

11 mars 2016 Marseille

Réseaux sociaux, objets connectés et big data : usages et enjeux du numérique en prévention et promotion de la santé

Colloque organisé par le Comité régional d'éducation pour la santé PACA, dans le cadre du Pôle régional de compétences en éducation et promotion de la santé



Suivez le colloque sur Twitter

[#prcpaca](#)

Veille sanitaire et Big data

- ◆ Daniel Eilstein
- ◆ Responsable de mission, Institut de veille sanitaire



avec le soutien de
l'ARS PACA et de l'INPES

Surveillance épidémiologique



« Collecte continue et systematique, analyse et interprétation de données de santé essentielles pour la planification, mise en place et l'évaluation des pratiques en santé publique, étroitement associée à la diffusion en temps opportun de ces données à ceux qui en ont besoin. L'étape finale du cycle de la surveillance est l'application de ces données au contrôle et à la prévention des maladies et accidents »

Thacker SB. Surveillance. In : Gregg MB, editor. Field Epidemiology. Oxford University Press New York, 1996: 16-32.

Surveillance épidémiologique



- Estimer l'**importance** d'un phénomène en santé publique
- Suivre les **tendances** des indicateurs épidémiologiques
- Suivre les **caractéristiques des personnes** atteintes ou menacées
- **Détecter** des phénomènes représentant un danger pour la santé
- Contribuer à l'**évaluation** des politiques de contrôle et de prévention
- Suggérer des **hypothèses** pour la recherche

Veille sanitaire



« Collecte et analyse en continu par les structures de santé publique de signaux pouvant représenter un risque pour la santé publique dans une perspective d'alerte et d'action précoce »

Dabis F, Desenclos JC. Epidémiologie de terrain. John Libbey Eds Paris, 2012

Veille sanitaire

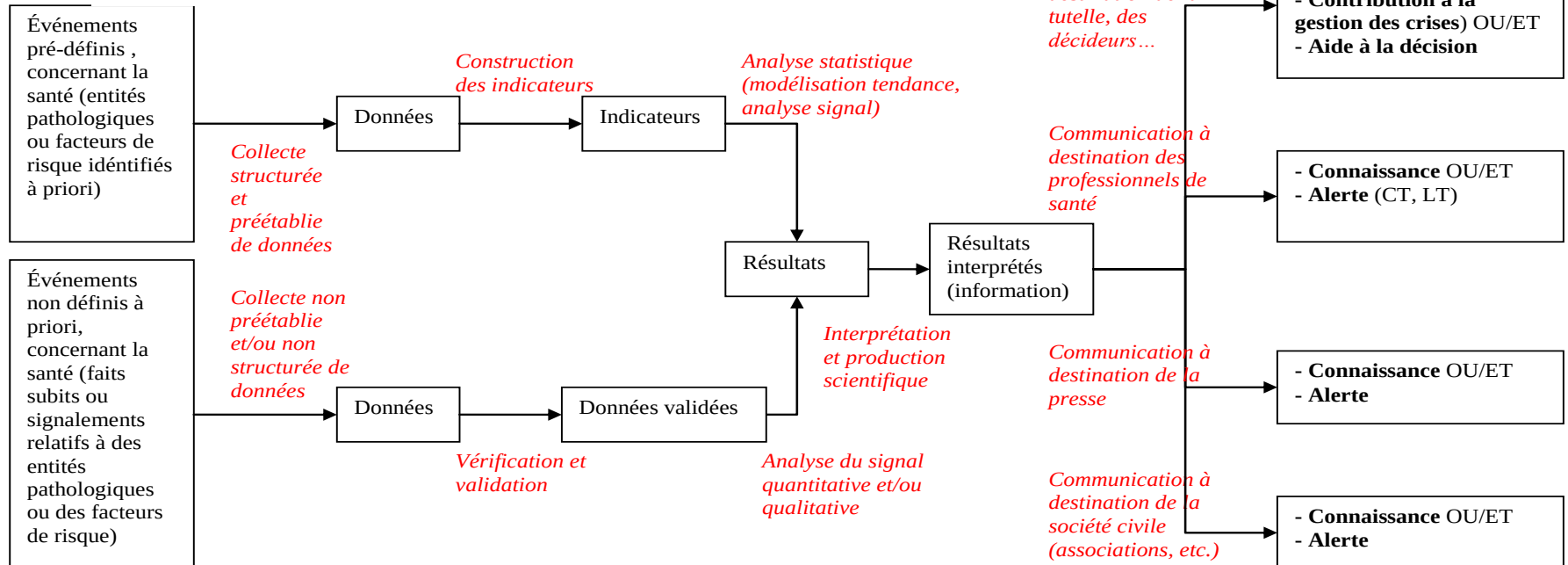


- Associe :
 - surveillance spécifique
 - surveillance non spécifique (syndromique)
 - signalement d'événements (approche qualitative)
 - veille internationale
 - veille documentaire
 - réseaux sociaux, moteurs de recherche, blogs...
- Sécurité sanitaire plus que santé publique
- Implique :
 - capacité permanente de réception, d'analyse, d'investigation des signaux dans une perspective d'alerte et d'action
 - filtre / tri dans l'afflux des signaux

Veille sanitaire



SURVEILLANCE ÉPIDÉMIOLOGIQUE (VEILLE PRÉ-FORMATÉE)



VEILLE NON PRÉ-FORMATÉE

Eilstein D, Salines G, Desenclos JC. Veille sanitaire : outils, fonctions, processus.
Rev Epidemiol Sante Publique 2012;60:401-11.

Sources et données actuelles



- Classiques

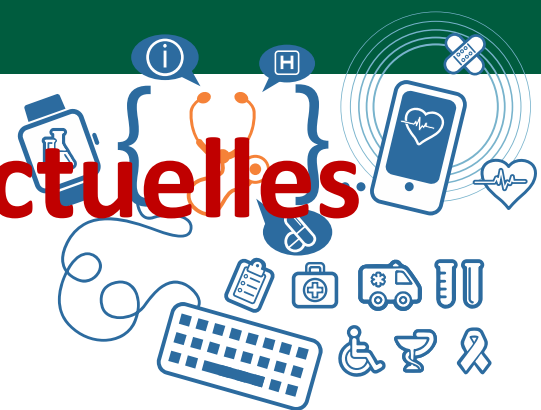
- Déclaration/notification obligatoire, certificats de santé...
- Données issus de services, de laboratoires, de médecins
- Réseau Sentinelle, Réseau des GROG
- Registres de morbidité
- Enquêtes périodiques en population
- (Très grandes) Cohortes
- Investigations
- Mesures environnementales
- Appels téléphoniques, e-mails...
- Prélèvements biologiques, biothèques

Sources et données actuelles



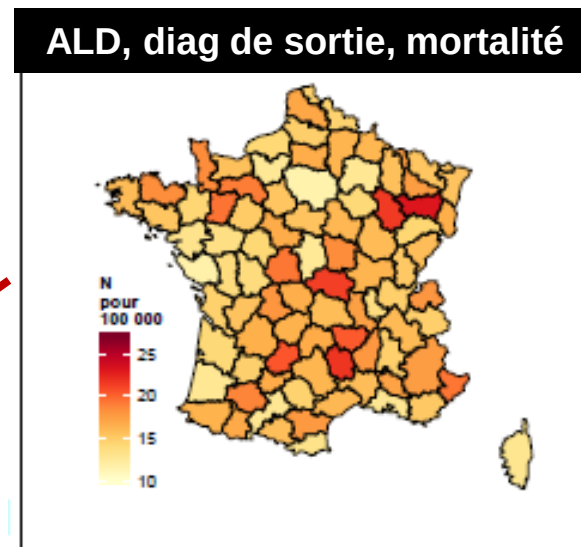
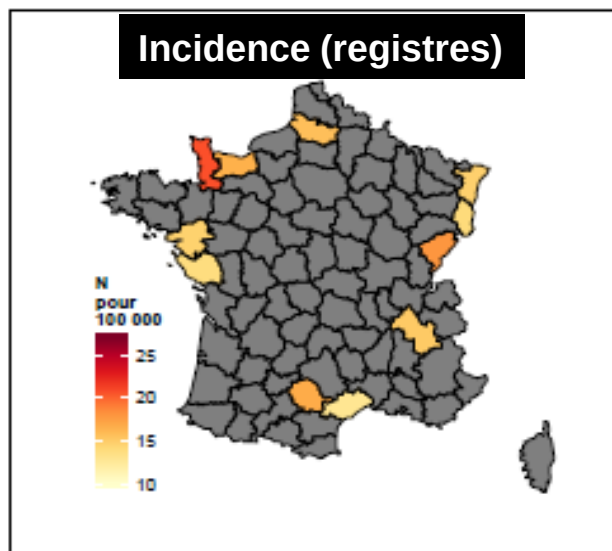
- Bases de données médico-administratives (SNDS)
 - Sniiram
 - > EGB
 - > DCIR (+ possibilité de croisement variables sensibles)
 - > PMSI
 - Certificats de décès
- Système Sursaud
 - Résumés de passage aux urgences (flux quotidien)
 - Décès (80 % (*communes informatisées*) ; flux quotidien)
 - Appels SOS médecins (quasi tous ; flux quotidien)
 - Certificats de mortalité électronique (7%, en cours)

Méthodes analytiques actuelles



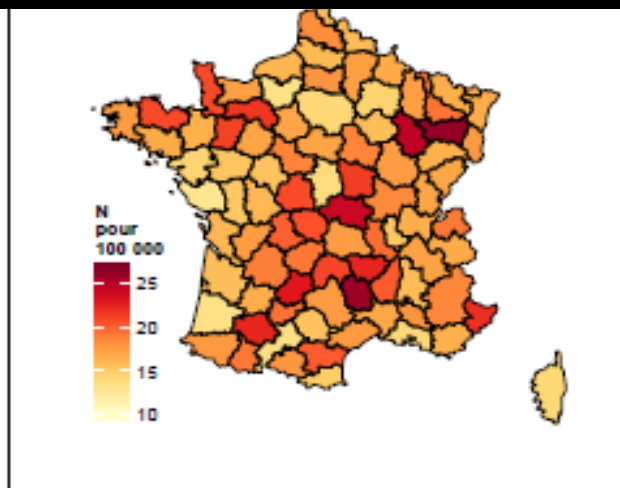
Objectif	Méthode
Description	Statistiques descriptives
Analyse de tendance et prévision	Modélisation
Associations indicateurs	Modélisation
Analyse du signal « événements » (signal qualitatif)	Jugements experts
Analyse du signal « indicateurs » (signal quantitatif)	Comparaison observé/attendu (modélisation seuils)
Evaluation de risque	Méthodes utilisant les résultats de l'expérimentation ou de l'épidémiologie (modélisation)

Méthodes analytiques actuelles



Modèle

Incidence estimée par département



- Modélisation du ratio **ALD**, **PMSI**, **mortalité**, **données des registres** dans les départements couverts par les registres
- Calibration et validation
- Estimation dans les région/départements sans registre

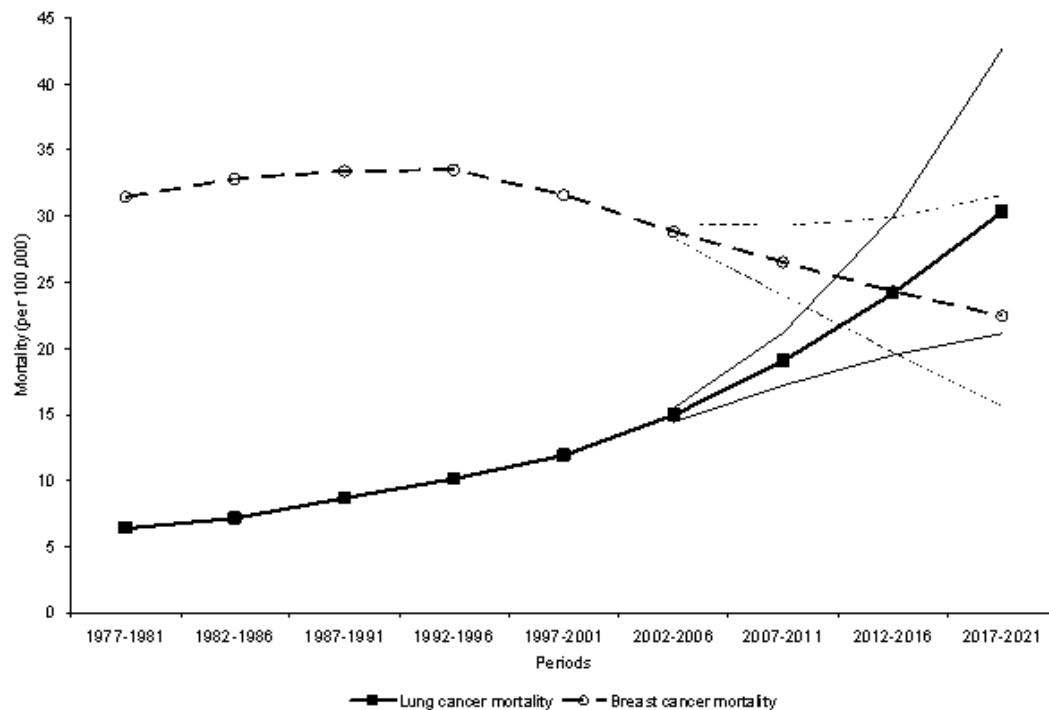
Estimation de l'incidence des cancers

Travail collaboratif
Francim/registres, HCL,
INCa, InVS

Méthodes analytiques actuelles



Projection de la mortalité par cancers du sein et du poumon chez les femmes de France métropolitaine

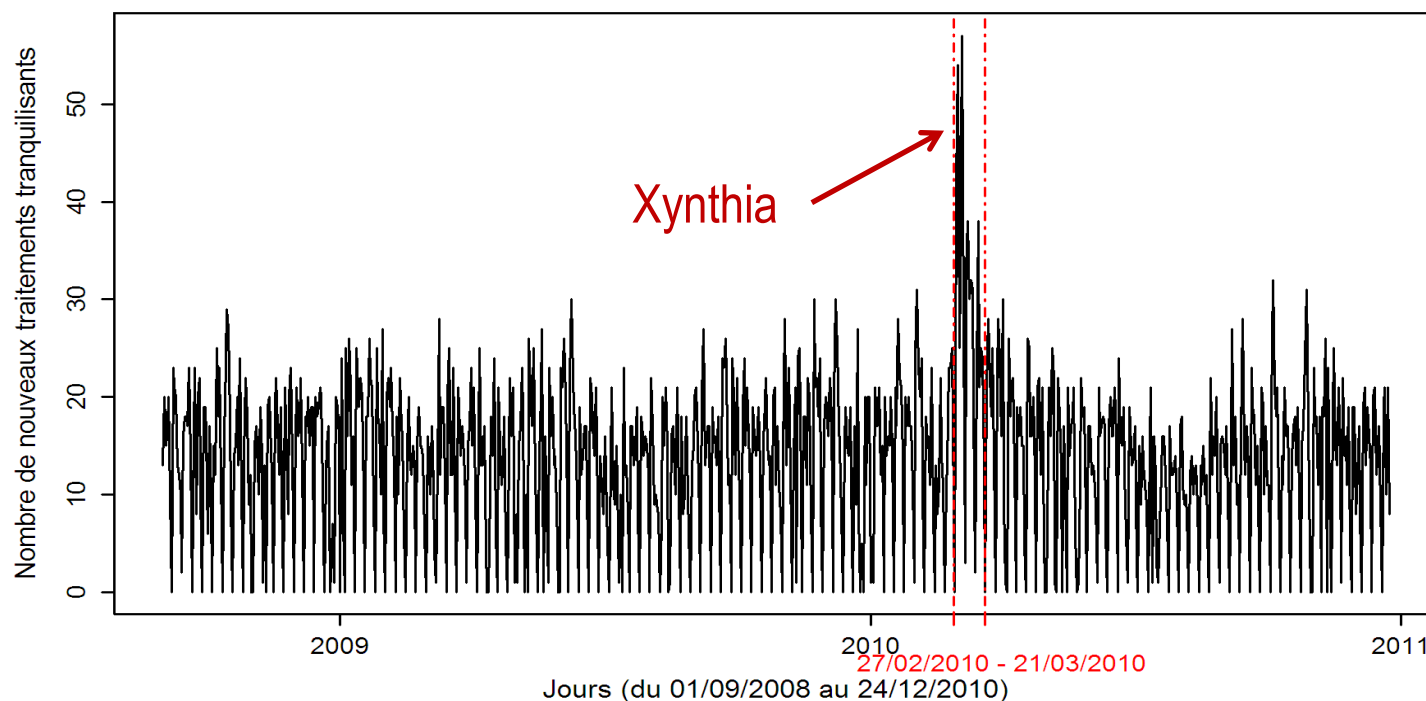


Eilstein D, Eshai K. Lung and breast cancer mortality among women in France: future trends. Cancer Epidemiology (2012), <http://dx.doi.org/10.1016/j.canep.2012.07.008>.

Méthodes analytiques actuelles



Impact sur la santé mentale de la tempête Xynthia de 2010 chez les hommes et les femmes



Motreff Y, Pirard P, Gorla S, Labrador B, Gourier-Fréry C, Nicolau J, Le Tertre A, Chan-Chee C. Increase in Psychotropic Drug Deliveries after the Xynthia Storm, France, 2010. Prehosp Disaster Med. 2013 Jun 27:1-5

Difficultés actuelles



- Sources et données

- Sources et données classiques
 - > Lourdes, difficiles...
- Bases de données médico-administratives
 - > Pas conçues pour une utilisation épidémiologique : validité , représentativité ?



- Méthodes

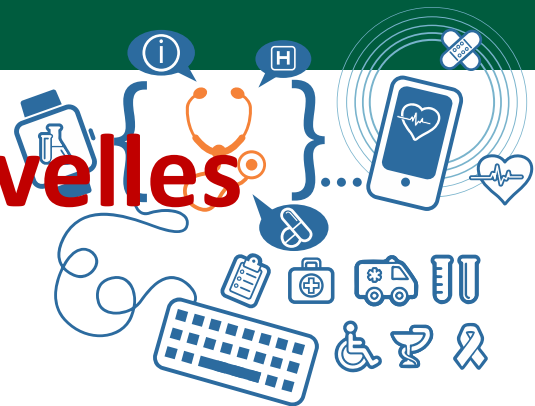
- Pas adaptées à l'afflux des données
- Trop rigides
- Pas assez sensibles/spécifiques

Montée en puissance des sources actuelles



- Systèmes de surveillance / détection syndromique
- Bases médico-administratives
- Cohortes à un niveau européen ou mondial
- Prélèvements biologiques

Sources et données nouvelles (et futures)



- Dossier médical personnel (Dossier médical personnel qui redevient partagé)
- Moteurs de recherche, forums, blogs
- *Crowd sourcing* délibéré, motivé et ciblé ?
- Objets connectés (santé et environnement) : *quantified self* et santé connectée
- Mesures par satellites
- Réseaux d'alerte s'appuyant sur des outils informatiques
- Systèmes de vigilance spécifique (nutriviigilance...)
- Courrier (médical...)
- « Web des données » : relier des données hétérogènes (personnes, lieux, produits, concepts...)

Nouvelles méthodes analytiques



- Doivent

- Croiser un grand nombre de données
 - > De sources différentes (environnementales, sanitaires, comportementales, socioéconomiques, démographiques, de veille informationnelle...)
 - > De natures différentes (qualitatives, quantitatives)
 - > De granularités différentes (temporelles et spatiales)
- Détecter/analyser les signaux faibles

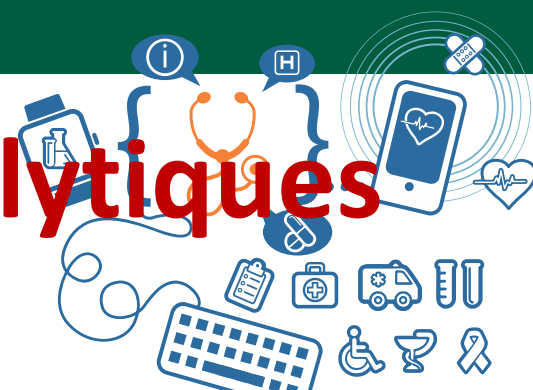
- Méthodes

- Fouille de données
- Télé-épidémiologie
- Analyse sémantique
- Décision : IA (programme Watson d'IBM)
- Veille prospective

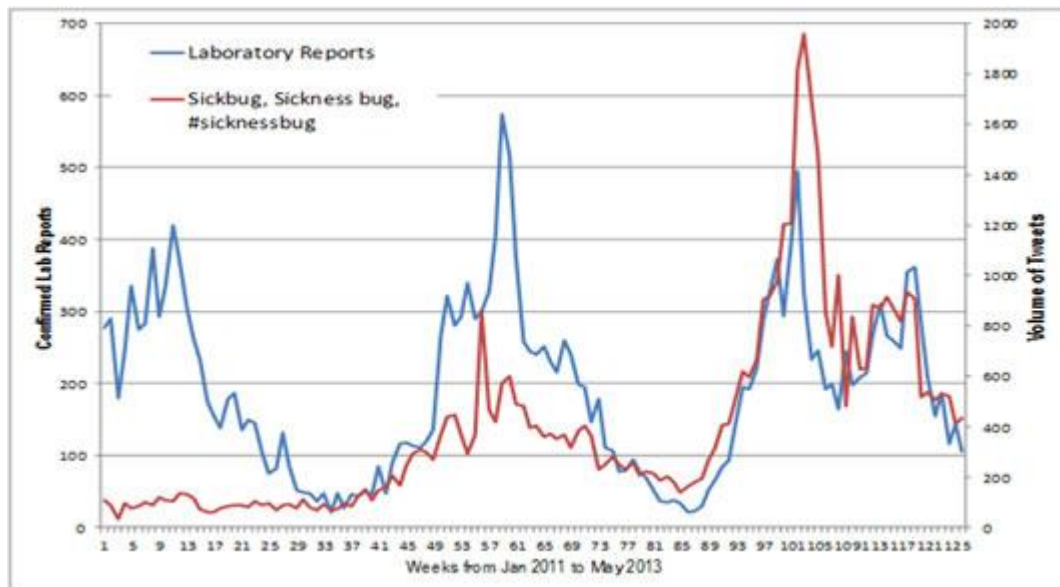


Nouvelles méthodes analytiques

Food Standards Agency : Twitter et gastroentérites



All weeks	Same week and past weeks	Future weeks
sick bug, sickness bug, stomach flu, vomiting, #sicknessbug	#winterbug, #norovirus, norovirus AND outbreak, norovirus, norovirus AND symptoms, sickness virus, winter AND sickness bug, winter sickness, winter virus, winter vomiting	#barf, #flu, barf, being sick, chuck up, feeling sick, flu, heave, puke, retch, sick AND fever, spit up, stomach pain, throw up, throwing up, upset stomach, vom



06/12/2013

Avantages



- Analyse multi-polluants de différentes natures
 - Sanitaires/environnementales
 - Quantitatives/qualitatives
 - Issues de modélisation et des mesures
- Augmentation de la qualité de détection
- Aide à l'expertise et à la décision

Difficultés et risques



- Techniques
 - Embouteillage informationnel : ↗ qualité de détection
⇒ ↗ nombre d'événements détectés et fausses alertes
 - Capacités de jugement des experts dépassées :
solutionnisme
- Socio-éthiques (et juridiques)
 - Non-respect de la vie privée (croisement)
 - > risque de ré-identification
 - > risque de profilage ultra fin / sélection population
 - Privation de liberté : pouvoir algorithmique*



* Séminaire « A qui profitent les Big Data ? Est-il possible de croire à l'innovation technologique et économique basée sur l'avènement des Big Data, tout en conciliant la préservation des libertés ? ». Institut rhônalpin des systèmes complexes, Lyon 28/04/2015

Difficultés et risques

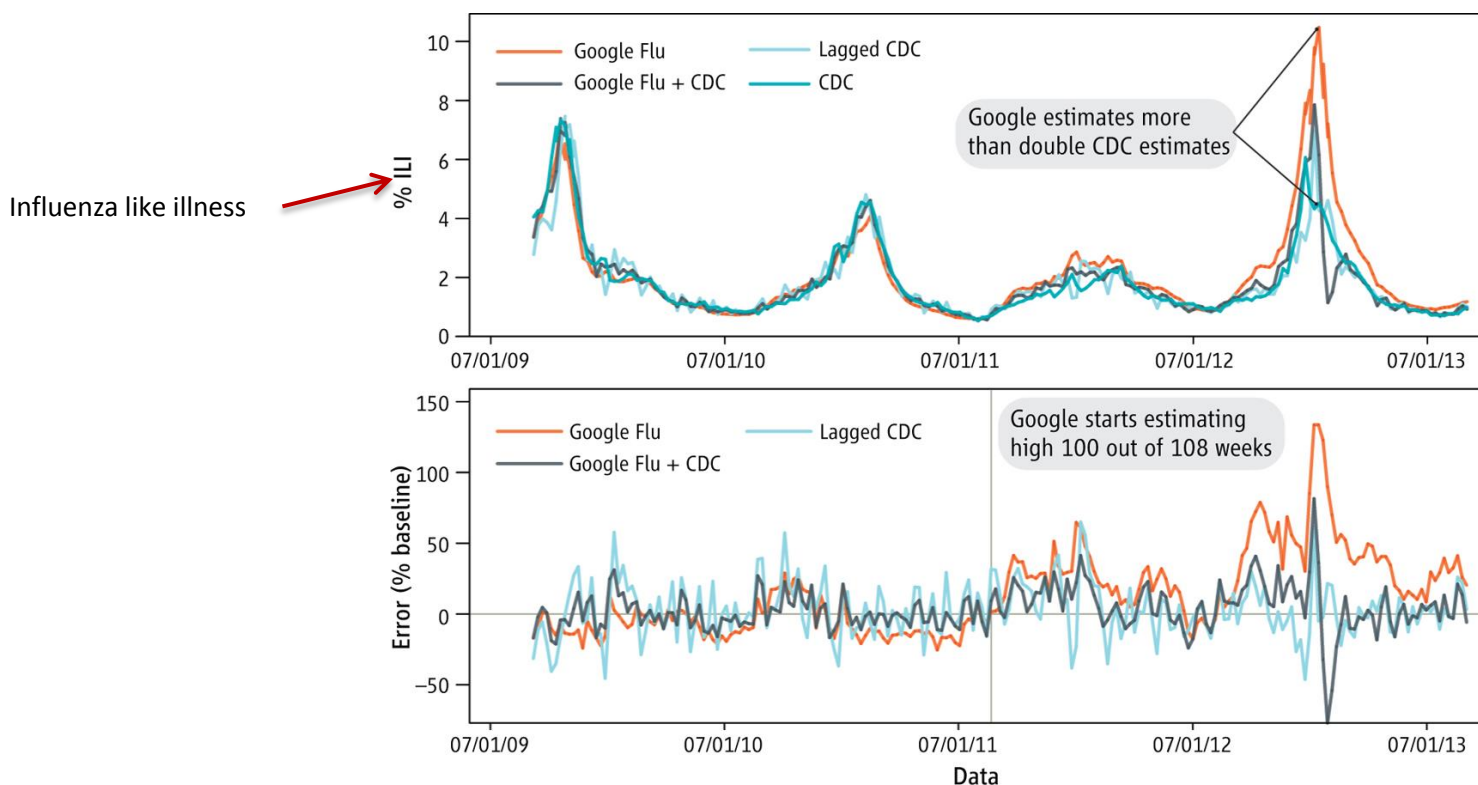


- Socio-économiques
 - Défavorable à l'assurance collective
 - > Les individus refuseraient de payer pour les autres
 - > L'assureur minimiserait sa prise de risque
 - Favorable aux inégalités sociales et territoriales de santé

Difficultés et risques



Google Flu Trends surestime la prevalence de la grippe en 2011–2012 (50%) et en 2012–2013 (100%)



Lazer D, Kennedy R, King G, Vespignani A. The Parable of Google Flu: Traps in Big Data Analysis. *Science* 2014; 343:1203-5. <http://science.sciencemag.org/content/343/6176/1203.full-text.pdf+html>.

Le futur vu de l'InVS



- Hypothèse 1 « tendancielle »
 - « Close Data » pour les données publiques mais essor du « Big data » dans l'industrie, le marketing, etc.
- Hypothèse 2 « Open Data »
 - Mise à disposition des données publiques
 - Investissent dans de grands projets liés au « Big data »
- Hypothèse 3 « Rendre les comportements vertueux »
 - Assurances, mutuelles, sécurité sociale utilisent les possibilités offertes par le « Big data » pour faire payer les assurés en fonction de leurs comportements (bonus/malus)

Eilstein D, Pozuelos J, Fuchs D, Gorza M, Therre H, Xerri B, et al. Risques pour la santé : évolution des conséquences pour la veille sanitaire à l'horizon 2024. Exercice de veille prospective InVS, 2014. Saint-Maurice : Institut de veille sanitaire ; 2015. 60 p.

Le futur vu du R31



- **Acteurs et/ou facteurs**
 - Développements technologiques (déclaration, analyse signalements)
 - Accroissement des signalements par citoyens
 - Montée en puissance des mesures par le citoyen individuel ou cadre associatif
 - Amélioration de la détection des expositions et du suivi des impacts sanitaires
 - Extension spatiale des systèmes de détection
- **Événements initiateurs de ruptures**
 - Multiplication des autres phénomènes menaçant la santé (pandémies, climat)
 - Décision politique de moins détecter
 - Intervention des acteurs privés dans le jeu : développement de détection/dépistage/diagnostic...

R31 : réseau d'organismes partenaires de l'Anses

Le futur vu du R31



- **Hypothèse 1 « tendancielle »**
 - État surveille moins (difficultés de financement)
 - Garde un certain monopole.
- **Hypothèse 2 « rupture »**
 - Évolution technologique, auto-surveillance et ↗ signalements
 - Surveillance noyée par les informations
 - Question de la propriété et de l'accès aux données obtenues
 - Nouveaux acteurs qui vendent des données
- **Hypothèse 3 « favorable »**
 - Surveillance bien maîtrisée (État et institutions)
 - Amélioration de la qualité de la détection

Merci



- Anne Doussin
- Javier Nicolau