bénéficiaire du RSI (7.2% des ordonnances du RSI ; le RSI couvrant 4% de la population).

L'ordonnance qui ne peut être classée en cas ou non cas n'apparaît pas dans l'entrepôt des cas. Cependant, les statistiques de DM correspondantes sont conservées par semestre et par département. En supposant que le taux de DM n'affectent pas plus les cas que les non cas, les comptes de cas peuvent être ainsi redressés. Le redressement peut être envisagé si le secteur d'étude inclut plusieurs départements et si l'étude porte sur le risque endémique.

5.2.2 DM empêchant la localisation des cas de GEAm*

Le principal problème identifié comme susceptible d'entraîner des biais non négligeables dans les conclusions est celui de la géolocalisation des cas (commune Insee bénéficiaire et prescripteurs ; tableau 4) avec des proportions de DM qui peuvent varier de 2 à 10% entre départements. Une diminution importante de la fréquence de ces DM est enregistrée en 2011 avec l'accès aux données des communes des pharmaciens. Une légère diminution tendancielle est ensuite enregistrée. L'absence de localisation des cas résulte de l'absence de l'information (DM) mais également d'un code qui ne correspond pas à une commune dans la base Géofla (respectivement 2,4% et 3,2% pour le code bénéficiaire en 2014).

I TABLEAU 4 I

Fraction des cas de GEAm présentant un problème de géolocalisation (code Insee de la commune manquant ou erroné, DCIR)

% données manquantes							Détail par régime (2014)			
	2010	2011	2012	2013	2014	2015	RG	MSA	RSI	Autres régimes ^a
Bénéficiaire	4,4	3,7	2,6	2,4	2,4	1,7	2,4	<0,01	0,1	30,9
Médecin prescripteur	12,8	6,1	6,3	6,6	7	7,6	7,6	5,1	6,9	11,1
Pharmacien dispensateur	ND	<0,01	0	0	0	0	0	0	0	0
Non classable résident/touriste	17,3	4,6	3,3	3,2	3,1	2,2	3,2	0,2	0,6	31,8

ND : Non disponible

^aAutres régimes = Régimes autres que RG, MSA et RSI (soit 7% de la population)

Les informations disponibles pour la géolocalisation du bénéficiaire et du praticien conditionnent le traitement du cas de GEAm. L'objectif du traitement des cas mal géolocalisés (tableau 5) est d'une part de trouver un compromis entre la perte d'information en cas d'exclusion et le risque d'erreur de classement sur l'exposition en cas d'inclusion avec imputation incertaine, d'autre part d'identifier, en cas d'exclusion, les possibilités de redressement des comptes de cas communaux pour disposer d'une estimation non biaisée de l'incidence. Il faut prévoir, le cas échéant, de produire les comptes de cas exclus nécessaires au redressement.

Nous prendrons ici comme exemple les études à caractère local ciblant le facteur hydrique (tableau 5) ; les données sont analysées à la commune et le secteur d'étude est infra-départemental ou bien couvre plusieurs départements. Les deux catégories de cas d'intérêt sont les résidents présents à leur lieu de résidence (RésPré) et les touristes (Tou). Quand toute l'information est disponible (commune du bénéficiaire et du praticien connue), le cas peut être localisé et imputé sans ambiguïté soit aux touristes soit aux résidents présents sans ambiguïté (cellule jaune du tableau 5). À l'opposé, quand les départements du bénéficiaire et du praticien sont inconnus, alors on ne peut pas localiser le cas ni le classer en résident présent ou touristes. Le tableau 5 indique en fonction de la précision des localisations du bénéficiaire et du praticien, les possibilités de classement catégoriel des cas, de localisation probabiliste du cas, et de redressement des comptes de cas communaux.

Si l'information est incomplète on peut raisonnablement imputer le cas aux résidents présents si le nombre de résidents dépasse localement *a priori* le nombre de touristes, en acceptant toutefois quelques erreurs de classement sur l'exposition. Le choix de l'imputation dépend du taux d'erreurs, c'est à dire du caractère touristique du lieu et de la période, les zones et les périodes touristiques favorisant les erreurs. Il existe une exception au classement par défaut en résident, si l'information de localisation suffit à conclure que le lieu de consultation (ou d'achat des médicaments) sont éloignés du lieu de résidence (départements disjoints par exemple). Dans un seul cas on peut raisonnablement imputer le cas aux touristes et le localiser à la commune (avec possibilité d'erreur résiduelle sur la localisation : cf. infra) : commune de consultation connue, département de résidence du bénéficiaire éloigné.

La situation la plus préoccupante est celle où la commune du bénéficiaire n'est pas connue. La géolocalisation du praticien à la commune permet éventuellement d'imputer le cas à la commune du praticien, en considérant le cas comme un cas résident présent. Cette imputation engendre un risque d'erreur de classement, si le bénéficiaire consulte (ou achète ses médicaments) en dehors de sa commune. Ce risque est plus élevé (i) en secteur rural (plus la taille de la commune de résidence est petite plus le risque est élevé), (ii) si on ne dispose que de la localisation de la commune du pharmacien (les pharmacies sont moins nombreuses que les médecins généralistes), et (iii) si le lieu ou la période d'incidence du cas sont touristiques. La situation doit donc être appréciée localement.

Si la proportion de DM sur la localisation est élevée pour un département donné (problème lié à une CPAM) et que le secteur d'étude s'étend sur plusieurs départements, il est utile de se servir des statistiques sur les DM_localisation pour redresser les comptes de cas communaux en fonction de leur département d'appartenance et limiter ainsi les biais différentiels entre les départements (par exemple en adoptant une correction au mois : $n_{c_{i,j,t}} = Tx_DM_{j,t} \times n_{i,j,t}$ où $n_{i,j,t}$ et $n_{c_{i,j,t}}$ sont respectivement les comptes brut et redressé pour la commune i du département j au mois t et $Tx_DM_{j,t}$ le taux de DM observé sur le département j au mois t).

I TABLEAU 5 I

Traitement du cas de GEAm mal géolocalisé dans les études à caractère local en fonction de l'information disponible pour la géolocalisation du bénéficiaire et du praticien de santé

Codes disponibles		Bénéficiaire (B)		
		Commune (C)	Département (D)	Rien
Praticien (P)	Commune (C)	RésPré ¹ , Tou	RésPré ³ , Tou	RésPré ⁴ , Tou
	Département (D)	RésPré ¹ , Tou	RésPré, Tou (R)	RésPré, Tou
	Rien	RésPré ² , Tou	RésPré, Tou (R)	RésPré, Tou

RésPré : résidents présents, Tou : touristes.

Non barré : classable et localisable à la commune

Barré : classable mais non localisable à la commune

Double barré : non classable et non localisable à la commune

La légende (tableau 5) indique les risques d'erreur de classement

	Préconisation	Erreur de classement
1	Si D_P = D_B ou si C_B proche de D_P, imputation possible des cas à la commune du bénéficiaire Classement : résident présent	Imputation des touristes à l'intérieur du département aux résidents présents (tx d'erreur << 4%, sauf fréquentation élevée d'un lieu de villégiature dans le même département, ex. Arcachon pour Bordeaux)
2	Imputation possible des cas à la commune du bénéficiaire Classement : résident présent	4% d'erreur en moyenne, avec de fortes variations selon la saison et le lieu (erreur max. dans les lieux touristique et pendant les vacances scolaires)
3	Si D_P = D_B ou si C_P proche de D_B, Imputation possible des cas, à la commune du praticien Classement : résident présent OU BIEN touriste si D_B éloigné de C_P	Taux d'erreur inversement liée à la taille de la commune (erreur si la commune de résidence ou de séjour n'a pas de praticien)
4	Si D_P = D_B, Imputation des cas à la commune du praticien Classement : résident présent	
R	Redressement. Si le bénéficiaire est localisé au département et le praticien n'est pas localisé à la commune (cellules (R) du tableau), alors le cas ne peut pas être imputé à une commune. La proportion p de ces cas au niveau départemental permet de redresser les comptes de cas communaux du département. Le redressement est préconisé (i) quand le secteur d'étude couvre plusieurs départements, (ii) quand p diffère sensiblement entre les départements concernés et (iii) selon l'objectif de l'étude, la détection de clusters étant moins exigeante que l'étude du bruit de fond.	

5.2.3 Cas de GEAm dont la géolocalisation est erronée*

La localisation de la résidence des **jeunes adultes** (16-30 ans) pose problème, une certaine proportion d'étudiants (ou de jeunes travailleurs précaires) négligeant de procéder à une information mise à jour concernant leurs changements d'adresse ou préférant fournir à l'Assurance maladie l'adresse de leurs parents. Ce problème n'est pas évalué.

D'autre part, les étudiants recourent à des sections locales mutualistes (SLM, par exemple les mutuelles des étudiants). Ces SLM prennent en charge le remboursement non seulement de la part mutuelle mais aussi de la part sécurité sociale. Dans ces conditions, elles se dispensaient de fournir jusqu'en 2005 l'adresse de leurs clients à l'Assurance maladie. Ce problème a été résolu progressivement et ne concerne pas les données de l'entrepôt.

6 DÉLAIS DE CONSOLIDATION ET DE MISE À DISPOSITION DES DONNÉES*

Les différents régimes d'assurance maladie n'alimentent pas le SNIIRAM à égale fréquence. Le régime général abonde le SNIIRAM à 72 heures pour ce qui concerne les dossiers transmis par voie électronique par les pharmacies. Les feuilles « maladie » sont saisies en quasi-totalité dans les 2 mois après leur transmission. Elles ne sont cependant périmées qu'au bout de 27 mois et les comptes de cas peuvent donc théoriquement évoluer dans les 3 années qui suivent un événement à investiguer, suite à l'envoi tardif de feuilles maladie. Les données des autres régimes de liquidation sont intégrées mensuellement au SNIIRAM.

Le DCIR est la partie du SNIIRAM directement accessible par Santé publique France. Le DCIR est normalement actualisé chaque mois (généralement en milieu de mois pour le RG, fin de mois pour les autres régimes) pour les feuilles de soins non rejetées par les processus qualité du SNIIRAM. Chaque régime émetteur est responsable de la qualité des données transmises. La dynamique d'intégration des données a été étudiée (tableau 6). La complétude d'un jeu de données mensuel se situe autour de **80% un mois** après le terme du mois étudié, **>98% deux mois** après et avoisine 100% à 6 mois (tableau 6). Aucune différence de dynamique d'intégration des données n'est observée entre les médicaments testés. **On peut donc utiliser les données à un mois pour confirmer une épidémie**, produire une première évaluation de l'impact et mettre en œuvre les investigations de terrain pour conforter l'origine et les circonstances. À deux mois les données peuvent être considérées comme exhaustives pour l'investigation d'épidémies. On considèrera que le jeu de données est définitif à 4 mois quel que soit l'usage.

Les statistiques extraites pour une période donnée évoluent cependant marginalement sur l'année voir 2 ans après cette période, pour prendre en compte les dossiers intégrés tardivement dans le SNIIRAM (feuilles de soins « papier »). Il est donc normal qu'une requête répétée à l'identique ne reproduise pas exactement le jeu de données initial.

I TABLEAU 6 I

Complétude (%) des données selon le délai la date de prescription et la date de requête

Pour chaque médicament, 12 dates de requête ont été simulées, en se fondant sur des dates de flux antérieures au premier de chaque mois de l'année 2012. Spécialité : TIORFAN 100MG Gélule 20, 1 845 624 boîtes remboursées⁴.

	à 1 mois (1)	à 2 mois	à 3 mois	à 6 mois	à 12 mois
Moyenne sur les 12 mois	81.3	99.0	99.4	99.8	100.0
minimum	(2) 71.5	98.7	99.3	99.8	99.9
maximum	84.7	99.1	99.5	99.8	100.0

(1) « À un mois » signifie « un mois après la fin du mois étudié »

(2) Valeur accidentelle. Valeur minimale en second : 78,9

⁴ La dynamique d'intégration des données a été étudiée pour un médicament cible en présentation adulte (TIORFAN 100MG GELULE 20) et pédiatrique (TIORFAN 30MG ENF PDR ORALE SACHET 30/3 G) et pour un antibiotique (AMOXICILLINE BIOGARAN 500MG GELULE 12 et AMOXICILLINE MYLAN 500MG). Ces médicaments ont été choisis car ils sont prescrits très largement (entre un et deux millions de boîtes vendues par année en France pour ces 4 spécialités) et car ils sont spécifiques du traitement de pathologies aiguës (sauf TIORFAN 100MG, aussi utilisé pour le soin de pathologies chroniques).

7 PRÉCONISATION D'UTILISATION DES DONNÉES**

7.1 Généralités**

La répartition par **âge** de la population est une donnée clé pour standardiser et comparer l'incidence des GEAm entre différents lieux, époques ou événements. Le ratio enfants/adultes observé parmi des cas rattachés à un agrégat peut aussi permettre de distinguer les Tiac survenant en restauration collective (maison de retraite, école) des épidémies portées par l'eau. Cette information pourrait aussi orienter la recherche de l'agent pathogène (ex. les enfants sont plus sensibles au rotavirus), mais ce sujet n'est pas évalué.

La **standardisation sur l'âge** en 2 classes est préconisée dès qu'il s'agit de comparer spatialement des taux d'incidence. Elle est indispensable pour la comparaison de taux d'incidence entre eux, notamment pour l'étude de signaux faibles (non épidémiques) et quand l'unité géographique d'étude est de taille réduite (commune). Elle peut se limiter à la prise en compte de 2 classes d'âge : enfants (1-15 ans) et adultes (16 ans et plus). D'une part les taux d'incidence sont très différents entre adultes et enfants (tableau 7) : l'incidence des cas chez les adultes (16 ans et plus) est en moyenne 4 fois inférieure à celle observée chez les enfants (1-15 ans). Le ratio $TI_{\text{enfants}} / TI_{\text{adultes}}$ des GEAm varie en fonction de la saison avec un maximum durant l'épidémie hivernale (de l'ordre de 10) et suivant les années en fonction de l'importance de l'épidémie hivernale. D'autre part le ratio des effectifs communaux adultes / enfants varie de moins de 3 à plus de 10 parmi les communes de 500 habitants et plus [15]. L'un dans l'autre, il y a en moyenne autant de cas pédiatriques que de cas adultes, avec des variations saisonnières et inter annuelles importantes.

I TABLEAU 7 I

Incidence de GEAm en fonction de l'âge (DCIR)

Données 2009	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	50 338 347	11 395 315	61 733 662
Nombre de cas de GEAm (France)	2 591 574	2 526 504	5 118 078
Tx d'incidence des GEAm (par an et pour 100 sujets)	5,1	22,2	8,3
Données 2010	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	50 611 506	11 483 242	62 094 748
Nombre de cas de GEAm (France)	3 008 858	3 709 172	6 718 030
Tx d'incidence des GEAm (par an et pour 100 sujets)	5,9	32,3	10,8
Données 2011	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	50 881 163	11 561 464	62 442 627
Nombre de cas de GEAm (France)	2 950 741	3 534 601	6 485 342
Tx d'incidence des GEAm (par an et pour 100 sujets)	5,8	30,6	10,4
Données 2012	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	51 163 853	11 616 737	62 780 590
Nombre de cas de GEAm (France)	3 059 327	3 512 947	6 572 274
Tx d'incidence des GEAm (par an et pour 100 sujets)	5,9	30,2	10,5
Données 2013	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	51 447 746	11 661 218	63 108 964
Nombre de cas de GEAm (France)	2 951 169	3 434 228	6 385 397
Tx d'incidence des GEAm (par an et pour 100 sujets)	5,7	29,4	10,1
Données 2014	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	51 716 870	11 711 352	63 428 222
Nombre de cas de GEAm (France)	2 651 945	3 086 516	5 738 461
Tx d'incidence des GEAm (par an et pour 100 sujets)	5,1	26,3	9
Données 2015[#]	Adultes (16 ans et plus)	Enfants (1-15 ans)	Ensemble
N (France)	51 993 645	11 745 208	63 738 853
Nombre de cas de GEAm (France)	2 507 666	2 926 668	5 434 334
Tx d'incidence des GEAm (par an et pour 100 sujets)	4,8	24,9	8,5

Les données de 2015 ne sont pas complètes

Le sexe n'est pas un caractère discriminant de telle ou telle étiologie ou type d'épidémies et **la prise en compte du sexe ne présente guère d'intérêt**. Le sexe ratio observé dans les épidémies ne s'écarte pas de 1/1 sauf dans de rares Tiac causées par des aliments dont la consommation varie en fonction du sexe (ex. salades).

La séparation des différents types de cas n'est pas préconisée en analyse épidémiologique de routine. Bien que les antiémétiques évoquent plutôt une étiologie virale, les qualités discriminantes des antiémétiques vis-à-vis de l'étiologie n'ont pas été confirmées par l'analyse d'une épidémie d'étiologie virale certaine [16].

La catégorisation en types de cas n'a pas actuellement d'usage épidémiologique. Elle est destinée à maintenir l'homogénéité de la définition de cas dans la durée car elle permet de revenir rétroactivement sur la définition de cas en recomposant les types de cas inclus, dans l'hypothèse de changements thérapeutiques ou de déremboursement survenant au sein de la période d'étude. La catégorie « divers » fondée sur des combinaisons de médicaments peu spécifiques (antispasmodiques, pansements intestinaux) pourrait par exemple moins bien résister à ces changements.

Co-prescription d'une **coproculture**. Dans l'entrepôt des cas, la prescription d'une coproculture apparaît dans 0,6% des cas de GEAm en 2014. Cette information peut servir à l'analyse d'un agrégat local, car la notion de coproculture pourrait constituer un indice en faveur d'une étiologie bactérienne. L'utilité de la pertinence de cette information n'a pas été évaluée.

7.2 Cas « résident présent » et cas « touriste »**

Quand la commune de domicile du bénéficiaire est distante de plus de 50 km de celle d'exercice du médecin (ou du pharmacien qui a délivré les médicaments, si la donnée médecin est manquante), le cas est considéré comme « touriste » (option prise par défaut dans la constitution des fichiers de comptes de cas). Dans certains secteurs ruraux ou de montagne, la distance que parcourt un malade pour aller consulter un médecin peut cependant dépasser 50 km. Il est toutefois possible de modifier la distance seuil ou de créer plusieurs classes de distance, en adaptation à des situations ou des objectifs particuliers (voir § sur l'entrepôt des cas aux données « utilisateurs »).

Géolocalisation et erreur de classement sur l'exposition. En secteur rural, l'imputation systématique du lieu d'exercice du praticien au lieu de résidence du bénéficiaire « touriste » au moment de la gastro-entérite est à proscrire car il est peu probable que la commune du praticien corresponde à celle du patient, notamment en ce qui concerne le pharmacien. Il y a en effet en moyenne en France un médecin (omnipraticien libéral) pour 1 050 habitants et une officine pour 2 900 habitants (Insee 2012) avec des variations géographiques en défaveur des secteurs ruraux. Il appartient donc à l'investigateur de terrain de juger de la faisabilité d'une imputation en fonction de l'implantation locale des acteurs de santé. Nous conseillons pour investiguer des épidémies sur un secteur donné de ne considérer que les résidents présents certains du secteur considéré (codes communes disponibles pour le bénéficiaire et le praticien).

Taux d'incidence inconnu des « cas touristes ». On ne peut pas estimer le taux d'incidence des « cas touristes » en l'absence de données sur la démographie touristique à la commune et à la semaine. Il est cependant possible de faire dans certaines situations des hypothèses raisonnables sur la démographie touristique. Par exemple en période de pointe (vacances de février dans les stations de ski, début août pour les estivants), on peut supposer la saturation de la capacité d'accueil touristique. La mise en œuvre de ce redressement nécessite toutefois de connaître la fraction des étrangers parmi la population

touristique, sachant que les touristes résidant à l'étranger ne sont pas couverts par l'Assurance maladie. S'il s'agit de détecter un excès de cas ponctuel, on peut aussi contourner le calcul d'incidence en prenant pour référence le nombre de cas touristes observés les années précédentes dans des conditions proches en termes de fréquentation touristique. Par exemple, comparer le nombre de cas pendant la première semaine de vacances d'été avec le nombre observé pendant la première semaine des vacances d'été précédentes. La mise en œuvre de cette méthode nécessite aussi l'hypothèse que la part des touristes résidant en France parmi les touristes est constante. Il est clair que la mise en œuvre de ces techniques de détection d'un excès de cas doit s'appuyer sur une bonne connaissance du terrain et qu'elles ne valent que pour la détection d'épidémies franches.

Erreur sur l'incidence des « résidents présents » : Les **résidents absents** (commune du bénéficiaire correspondant au secteur d'étude, mais éloignée de la commune du praticien) sont à la source d'un risque de sous-évaluation de l'incidence parmi les résidents si on rapporte le compte des cas « résidents présents » aux données des recensements. Ce risque s'exprime pendant les vacances scolaires et particulièrement dans les villes. En l'absence de données sur la part de la population absente, la fraction de la population résidente absente ne peut être estimée par la fraction des cas domiciliés sur le secteur d'étude pour lesquels la commune du praticien est éloignée de celle de résidence du cas⁵. En effet, les individus en vacances à l'étranger ne sont pas comptabilisés.

⁵ En 2004 [8] Beudeau P, Bounoure F. *Évaluation épidémiologique d'indicateurs d'incidence des gastroentérites fondés sur les données de l'Assurance Maladie*. [Epidemiological valuation of indices of gastroenteritis based upon French Health Insurance data]. *Environnement, Risques & Santé*. 2006;5(5):1-10., la proportion des ordonnances délivrées hors du département de résidence de l'assuré est limitée en province (3,7 % des cas en moyenne) mais importante en région parisienne (12,5 %). Cette proportion reste stable en période scolaire : le niveau de base oppose les sites de province, où très peu d'ordonnances sont délivrées hors département en période scolaire (1 % des cas), à la région parisienne (9 %). Cette proportion augmente brutalement pendant les vacances quel que soit le site. Elle culmine à 40 % des enfants de 1 à 4 ans absents en région parisienne vers le 15 août.

8 DE L'ENTREPÔT DES CAS AUX DONNÉES « UTILISATEURS »⁶

Les données de l'entrepôt sont mises à disposition des Cire. Une interface d'extraction des données de l'entrepôt des cas individuels est proposée aux Cire en libre-service, à partir de 2016 pour le paramétrage standard en libre-service. Les autres demandes d'extraction sont effectuées par l'unité Rariéc de la DSE.

Trois niveaux d'utilisation pourraient être envisagés à terme :

- Extraction directe avec paramétrage standard libre-service (prédéfini). Cette méthode devrait répondre aux besoins d'investigation d'événements épidémiques suspectés en lien avec l'eau distribuée. Le recours au paramétrage standard favorise la comparaison entre événements (présentation similaire des données) dans une optique de consolidation nationale. Cette fonctionnalité a été mise en place en octobre 2015.
- Extraction avec paramétrage fixé par l'utilisateur, à partir de données proposées directement « requêttables » dans l'entrepôt. Cette méthode convient pour des applications plus spécialisées, comme l'investigation d'événements d'origine inconnue. Il est conseillé de prendre l'attache de la DSE afin de préserver l'homogénéité des méthodes et les possibilités de comparaison avec d'autres événements au niveau national ; cette fonctionnalité pourrait être appliquée après 2017.
- Des applications particulières sont réalisables sur demande à adresser à la DSE. Elles concernent des applications particulières, par exemple celles qui demandent un ajustement de la définition de cas ou un traitement spécifique des données manquantes. Cette offre est déjà effective depuis 2010.

8.1 Paramétrage standard en libre-service

Comptes de cas

En utilisation standard, les comptes de cas couvrent uniquement les résidents présents. Ils sont séparés en 2 classes d'âge (1-15 vs. ≥ 16 ans) (tableau 8). Les données sont fournies par commune et par jour, à partir d'une liste de communes entrées par l'utilisateur (ou code département si l'utilisateur souhaite extraire les données pour la totalité des communes) et des date de début et de fin de la période d'étude.

⁶ Cette section est provisoire. C'est une première proposition à revoir avec les utilisateurs en 2017

I TABLEAU 8 I

Paramétrage de l'utilisateur pour l'extraction des fichiers de comptes de cas en mode standard. Au total, 2 tableaux « date X commune » sont produits correspondant aux 2 classes d'âge

Variable	Commentaire	Entrées	Modalités
date	Date de prescription de l'ordonnance	Dates début et fin	
Commune	Commune de résidence du bénéficiaire (cas résident)	Liste codes Insee ou code département	
Âge_cl	Classe d'âge du bénéficiaire	Prédéfini	[1-15 ans] 16 et plus

Statistiques de qualité du jeu de données

Des données supplémentaires seront mises à disposition sur l'espace collaboratif :

- la proportion de **données inclassables** en cas/non cas par département et par mois ;
- la fréquence des **DM** de géolocalisation par département (tableau 9) ;
- les **données démographiques** en correspondance avec les comptes de cas, c'est-à-dire par jour et par commune, pour les 2 classes d'âge pour l'ensemble du secteur et de la période d'étude. Les données sont des estimations de l'Insee pour l'année correspondant au début de la période d'étude (ou l'année la plus proche en cas d'indisponibilité). Ces données permettent de calculer les taux d'incidence chez les résidents et, au besoin, d'opérer une standardisation sur l'âge de ces taux entre le secteur cible et le secteur témoin.

I TABLEAU 9 I

Comptes de données disponibles pour la localisation des cas dans les fichiers de comptes de cas destinés aux utilisateurs. Modèle de tableau

Géolocalisation : compte des codes valides disponibles		Bénéficiaire (B)			
		Commune (C)	Département (D)	Rien	TOTAL
Praticien (P)	Commune (C)				
	Département (D)				
	Rien			(1)	
	TOTAL				

(1) Compte disponible pour la France, estimation possible au prorata de la population du jeu de données

Les données de qualité permettent d'identifier des problèmes spécifiques à la période et au lieu étudié en complément des opérations de validation faites en routine sur l'entrepôt. Dans le cas général, cette étape ne débouche sur aucune action particulière sinon la validation du jeu de données. Si les DM sont nombreuses (notion qui sera précisée au fil de l'expérience), d'autres contrôles (sur le régime par exemple) sont nécessaires pour définir un éventuel traitement de DM comme :

- imputation (tableau 5) ;
- redressement (tableau 5) ;
- filtrage d'un régime particulier ou de catégories de cas particulières.

Cette dernière manipulation sort alors du cadre du paramétrage standard et du libre-service.

8.2 Paramétrage de données commandées

Les tableaux 10 et 11 fournissent la liste complète des variables susceptibles d'être paramétrées pour l'obtention de données dans le cadre d'une utilisation experte (hors investigation basique d'épidémies).

+

Le tableau 10 donne les variables catégorielles qui pourraient être paramétrées par l'utilisateur quand l'entrepôt aura évolué (V2). Des variables catégorielles peuvent être créées librement à partir de variables quantitatives (âge et distance entre la commune de résidence et celle de consultation). Le nombre et les bornes des classes sont alors paramétrables.

Le tableau 11 fournit des informations utiles pour des applications spécialisées comme celles touchant à la définition de cas (par exemple pour viser une étiologie particulière ou corriger une dérive dans le temps). Le traitement des données manquantes présentant une fréquence jugée anormale relève aussi des traitements particuliers.

I TABLEAU 10 I

Variables catégorielles paramétrables par l'utilisateur

Variable	Commentaire	Format paramétrage	du Exemples modalités de
Date	Date de prescription de l'ordonnance	Dates début et fin	
Commune	Commune de résidence du bénéficiaire (cas résident), soit la commune du praticien (cas touriste)	Listes Code Insee	
Âge_cl	Classe d'âge du bénéficiaire	Bornes des classes	[0-11 mois] [1-5 ans] [6-15 ans] [16-65 ans] [66-75 ans] [76 ans +]
Sexe	Sexe du bénéficiaire	Global / séparés	Homme Femme
StatutResid	Statut de résident ou de touriste Si distance \geq borne alors StatutResid = « touriste »		Par défaut, 2 classes avec borne = 50 km
Distance	Distance « commune bénéficiaire-commune praticien ». Permet de modifier StatutResid	Bornes (km)	Nombre et valeurs des bornes libres

I TABLEAU 11 I

Variables catégorielles non accessible à l'utilisateur mais disponibles pour des applications particulières

Variable	Commentaire	Modalités disponibles
B_Rg1	Régime d'assurance maladie	Régime général
		MSA
		RSI
		Autres
		SLM
cas_type#	Type de cas de GEAm en fonction des médicaments présents sur l'ordonnance	Anti-émétique et Ralentisseur de transit
		Anti-émétique seul
		Ralentisseur de transit seul
		Soluté de Réhydratation Orale
		Divers
copro	Prescription d'une coproculture	Avec/Sans

Détails en § Algorithme

8.3 Fichiers de sortie

Pour la V0 du libre-service :

- les fichiers de sortie sont des fichiers Excel (.xls). Ils comportent 3 onglets : total des cas, cas [1ans ; 15 ans], cas >= 16 ans ;
- chaque onglet est composé d'une première colonne avec les dates de la période d'étude.
- Les autres colonnes correspondent aux différentes communes de la zone d'étude, avec le nombre de cas de GEAm par jour et commune ;
- Les noms des fichiers sont générés automatiquement sous la forme suivante :
 - o gea_cireX_date[JJMMMAA]_HHMMSS.xls
 - o X : numéro de Cire paramétré dans l'application Madispo (ex : Île-de-France = 7)
 - o JJ : jour de l'extraction sur 2 caractères (ex : 01)
 - o MMM : mois de l'extraction (mois anglais) sur 3 caractères (ex : Aout = AUG)
 - o AA : année de l'extraction sur 2 caractères (ex : 2015 = 15)
 - o HH : heure de l'extraction sur 2 caractères (ex : 10)
 - o MM : minute de l'extraction sur 2 caractères (ex : 51)
 - o SS : seconde de l'extraction sur 2 caractères (ex : 23)

Références bibliographiques

- [1] Majowicz SE, Hall G, Scallan E, Adak GK, Gauci C, Jones TF, *et al.* A common, symptom-based case definition for gastroenteritis. *Epidemiol Infect.* 2008;136(7):886-94.
- [2] Galey C. *Production et validation par Santé publique France des cas de Gastro-entérites aiguës médicalisées: de l'extraction du DCIR à l'entrepôt des cas. Procédure du Département santé-environnement.* Saint-Maurice, France ; 2014. 1-39 p.
- [3] Beaudreau P. *Surveillance syndromique des gastro-entérites aiguës : une opportunité pour la prévention du risque infectieux attribuable à l'ingestion d'eau du robinet.* Université de Rennes 1; 2012. 1-239 p.
- [4] Bounoure F, Beaudreau P, Mouly D, Skiba M, Lahiani-Skiba M. *Syndromic surveillance of acute gastroenteritis based on drug consumption.* *Epidemiol Infect.* 2010;139(9):1388-95.
- [5] Penit P. *Étude épidémiologique des gastro-entérites aiguës médicalisées et spécificité chez l'enfant.* Thèse de pharmacie. : Université de Rouen; 2014. 1-139 p.
- [6] Gofti-Laroche L. *Évaluation du risque microbiologique hydrique : validation épidémiologique des fonctions dose-réponse du risque viral et parasitaire.* Étude EMIRA. Thèse. : Université Joseph Fourier - Grenoble1, Faculté de Pharmacie; 2001. 1-259 p.
- [7] McCarthy TA, Barrett NL, Hadler JL, Salsbury B, Howard RT, Dingman DW, *et al.* Hemolytic-Uremic Syndrome and *Escherichia coli* O121 at a Lake in Connecticut, 1999. *Pediatrics.* 2001;108(4):E59.
- [8] Beaudreau P, Bounoure F. *Évaluation épidémiologique d'indicateurs d'incidence des gastro-entérites fondés sur les données de l'Assurance maladie.* *Environnement, Risques & Santé.* 2006;5(5):1-10.
- [9] Van CD, de VH, Vaux S, Le SY, Vaillant V. *Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study.* *Epidemiol Infect.* 2012;140(4):697-705.
- [10] Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, *et al.* Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *BMJ.* 1999;318(7190):1046-50.
- [11] Beaudreau P, de VH, Vaillant V, Mouly D, Ledrans M, Mannschott C, *et al.* Lessons learned from ten investigations of waterborne gastroenteritis outbreaks, France, 1998-2006. *J Water Health.* 2008;6(4):491-503.
- [12] Galey C, L'Azon M, Duchon C, Beaudreau P. *Épidémie de gastro-entérites aiguës dans un camping, Ardèche, France, août 2008.* *BEH.* 2012(33):379-82.
- [13] Pirard P, Gorla S, Nguengang Wakap S, Galey C, Motreff Y, Guillet A, *et al.* No increase in drug dispensing for acute gastroenteritis after Storm Klaus, France 2009. *J Water Health.* 2015;13(3):737-45.

- [14] Mouly D, D VANC, Vincent N, Vaissiere E, Beaudreau P, Ducrot C, *et al.* *Description of two waterborne disease outbreaks in France: a comparative study with data from cohort studies and from health administrative databases.* *Epidemiol Infect.* 2015;1-11.
- [15] Rambaud L, Galey C, Beaudreau P. *Automated detection of case clusters of waterborne acute gastroenteritis from health insurance data - pilot study in three French districts.* *J Water Health.* 2016;14(2):306-16.
- [16] Rambaud L, Mouly D, Schmitt M, Kerrien F, Beaudreau P. *Utilisation des données de remboursement des médicaments de l'Assurance maladie pour identifier et caractériser une épidémie de gastro-entérites d'origine hydrique, Bourg Saint-Maurice (Arcs 1800), 2006.* *BEH.* 2011(31):339-43.
- [17] Beaudreau P, Bounoure F, Genet M. *Estimation de l'incidence des gastro-entérites aiguës à partir des bases de données de l'Assurance maladie.* *Pratiques et Organisation des Soins.* 2006;37(4):313-23.
- [18] Beaudreau P, Mouly D, Lauzeille D. *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.* Rapport de l'Institut de veille sanitaire. Saint-Maurice ; 2006. 1-34 p.

Annexe 1. Historique du projet

Le tableau 12 retrace l'historique du projet.

Les premiers travaux utilisant les données fournies par la Cnamts pour l'épidémiologie des GEA remontent à 1999 (F. Bounoure, 2000, thèse de pharmacie, Université de Rouen). Les données étaient extraites d'Érasme par la CPAM du Havre. La Ville du Havre, partenaire du projet souhaitait examiner le potentiel des données de la Cnamts comme alternative à celles fournies par un réseau de 35 pharmaciens d'officine livrant à la Ville leurs statistiques de vente des médicaments.

I TABLEAU 12 I

Historique de l'exploitation des données de la Cnamts pour l'étude et la surveillance des GEA

Années	Données	Base de Données	Algo rithme GEAm	Épidémiologie
2001 2002-2007	Première convention Cnamts /InVS Livraison de données sur 11 villes	Erasme régionales	V1	Début étude T&G
2007-8	Déremboursement des antiseptiques et produits microbiens			
2008 2009-2014	Deuxième convention : Livraison Datamart France entière par semestre par la Cnamts	SNIIRAM	V2 V2	Fournitures de données aux Cire à la demande pour l'investigation d'épidémies, ou des études particulières Fin étude T&G
2013 ≥2014	Accès InVS aux données sensibles ; requêtage direct par l'InVS	DCIR	V2	Fournitures de données aux Cire à la demande pour l'investigation d'épidémies, ou des études particulières
≥2015				Recherche automatisée d'épidémies, France entière

Les résultats se montrèrent encourageants quant à la possibilité de construire une méthode de discrimination des « vraies » GEA sur la base des informations contenues dans l'ordonnance, et au-delà quant au potentiel épidémiologique des données.

En 2001, l'InVS s'intéressa à ces travaux et décida de construire un système de surveillance des GEA principalement dédié à la surveillance du risque infectieux d'origine fécale porté par l'eau du robinet. Une convention fut passée avec l'Université de Rouen pour assurer le développement et la mise à jour de l'algorithme de discrimination des cas de GEAm fondé sur des enquêtes annuelles en pharmacies. La convention a été constamment renouvelée depuis. La première version de l'algorithme GEAm (V1 du tableau 12) a été publiée [17] ainsi que la description des données qui en résultent correspondant [8]. Ce dernier travail a mis en évidence le manque de spécificité de deux sous-catégories de cas (concernant les monothérapies et les personnes âgées) qui furent exclus dans les applications épidémiologiques.

Une convention fut passée par ailleurs avec la Cnamts pour la livraison de données extraites des BD Erasme régionales pour 11 sites qui avaient été sélectionnés pour l'étude de séries temporelles « turbidité et gastro-entérites » (T&G) visant à modéliser le risque endémique résiduel (cas sporadiques de GEAm) attribuable à des adductions d'eau « microbiologiquement conformes », au sens de la réglementation.

La collecte de données T&G s'est effectuée de 2002 à 2007 et leur exploitation s'est achevée en 2014.

En 2008, le développement du SNIIRAM amena la Cnamts à proposer à l'InVS la fourniture d'un datamart couvrant toute la France. Une deuxième version de l'algorithme GEAm fut élaborée pour prendre en compte (i) le manque de spécificité des cas issus du 1^{er} algorithme et (ii) le déremboursement des produits microbiens et des antiseptiques intestinaux.

L'homogénéité avec l'indicateur passé n'a pu être évaluée (les données de 2008 issues du SNIIRAM comportant de nombreuses données manquantes).

L'extension territoriale de la collecte permet désormais de répondre aux demandes des Cire qui souhaitent investiguer des suspicions d'épidémies de GEA d'origine hydrique. On peut en effet estimer à deux journées d'épidémiologiste, l'exploitation des données fournies contre un mois et plus pour la réalisation d'une enquête en population. Le recours aux données de la Cnamts permet la description spatiotemporelle de l'impact, l'émission d'hypothèse sur le vecteur dans certains cas (eau, aliment), mais l'investigation complète de ces facteurs de risque nécessite toujours une enquête sur le terrain.

En 2013, la Cnil et l'Assurance maladie ont autorisé l'accès de l'InVS à l'ensemble des données « sensibles » (date de naissance du bénéficiaire, commune du bénéficiaire et date de prescription) et leur croisement. L'InVS décida alors de pérenniser et de consolider le système de surveillance en pratiquant les extractions de façon autonome à partir du DCIR et en mettant le système sous assurance qualité. L'algorithme GEAm n'est pas modifié mais une mise à jour des listes des médicaments rétrospective est opérée (portant essentiellement sur les médicaments dont la présence est incompatible avec une GEA).

En 2016, les données de l'entrepôt sont devenues directement accessibles aux Cires.

Annexe 2. Ressources alternatives

Diverses ressources alternatives au SNIIRAM ont été envisagées [18].

Le **Réseau sentinelles** des médecins généralistes présente l'avantage d'une mise à jour hebdomadaire des données. Le taux d'échantillonnage convient aux études nationales voire régionales pour les régions bien dotées de médecins sentinelles mais est largement insuffisante pour les investigations à la commune.

Le **dossier pharmaceutique** pourrait permettre d'accéder à des données comparables à celles obtenues avec le SNIIRAM. La mise à disposition des données est décalée d'une journée. De plus les données incluent les ventes conseils. La couverture atteint 50% de la population en 2013, mais ce taux dépend localement de la participation des pharmacies, de la saisie systématique d'informations client (par lecture de la carte Vitale) pour l'achat de médicaments en vente conseil et de l'acceptation de la clientèle. La résolution spatiale pose également problème avec une unité géographique minimale correspondant à plusieurs pharmacies (les données sont agrégées pour préserver la discrétion sur le niveau de leur activité individuelle). Cette résolution n'est pas compatible en secteur rural avec l'étude du facteur hydrique. Il en va de même des données produites en ligne par le réseau Celtipharm.

Des **réseaux de pharmaciens** ont été mis en place dans le passé dans un objectif épidémiologique. Les données sont alors des **ventes de médicaments**. La maintenance de tels réseaux demande un effort important et les données sont des ventes de médicaments (prescrits et conseils). Il n'est pas possible de construire un indicateur issu des ventes à la fois sensible et spécifique des GEA [4].

Les **ventes conseil** apparaissent dans la littérature scientifique d'origine américaine comme une ressource intéressante. Cette position est subordonnée au système de santé américain qui ne couvre que partiellement la population comparativement à la France. Dans ce contexte, la population recourt beaucoup plus à la vente conseil. En France, les ventes conseils concernent 9% des cas de GEA [3]. Les ventes en ligne de médicaments « conseil » et de parapharmacie représenteraient 10% des ventes (chiffre d'affaire) en Europe, mais seraient moins développées en France (www.lafnim.com/fnim-rendez-vous.asp?id=27). Nous ne disposons d'aucune information relative à l'utilisation future de ces statistiques pour l'épidémiologie.

Les ventes des grossistes sont utilisées avec certains avantages (réactivité) mais les inconvénients (faible résolution géographique, couverture faible de la population) empêchent l'étude des facteurs de risque au niveau local.

Les statistiques de visite aux urgences hospitalières pour GEA sont utilisées aux États-Unis pour l'épidémiologie des infections entériques, notamment celles concernant les personnes âgées, jugées fiables. Ces données sont très peu sensibles comparativement aux consultations en France (facteur 10 à 100 selon la base de comparaison, $\geq 65Fr/\geq 65US$ et $TotFr/\geq 65US$). Les données de visite aux urgences n'ont pas été systématiquement évaluées pour la France, ni celles de SOS médecin, mais leur utilité paraît réduite comparativement aux indicateurs issus du DCIR. Leur intérêt provient de leur disponibilité à un jour qui permettrait leur utilisation pour la mise en place d'une prévention rapprochée (promulgation d'avis de non consommation). Cependant les visites aux urgences ne concernent que 1% des cas de GEA symptomatiques [9] et leur utilité est limitée à la détection d'événements majeurs. Les données de SOS médecin ne peuvent convenir dans la mesure (1) où la qualité de la localisation des patients est disparate dans les données et (2) SOS médecins couvre principalement les villes alors que les épidémies d'origine hydrique frappent principalement les secteurs ruraux. Là encore l'utilité se réduit à l'aide à la détection d'événements majeurs impactant les agglomérations.