



# Stratégie d'évaluation des risques servant à la détection précoce et à la prédiction des éclosions de maladies infectieuses associées aux changements climatiques

EE Rees<sup>1\*</sup>, V Ng<sup>2</sup>, P Gachon<sup>3</sup>, A Mawudeku<sup>4</sup>, D McKenney<sup>5</sup>, J Pedlar<sup>5</sup>, D Yemshanov<sup>5</sup>, J Parmely<sup>6</sup>, J Knox<sup>1,2</sup>

## Résumé

Une nouvelle gamme de stratégies de surveillance visant à détecter les infections émergentes et à reconnaître les risques accrus d'éclosions de maladies infectieuses qui seraient causées par les changements climatiques est en cours de développement. Ces stratégies comportent des systèmes de surveillance événementielle et la modélisation des risques. Les systèmes de surveillance événementielle font appel à des données Internet à code source libre, comme des rapports officiels ou des médias, et des réseaux sociaux (p.ex. Twitter), pour détecter les signes d'une menace imminente et ils peuvent servir, en association avec les systèmes de surveillance traditionnels, à faciliter l'émission d'avis précoces de menace pour la santé publique. Depuis quelque temps, ces systèmes comprennent des applications d'intelligence artificielle, comme l'apprentissage automatique et le traitement des langues naturelles, qui accroissent la rapidité, la capacité et l'exactitude du filtrage, du classement et de l'analyse des données Internet portant sur la santé. La modélisation des risques recourt à des méthodes statistiques et mathématiques pour évaluer la gravité de l'émergence d'une maladie et pour diffuser certains facteurs portant sur l'hôte (p.ex. le nombre de cas signalés), le pathogène (p.ex. la pathogénicité) et l'environnement (p.ex. si le climat est propice aux espèces réservoirs). On élargit les types de données de ces modèles en vue d'y inclure les renseignements sur la santé tirés de données Internet à code source libre et l'information sur les tendances des êtres humains et des biens en matière de mobilité. Ces renseignements contribuent à cerner les populations à risque et à prédire par quelles voies les infections pourraient se répandre dans de nouvelles régions et de nouveaux pays. Outils très efficaces complétant les stratégies traditionnelles de surveillance qui déterminent ce qui s'est déjà produit, les systèmes de surveillance événementielle et la modélisation des risques devraient servir de plus en plus à éclairer les mesures de santé publique visant à prévenir, à détecter et à contrôler les hausses de cas de maladies infectieuses causées par les changements climatiques.

**Citation suggérée :** Rees EE, Ng V, Gachon P, Mawudeku A, McKenney D, Pedlar J, Yemshanov D, Parmely J, Knox J. Stratégie d'évaluation des risques servant à la détection précoce et à la prédiction des éclosions de maladies infectieuses associées aux changements climatiques. *Relevé des maladies transmissibles au Canada* 2019;45(5):132–40. <https://doi.org/10.14745/ccdr.v45i05a02f>

**Mots-clés :** changements climatiques, évaluation des risques, systèmes de surveillance événementielle, intelligence artificielle, apprentissage automatique, traitement des langues naturelles, modélisation des risques

## Introduction

Le réchauffement climatique tend à accélérer depuis quelques décennies. En effet, durant la période de 1850 à 2017, les neuf années les plus chaudes ont toutes été observées au cours des douze dernières années. Au total, la température annuelle moyenne de l'atmosphère a augmenté d'environ 0,97 °C entre 1880 et 2017 (1). Aussi infime soit-il, cet accroissement de la température moyenne mondiale est la cause de changements importants dans le climat planétaire et il produit des effets non moins importants sur la société : la hausse du niveau de la mer (et l'érosion qu'elle

Cette oeuvre est mise à la disposition selon les termes de la licence internationale Creative Commons Attribution 4.0



## Affiliations

<sup>1</sup> Division des sciences des risques pour la santé publique, Laboratoire national de microbiologie, Agence de la santé publique du Canada, Saint-Hyacinthe (Québec)

<sup>2</sup> Division des sciences des risques pour la santé publique, Laboratoire national de microbiologie, Agence de la santé publique du Canada, Guelph (Ontario)

<sup>3</sup> Centre pour l'étude et la simulation du climat à l'échelle régionale (ESCIER), Université du Québec à Montréal (UQAM), Montréal (Québec)

<sup>4</sup> Bureau de la connaissance de la situation et des opérations, Centre de mesures et d'interventions d'urgence, Agence de la santé publique du Canada, Ottawa (Ontario)

<sup>5</sup> Ressources naturelles Canada, Service canadien des forêts, Centre de foresterie des Grands Lacs, Sault Ste. Marie (Ontario)

<sup>6</sup> Réseau canadien pour la santé de la faune, Université de Guelph, Guelph (Ontario)

\*Correspondance :  
[erin.rees@canada.ca](mailto:erin.rees@canada.ca)



provoque) et l'augmentation de la fréquence et de l'intensité des inondations, des sécheresses (et les incendies et les mauvaises récoltes qu'elles causent) et des épisodes de pluie verglaçante (2). Fait qui revêt une importance particulière pour le Canada, le réchauffement climatique est encore plus sensible sous les hautes latitudes et pendant l'hiver (3). En effet, depuis 70 ans, la température annuelle moyenne globale au pays a grimpé de 1,8 °C (4) et la température hivernale moyenne, de 3,4 °C (4). Dans certaines parties du Nord-Ouest, cette hausse est encore plus marquée. Étant donné que les changements climatiques influent non seulement sur les températures, mais aussi sur les régimes de précipitations, le Canada a tendance à devenir plus aride dans l'Ouest et plus humide dans l'Est (4).

On sait que les changements de température et de précipitations provoqués par les changements climatiques augmentent les risques de transmission de maladies infectieuses. Les changements climatiques modifient en effet les aires de distribution des vecteurs de maladies (comme les tiques et les moustiques) et des espèces réservoirs (les oiseaux, les rongeurs et les cerfs), qui contribuent à la transmission aux humains d'agents pathogènes infectant les tiques et les moustiques, à mesure que le climat devient plus propice aux vecteurs et aux espèces réservoirs (5,6). Par exemple, la hausse des cas de maladie de Lyme au Canada reflète l'expansion de son vecteur, la tique à pattes noires (*Ixodes scapularis*), dont l'aire de répartition se déplace vers le nord aux États-Unis et couvre désormais aussi le sud du Canada, les changements climatiques ayant rendu le pays plus hospitalier pour les tiques (7,8). Cette expansion dans une région où les vecteurs et les espèces réservoirs peuvent se multiplier se traduit non seulement par la hausse des risques de maladies infectieuses sporadiques, mais aussi par la probabilité accrue que ces vecteurs et les maladies qu'ils transmettent deviennent endémiques (6,9–11).

En outre, les changements climatiques influent sur la mobilité des personnes et des biens. On s'attend à voir une augmentation du nombre de réfugiés climatiques, ces personnes déplacées fuyant les phénomènes météorologiques extrêmes qui menacent leur existence ou leurs moyens de subsistance (11). Ces réfugiés, souvent issus de régions du monde où les maladies infectieuses sont plus répandues et ayant des pratiques et des calendriers de vaccination différents risquent, sans le vouloir, de transmettre ces maladies aux Canadiens (12). Le tourisme, tant dans les pays d'origine que dans les pays de destination, est également touché par les changements climatiques, qui modifient les facteurs antinomiques incitant les touristes à voyager et le potentiel de propagation des maladies (13–15). Sans le vouloir, on risque de transporter des vecteurs et des pathogènes dans le cadre des importations canadiennes par la voie des airs, par la voie terrestre et par bateau (16–18). Comme on le sait, les conteneurs terrestres et maritimes facilitent les invasions de moustiques parce que les larves peuvent se développer dans l'eau stagnante qu'ils contiennent et que, même en l'absence de liquide, leurs

œufs résistent à la déshydratation pendant des semaines, voire des mois (19,20). De plus, les passagers des lignes aériennes transmettent les infections dans de nouvelles régions. Ainsi, au Canada, les passagers de retour au pays ont transmis le virus Zika et provoqué une éclosion du coronavirus du syndrome respiratoire aigu sévère (SRAS) (15,21,22).

Les risques accrus de maladies infectieuses causées par les changements climatiques posent ainsi des risques importants pour la santé publique, et des efforts sont en cours pour surveiller, évaluer et prédire leurs répercussions. Auparavant, la gestion de la santé publique reposait sur des systèmes de déclaration et de surveillance des maladies à déclaration obligatoire, qui visaient à détecter les éclosions, à suivre la progression des maladies et à éclairer les politiques de prévention et de contrôle. Malheureusement, les systèmes traditionnels de surveillance se caractérisent par des retards dans la transmission et l'analyse des données, et dans la communication des résultats.

Compte tenu de la nécessité d'avoir un système de surveillance des problèmes émergents qui soit plus en temps réel et de comprendre plus tôt leurs répercussions possibles sur la santé, on a élaboré et l'on continue d'élaborer deux stratégies d'évaluation des risques : les systèmes de surveillance événementielle qui, de plus en plus, font appel à l'intelligence artificielle et la modélisation des risques. Le présent aperçu a pour objectif de décrire ces deux stratégies et d'expliquer en quoi elles peuvent éclairer les mesures de santé publique prises afin de prévenir, de détecter et de contrôler les augmentations de cas de maladies infectieuses causées par les changements climatiques.

## Systemes de surveillance événementielle

Les systèmes de surveillance événementielle font appel à plusieurs données à code source libre et à plusieurs techniques d'évaluation permettant de reconnaître les menaces pour la santé (23,24). On compte, parmi ces données à code source libre, les fils de presse en ligne, les réseaux sociaux et les autres flux de données Internet, en de nombreuses langues, qui permettent de détecter de façon précoce les menaces pour la santé publique. On a démontré que ces systèmes sont plus rapides que les données de surveillance traditionnelles transmises par les laboratoires ou les hôpitaux (25) et qu'ils peuvent servir, en association avec les systèmes traditionnels de surveillance, à faciliter la communication précoce de menaces pour la santé publique (26). Plus on reconnaît les signes d'une éclosion en pleine évolution rapidement, plus on peut, sans tarder, en assurer le suivi, ainsi que planifier et mettre en œuvre les mesures d'intervention nécessaires afin de protéger la santé publique (27).



Il existe trois types de systèmes de surveillance événementielle : modéré, partiellement modéré et entièrement automatisé (28). Le niveau d'automatisation influe sur la gestion du flux de l'information dans les systèmes à partir des données à code source libre provenant des agrégateurs (Factiva, Google News, Moreover Baidu, etc.), du format RSS et des réseaux sociaux, qu'ils soient de sources officielles ou officieuses (Twitter pour les Centers for Disease Control américains et le grand public, etc.), et à partir des rapports officiels validés (Organisation mondiale de la Santé, Centers for Disease Control américains, etc.). Programme de suivi des maladies émergentes, ProMED constitue un exemple de système modéré. Il était au cœur de l'élaboration de systèmes de surveillance événementielle il y a plus de 25 ans (29,30). ProMED est exécuté par des analystes bénévoles (qui constituent en quelque sorte des conservateurs experts des données) qui cherchent et choisissent des articles de presse, en valident le contenu et avertissent les abonnés en cas d'éclosion importante de maladies infectieuses. Les avantages de ce programme sont son faible rapport signal sur bruit, son accès ouvert et sa grande portée. Par contre, ces bénévoles ne couvrent pas toutes les populations à risque, leurs partis pris peuvent influencer sur la modération des cas et, ce qui est normal, ils ne disposent pas de ressources suffisantes pour fournir des renseignements détaillés permettant de connaître la situation et ainsi d'évaluer avec précision le niveau de danger (29).

Le Réseau mondial d'intelligence en santé publique (RMISP) est un système partiellement modéré créé par le gouvernement du Canada, en collaboration avec l'Organisation mondiale de la Santé, quatre ans après ProMED (31–33). Son accès est réservé aux organismes ayant un mandat lié à la santé. Les algorithmes d'intelligence artificielle du RMISP automatisent un flux de 2000 à 3000 articles de presse par jour, modérés par 12 analystes experts qui reconnaissent les menaces et lancent une alerte au besoin, à la lumière de données contextuelles tacites (contexte historique, tendances du marché, interdictions de voyager, anomalies climatiques, etc.). L'exemple suivant permet de voir à quel point ce réseau peut être utile : au début de 2003, les analystes ont pu lire des rapports chinois mentionnant la hausse des ventes de traitements antiviraux immédiatement avant le déclenchement de l'épidémie mondiale de SRAS (34). Contrairement à ProMED, le RMISP est filtré à plusieurs étapes au moyen de l'intelligence artificielle et d'analystes qualifiés. L'intelligence artificielle permet de traiter un flux de données plus important, tandis que les analystes disposent des ressources nécessaires pour faire connaître la situation. Bien que ProMED et le RMISP puissent tous deux analyser plusieurs langues, il est coûteux de le faire dans le cas du RMISP, car il faut engager des analystes parlant couramment des langues étrangères (33).

Parmi les systèmes entièrement automatisés, on compte le système d'information médicale de la Commission européenne (MedISys), le Pattern-based Understanding and Learning System (PULS) et HealthMap. Bien que ces systèmes soient accessibles au public, ils comportent aussi des zones réservées aux organismes sanitaires : forums de discussion privés,

fonctionnalité accrue et traitement des données de sources commerciales (35,36). Les systèmes entièrement automatisés traitent les données plus rapidement et leur utilisation est moins coûteuse que les systèmes modérés. Leur principal désavantage est leur rapport signal sur bruit plus élevé, qui se traduit par un risque accru de détecter de fausses menaces (37,38). On peut toutefois relier les systèmes de surveillance événementielle et les faire fonctionner en synergie pour atténuer ce risque (39). Par exemple, MedISys recueille les données, au faible rapport signal sur bruit, de ProMED et du RMISP, et il fait appel aux algorithmes du PULS, qui traitent mieux les langues. Le PULS extrait l'information sur les cas repérés par MedISys, puis il la renvoie à MedISys (36,40). Le **tableau 1** résume les types de systèmes de surveillance événementielle.

### Applications de l'intelligence artificielle

Les applications de l'intelligence artificielle au traitement des données ont révolutionné la capacité des systèmes de surveillance événementielle à détecter rapidement et sans faute les menaces telles que les éclosions de maladies infectieuses. Les données Internet à code source libre sont considérées comme des données non structurées, en ce sens que les articles de presse, les blogues, les gazouillis, etc., racontent et décrivent des faits. Comme le texte, les chiffres et les dates ne suivent pas les modèles de données (p.ex. une base de données) pouvant servir à la détection automatisée des cas et à la modélisation des risques, il faut traiter les données à code source libre en vue d'extraire et de structurer l'information sur la nature du cas, le lieu et la date où il s'est produit, et les personnes touchées. Les systèmes de surveillance événementielle recourent à des méthodes de traitement des langues naturelles pour traiter et comprendre les cas (46–48). Le traitement des langues naturelles est un domaine des sciences qui se consacre à la compréhension du discours humain (49). L'analyse du sous-langage est l'une des premières méthodes. On se sert des règles et des modèles pour interpréter et classer le vocabulaire, la syntaxe et la sémantique d'un texte descriptif non structuré. Les systèmes de surveillance événementielle ont des taxonomies associant des termes prédéfinis et leurs synonymes aux termes trouvés dans les sources de données. Tout comme la recherche documentaire traditionnelle, la classification taxinomique de textes descriptifs peut repérer des articles sur la santé en cherchant des termes connexes. Par exemple, les synonymes de la grippe A sont H1N1, grippe H1N1, grippe de type A, influenza A, influenza de type A, grippe porcine, grippe du porc et grippe de Hong Kong (50). Si la méthode du sous-langage de détection des données sur la santé dans les systèmes de surveillance événementielle est efficace, elle comporte aussi des inconvénients. Les taxonomies ne sont pas faciles à généraliser. Il faut en créer une pour chaque maladie dont on assure le suivi et aussi la mettre à jour puisque la langue évolue et que les scientifiques font sans cesse de nouvelles découvertes sur les maladies. C'est pourquoi le traitement des langues naturelles a posé une assise solide en matière de recours aux méthodes d'apprentissage automatique.



Tableau 1 : Résumé de certains systèmes de surveillance événementielle

Type	Exemple	Année de création	Accessibilité au public
Système modéré <sup>a</sup>	Program for Monitoring Emerging Disease (ProMED) (29,30)	En 1994 à titre d'organisme sans but lucratif	Oui
Système partiellement modéré <sup>b</sup>	Réseau mondial de renseignement de santé publique (RMISP) (31–33)	En 1998 grâce à la collaboration du gouvernement du Canada et de l'Organisation mondiale de la Santé	Non; réservé aux organismes sanitaires membres du Réseau
Système entièrement automatisé <sup>c</sup>	Medical Information System (MedISys) (36,41,42)	En 2004 par la Commission européenne	Oui
	HealthMap (35,38,40,43)	En 2006 par le Boston Children's Hospital	Oui
	Pattern-based Understanding and Learning System (PULS) (36,44,45)	En 2007 par le Département d'informatique de l'Université d'Helsinki, en Finlande	Oui

<sup>a</sup> Dans un système modéré, des « conservateurs » experts bénévoles reconnaissent, examinent et valident les sources, et produisent les rapports

<sup>b</sup> Le système partiellement modéré acquiert, classe et filtre les sources automatiquement. Ses conservateurs experts modèrent le sous-ensemble des sources et produisent les rapports

<sup>c</sup> Le système entièrement automatisé acquiert, classe et filtre les sources automatiquement, et produit aussi les rapports automatiquement

L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle qui fait appel à des algorithmes, comme des modèles statistiques, pour exécuter une tâche donnée sans instructions explicites et en recourant plutôt à des modèles et à l'inférence. Les systèmes de surveillance événementielle recueillent des données Internet à code source libre (les flux et les requêtes), puis les filtrent par une combinaison d'analyse du sous-langage et d'apprentissage automatique, ce dernier servant aux tâches plus complexes, soit l'analyse de la syntaxe, de la sémantique, de la morphologie, de la pragmatique et du discours (51). L'apprentissage automatique peut ainsi servir à déterminer la différence entre les articles ne portant pas sur la santé (p.ex. la « fièvre Justin Bieber » fait référence aux fans du chanteur) et les articles portant sur l'écllosion d'une maladie infectieuse (43,51,52). L'apprentissage automatique peut également servir à éclaircir les ambiguïtés dans les dates et les lieux, comme les écllosions passées et les écllosions en cours, dans les articles discutant du contexte historique (53,54). De plus, on élabore des applications avant-gardistes de l'apprentissage automatique, telles que la structuration des données sur les cas sous forme de listes de parcours des éléments épidémiologiques (la liste des patients infectés et l'information connexe : état de santé, sexe, lieu, date d'apparition de la maladie, hospitalisation, etc.) pouvant servir à enquêter sur les écllosions et à réaliser une modélisation des risques (55). Après avoir traité les données Internet à code source libre et créé un modèle, on peut, le cas échéant, examiner l'événement et produire un rapport à son sujet. Qui plus est, on peut analyser des données supplémentaires afin de communiquer l'incidence, actuelle et prévue, de la menace pour la santé. Le **tableau 2** constitue le résumé du flux d'information provenant de la collecte de données par les systèmes de surveillance événementielle, de leur traitement, de leur analyse et de la production de rapports.

## Modélisation des risques

La plus grande variété des données de modélisation constitue un progrès important dans l'évaluation des risques. Dans le contexte des maladies infectieuses, la modélisation des risques est le processus par lequel on reconnaît et l'on caractérise,

chez les individus ou les populations, les facteurs accroissant leur vulnérabilité à contracter une maladie (leur âge, leur proximité du lieu de l'écllosion, etc.). L'induction statistique est une méthode solide et instructive de modélisation des risques, qui comprend l'analyse de régression. Elle sert à établir le lien entre les facteurs de risque (variables explicatives) et le résultat (p.ex. le nombre de cas signalés). Les modèles de régression, et l'induction statistique en général, s'élargissent et comprennent désormais l'information tirée de données Internet à code source libre. En voici un exemple remontant aux tout débuts : on recourait aux données recueillies par le moteur de recherche de Google Flu Trends pour prédire le résultat du nombre de consultations de médecins pour des maladies d'allure grippale (56). On se servait ensuite du modèle résultant pour prédire le nombre de cas de grippe saisonnière une ou deux semaines à l'avance. Malheureusement, cette approche prédisait mal les écllosions survenant à des moments autres que pendant la saison de la grippe, en raison des associations de mots recherchées par le moteur de recherche qui n'avaient rien à voir avec la grippe saisonnière (p.ex. la saison hivernale du basketball) (57). On a ensuite amélioré l'exactitude du système pour ce qui est de la prédiction des tendances de grippe saisonnière en recourant à des sources supplémentaires de données à code source libre (p.ex. Twitter) et en élargissant la méthode de régression de façon à ce qu'elle profite des algorithmes d'apprentissage automatique qui peuvent trouver des associations complexes entre le résultat et les variables explicatives (58). Qui plus est, on a amélioré la modélisation de la régression pour les risques d'infection en y ajoutant, en plus des données Internet à code source libre, d'autres variables explicatives (p.ex. les données climatiques et météorologiques transmises par les satellites) sur la présence, sur les déplacements et sur la distribution des pathogènes, des vecteurs, des espèces réservoirs et des personnes infectées (59,60). Par exemple, en Chine, on a amélioré les prédictions sur le nombre de cas de syndrome pieds-mains-bouche chez les enfants en incluant les données hebdomadaires sur la température et sur les précipitations, ainsi que les données sur les requêtes sur le syndrome pieds-mains-bouche faites au moyen de Baidu, le moteur de recherche chinois (61).



Autre approche fréquente de modélisation des risques, les modèles compartimentaux simulent de façon mathématique la dynamique de transmission au sein d'une population – autrement dit, le flux des états de santé chez les individus qu'on classe en patients sains (S), infectés (I) et guéris (R). Par exemple, les modèles SIR requièrent qu'on définisse les paramètres pour le taux de patients infectés (ou, au contraire, la période d'infection) et le taux de contacts infectés. Il est alors possible d'estimer si l'infection d'une population se transformera en épidémie et de caractériser la prévalence de la maladie au fil du temps. L'approche des modèles compartimentaux s'est récemment élargie pour inclure la simulation de la dynamique de transmission au sein de plusieurs populations (métapopulations). Il faut y ajouter les données sur la mobilité afin de définir le taux des personnes se déplaçant d'une population à l'autre (62). Sur le plan de la métapopulation, on peut considérer la mobilité humaine comme le déplacement des gens dans un réseau interrelié de villes et de pays. On peut tirer ces données des relevés des appels faits par des téléphones cellulaires et des volumes de passagers se déplaçant en avion (63,64). Grâce à la modélisation de la métapopulation, on peut savoir par quels chemins les pathogènes se propageront au Canada et en calculer la probabilité (65,66). Par exemple, le virus Zika serait apparu pour la première fois au Brésil entre août 2013 et avril 2014. Des voyageurs infectés seraient arrivés à Rio de Janeiro, à Brasilia, à Fortaleza ou à Salvador, et leur arrivée a été suivie d'épidémies à Haïti, au Honduras, au Vénézuéla, puis en Colombie (21).

## Discussion

On ne sait trop si les changements climatiques influenceront sur les nombreux facteurs liés à l'occurrence et à la diffusion des maladies infectieuses. Ces facteurs comprendront sans aucun

doute les fluctuations de l'aire de distribution des vecteurs et des espèces réservoirs, la mobilité des gens et des biens, ainsi que l'arrivée possible de pathogènes qui en découle, ce qui entraînera des répercussions sur l'exposition et les risques de transmission. Afin d'assurer le suivi des éclosions de maladies infectieuses de façon rapide et efficace, les professionnels de la santé publique ont besoin d'un meilleur accès à des données de surveillance à jour. Pour ce faire, on complète de plus en plus par des systèmes de surveillance événementielle les données traditionnelles, par exemple celles tirées des systèmes de surveillance des maladies à déclaration obligatoire. Même si les systèmes de surveillance événementielle profitent de l'apprentissage automatique et du traitement des langues naturelles qui optimisent l'exploitation des données, il reste néanmoins des obstacles à surmonter (59). Ainsi, il existe encore des problèmes de partage et de confidentialité des données à résoudre. Par exemple, à quel niveau peut-on utiliser et divulguer des données personnelles dans la détection de cas de maladie? Google et Twitter fournissent librement au public des données finement agrégées par semaine et par ville. Cela dit, des renseignements plus précis sur la date et le lieu de la source assureraient une détection plus détaillée (26). Qui plus est, les habitants de cette planète utilisent Internet et les réseaux sociaux chacun à leur façon et à des moments différents. En effet, on a constaté des écarts dans l'usage d'Internet et du téléphone cellulaire en Afrique (67); en Chine, c'est Baidu et non Google, qui est le principal moteur de recherche (61); et la tendance des gens à se servir de Twitter pour déclarer des maladies dépend de leur âge et de leur statut socio-économique (68).

La modélisation des risques permet d'estimer l'incidence des maladies infectieuses émergentes sur la santé. Les progrès accomplis dans cette modélisation sont, entre autres, les

**Tableau 2 : Flux d'information des données Internet à code source libre dans les systèmes de surveillance événementielle**

Systèmes de surveillance événementielle	Collecte de données	Traitement	Analyse	Production de rapports
Systèmes modérés	Des analystes humains repèrent les données Internet à code source libre portant sur la santé	Des analystes humains examinent le cas, procèdent au filtrage et établissent son niveau de danger	Aucun	On envoie par courriel des rapports sur les menaces pour la santé et on les affiche sur le site Web du système de surveillance événementielle
Systèmes partiellement modérés ou entièrement automatisés	Source automatique de données Internet à code source libre	La classification taxonomique et les algorithmes d'apprentissage automatique filtrent et classent les cas selon leurs métadonnées (p.ex. le type de menace, le lieu et la date). Les algorithmes d'apprentissage automatiques évaluent leur niveau de pertinence. Dans les systèmes partiellement modérés, ce sont des analystes humains qui organisent les sources de données présentant un score élevé	Les techniques analytiques évoluent au fil du temps et diffèrent d'un système de surveillance événementielle à l'autre. On compte, parmi les techniques courantes, le géomarketing des cas; les diagrammes à barres montrant les fluctuations du nombre de mots-clés au fil du temps, du nombre d'articles détectés et du nombre (attendu et observé) de cas de la maladie en question; les nuages de mots-clés illustrant l'importance des termes-clés; les alertes en cas de hausse soudaine du nombre de cas, de la fiabilité des sources ou du nombre de sources uniques	On envoie par courriel des rapports sur les menaces pour la santé, on les affiche sur le site Web du système de surveillance événementielle et l'on avise les communautés pertinentes d'utilisateurs d'applications Web





suivants : l'intégration des données climatiques et des données Internet à code source libre qui éclairent ces estimations, et le calcul de la mobilité des êtres humains, qui propagent les maladies infectieuses aux quatre coins du globe. À l'instar des systèmes de surveillance événementielle, les approches de modélisation des risques sont limitées par la disponibilité des données obtenues. Par exemple, les relevés des appels faits avec des téléphones cellulaires et les données liées au trafic aérien fournissent respectivement des renseignements à la tour de téléphonie cellulaire la plus proche et à l'aéroport le plus proche, mais des données plus précises sur l'emplacement des patients sont fournies, sous réserve des questions de confidentialité, par le système de localisation GPS des téléphones mobiles. Les données sur chaque individu pourraient améliorer énormément notre compréhension des facteurs liés à l'occurrence des maladies et à la diffusion des pathogènes, par exemple le rôle joué par certains patients dans l'écllosion de SRAS en 2003 (69).

## Conclusion

Les progrès accomplis dans l'évaluation des changements associés aux vecteurs, aux espèces réservoirs et à l'activité humaine, de même que leurs répercussions sur les maladies infectieuses, sont suivis par plusieurs stratégies de surveillance et d'analyse. Les systèmes de surveillance événementielle ont recours à des données à code source libre pour recueillir de l'information sur les maladies infectieuses. Ces systèmes sont soit modérés, soit partiellement modérés, soit entièrement automatisés, chaque type comportant des avantages et des désavantages. On observe une tendance à l'automatisation, car elle permet de traiter des volumes importants de données. De plus, l'apprentissage automatique et le traitement des langues naturelles deviennent de plus en plus précis au point qu'un jour, leurs capacités dépasseront peut-être celles des modérateurs humains. On se sert souvent de la modélisation des risques pour comprendre et prédire les répercussions des maladies infectieuses sur la santé à l'aide de l'induction statistique et des modèles compartimentaux. Ces méthodes améliorent les capacités de détecter les populations vulnérables aux maladies émergentes et de prédire par quelles voies elles arriveront au

pays et quelle sera leur incidence sur la santé, en intégrant les données Internet à code source libre et les données sur la mobilité humaine aux données plus traditionnelles sur le climat et les éclosions de maladies infectieuses. Les méthodes que nous avons présentées ici sont des développements prometteurs qui nous aideront à affronter les menaces changeantes posées par des maladies dont l'évolution est influencée par les changements climatiques. En ayant en main plus rapidement une plus grande quantité de données plus précises, les professionnels de la santé publique pourront confirmer et évaluer en moins de temps les éventuelles éclosions de maladies infectieuses et, de ce fait, en élaborer et en entreprendre le traitement (et concevoir et déployer d'autres stratégies de contrôle) de manière plus rapide.

## Déclaration des auteurs

EER — Conceptualisation, Enquête, Rédaction — ébauche originale, supervision et administration du projet  
VN — Enquête, Rédaction — examen et révision  
PG — Rédaction — examen et révision  
AM — Rédaction — examen et révision  
DM — Rédaction — examen et révision  
JP — Rédaction — examen et révision  
DY — Rédaction — examen et révision  
JP — Rédaction — examen et révision  
JK — Enquête, Rédaction — examen et révision

## Conflit d'intérêts

Aucun.

## Financement

Le présent travail a été soutenu par l'Agence de la santé publique du Canada.



## Références

1. National Oceanic and Atmospheric Administration. National Centers for Environmental Information. Climate at a Glance: Global Mapping. NOAA; 2018 ( Accessed 2018-11-14). [www.ncdc.noaa.gov/cag/](http://www.ncdc.noaa.gov/cag/)
2. O'Neill BC, Oppenheimer M, Warren R, Hallegatte S, Kopp RE, Pörtner HO, Scholes R, K, van Ypersele JP, Yohe G. IPCC reasons for concern regarding climate change risks. *Nature Climate Change* 2017;7:28–37. DOI
3. Environnement et Changement climatique Canada. Données et scénarios climatiques : Synthèse des observations et des résultats récents de modélisation. Gatineau(QC) : 2016. <https://www.canada.ca/fr/environnement-changement-climatique/services/changements-climatiques/publications/donnees-scenarios-synthese-observations-recents.html>
4. Blunden J, Anrtd DS, Hartfield G, editors. State of the Climate in 2017. *Bull Am Meteor Soc* 2018;99(8):Si-S332. DOI
5. Myers P, Lundrigan BL, Hoffman SM, Haraminac AP, Seto SH. Climate-induced changes in the small mammal communities of the Northern Great Lakes Region. *Glob Change Biol* 2009;15(6):1434–54. DOI
6. Roy-Dufresne E, Logan T, Simon JA, Chmura GL, Millien V. Poleward expansion of the white-footed mouse (*Peromyscus leucopus*) under climate change: implications for the spread of lyme disease. *PLoS One* 2013 Nov;8(11):e80724. DOI PubMed
7. Gasmi S, Ogden NH, Lindsay LR, Burns S, Fleming S, Badcock J, Hanan S, Gaulin C, Leblanc MA, Russell C, Nelder M, Hobbs L, Graham-Derham S, Lachance L, Scott AN, Galanis E, Koffi JK. Maladie de Lyme au Canada, de 2009 à 2015. Relevé des maladies transmissibles au Canada. 2017;43(10):219-25. DOI
8. McPherson M, García-García A, Cuesta-Valero FJ, Beltrami H, Hansen-Ketchum P, MacDougall D, Ogden NH. Expansion of the Lyme disease vector *Ixodes Scapularis* in Canada inferred from CMIP5 climate projections. *Environ Health Perspect* 2017 May;125(5):057008. DOI PubMed
9. Ogden NH, Mechai S, Margos G. Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. *Front Cell Infect Microbiol* 2013 Aug;3:46. DOI PubMed
10. Lafferty KD. The ecology of climate change and infectious diseases. *Ecology* 2009 Apr;90(4):888–900. DOI PubMed
11. Berchin II. Climate change and forced migrations: an effort towards recognizing climate refugees. *Geoforum* 2017;84:147–50. DOI
12. McMichael C. Climate change-related migration and infectious disease. *Virulence* 2015;6(6):548–53. DOI PubMed
13. Scott D, McBoyle G. Using a 'tourism climate index' to examine the implications of climate change for climate as a tourism resource. In: Matzarakis A, de Freitas CR, editors. *Proceedings of First International Workshop on Climate Tourism and Recreation*. 2001 Oct 5- Oct 10; Porto Carras, Greece. International Society of Biometeorology; 2001, p. 69–88. [www.academia.edu/2876399/Using\\_a\\_tourism\\_climate\\_index\\_to\\_examine\\_the\\_implications\\_of\\_climate\\_change\\_for\\_climate\\_as\\_a\\_tourism\\_resource](http://www.academia.edu/2876399/Using_a_tourism_climate_index_to_examine_the_implications_of_climate_change_for_climate_as_a_tourism_resource)
14. Scott D, McBoyle G, Schwartzentruber M. Climate change and the distribution of climatic resources for tourism in North America. *Clim Res* 2004;27(2):105–17. DOI
15. Ogden NH, Fazil A, Safronetz D, Drebot MA, Wallace J, Rees EE, Decock K, Ng V. Risk of travel-related cases of Zika virus infection is predicted by transmission intensity in outbreak-affected countries. *Parasit Vectors* 2017 Jan;10(1):41–9. DOI PubMed
16. Lounibos LP. Invasions by insect vectors of human disease. *Annu Rev Entomol* 2002;47(1):233–66. DOI PubMed
17. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Adv Parasitol* 2006;62:293–343. DOI PubMed
18. Tuite AR, Thomas-Bachli A, Acosta H, Bhatia D, Huber C, Petrasko K, Watts A, Yong JHE, Bogoch II, Khan K. Infectious disease implications of large-scale migration of Venezuelan nationals. *J Travel Med*. 2018;25(1). DOI
19. Manguin S, Boete C. The Importance of Biological Interactions in the Study of Biodiversity. Rijeka, Croatia: InTech; 2011. Chapter 3, Global Impact of Mosquito Biodiversity, Human Vector-Borne Diseases and Environmental Change; p. 27–50. DOI
20. Egizi A, Kiser J, Abadam C, Fonseca DM. The hitchhiker's guide to becoming invasive: exotic mosquitoes spread across a US state by human transport not autonomous flight. *Mol Ecol* 2016 Jul;25(13):3033–47. DOI PubMed
21. Zhang Q, Sun K, Chinazzi M, Pastore Y Piontti A, Dean NE, Rojas DP, Merler S, Mistry D, Poletti P, Rossi L, Bray M, Halloran ME, Longini IM Jr, Vespignani A. Spread of Zika virus in the Americas. *Proc Natl Acad Sci USA* 2017 May;114(22):E4334–43. DOI PubMed
22. Ruan S, Wang W, Levin SA. The effect of global travel on the spread of sars. *Math Biosci Eng* 2006 Jan;3(1):205–18. DOI PubMed
23. Barboza P, Vaillant L, Le Strat Y, Hartley DM, Nelson NP, Mawudeku A, Madoff LC, Linge JP, Collier N, Brownstein JS, Astagneau P. Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. *PLoS One* 2014 Mar;9(3):e90536. DOI PubMed
24. Hartley D, Nelson N, Walters R, Arthur R, Yangarber R, Madoff L, Linge J, Mawudeku A, Collier N, Brownstein J, Thinus G, Lightfoot N. Landscape of international event-based biosurveillance. *Emerg Health Threats J* 2010;3(1):7096. DOI PubMed
25. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, Eysenbach G, Brownstein JS. Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis* 2009 May;15(5):689–95. DOI PubMed



26. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014 Mar;343(6176):1203–5. [DOI PubMed](#)
27. Heymann DL. SARS and emerging infectious diseases: a challenge to place global solidarity above national sovereignty. *Ann Acad Med Singapore* 2006 May;35(5):350–3. [PubMed](#)
28. Linge JP, Steinberger R, Weber TP, Yangarber R, van der Goot E, Al Khudhairy DH, Stilianakis NI. Internet surveillance systems for early alerting of health threats. *Euro Surveill* 2009 Apr;14(13):19162. [PubMed](#)
29. Carrion M, Madoff LC. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. *Int Health* 2017 May;9(3):177–83. [DOI PubMed](#)
30. Yu VL, Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clin Infect Dis* 2004 Jul;39(2):227–32. [DOI PubMed](#)
31. Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. *Can J Public Health* 2006 Jan-Feb;97(1):42–4. [PubMed](#)
32. Mawudeku A, Blench M, Boily L, St. John R, Andraghetti R, Ruben M. *Infectious Disease Surveillance*. 2nd ed. New York, NY: John Wiley & Sons; 2013. Chapter 31, The Global Public Health Intelligence Network; p. 457–69. [www.wiley.com/en-us/Infectious+Disease+Surveillance%2C+2nd+Edition-p-9780470654675](http://www.wiley.com/en-us/Infectious+Disease+Surveillance%2C+2nd+Edition-p-9780470654675)
33. Dion M, AbdelMalik P, Mawudeku A. Les données massives et le Réseau mondial d'information en santé publique (RMISP). *Relevé des maladies transmissibles au Canada* 2015;41(9):209–14. [DOI](#)
34. Knobler S, Mahmoud A, Lemon S, Sivitz L, Oberholtzer K, editors. *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary* (Institute of Medicine). Washington, DC: The National Academies Press; 2004. p. 376. [www.nap.edu/catalog/10915/learning-from-sars-preparing-for-the-next-disease-outbreak-workshop](http://www.nap.edu/catalog/10915/learning-from-sars-preparing-for-the-next-disease-outbreak-workshop)
35. Harris JK, Hinyard L, Beatty K, Hawkins JB, Nsoesie EO, Mansour R, Brownstein JS; Louis Department of Health. Evaluating the implementation of a twitter-based foodborne illness reporting tool in the city of St. *Int J Environ Res Public Health* 2018 Apr;15(5):833. [DOI PubMed](#)
36. Linge JP, Steinberger R, Fuat F, Bucci S, Belyaeva J, Gemo M, Al-Khudhairy D, Yangarber R, van der Goot E. *Advanced ICTs for Disaster Management and Threat Detection: Collaborative and Distributed Frameworks*. Hershey, PA: IGI Publishing; 2011. Chapter: 9, MedISys: Medical Information System; p.131-142. [DOI](#)
37. Mantero J, Belyaeva J, Linge JP. How to maximise event-based surveillance web- systems : the example of ECDC/JRC collaboration to improve the performance of MedISys. Luxembourg; European Commission Joint Research Centre; 2011. Report No.: JRC 63805. <http://publications.jrc.ec.europa.eu/repository/bitstream/11111111/16206/1/lb-na-24763-en-c.pdf>
38. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 2008 Jul;5(7):e151. [DOI PubMed](#)
39. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P; Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS One* 2013;8(3):e57252. [DOI PubMed](#)
40. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008 Mar-Apr;15(2):150–7. [DOI PubMed](#)
41. Rortais A, Belyaeva J, Gemo M, van der Goot E, Linge JP. MedISys: an early-warning system for the detection of (re-) emerging food- and feed-borne hazards. *Food Res Int* 2010;43(5):1553–6. [DOI](#)
42. Alomar O, Batlle A, Brunetti JM, García R, Gil R, Granollers A, Jimenez S, Lavina A. Llinge JP, Pautasso M, Reverte C, Riudavets J, Rortais A, Sancanelli G, Volani S, Vos S. Development and testing of the media monitoring tool MedISys for early identification and reporting of existing and emerging plant health threats. *Bull OEPP* 2015;45(2):288–93. [DOI](#)
43. Hawkins JB, Tuli G, Kluberg S, Harris J, Brownstein JS, Nsoesie E. A Digital Platform for Local Foodborne Illness and Outbreak Surveillance. *Online J Public Health Inform* 2016;8(1):e60. [DOI](#)
44. Hartley DM, Nelson NP, Arthur RR, Barboza P, Collier N, Lightfoot N, Linge JP, van der Goot E, Mawudeku A, Madoff LC, Vaillant L, Walters R, Yangarber R, Mantero J, Corley CD, Brownstein JS. An overview of internet biosurveillance. *Clin Microbiol Infect* 2013 Nov;19(11):1006–13. [DOI PubMed](#)
45. Yangarber R, Jokipii L, Rauramo A, Huttunen S. Extracting information about outbreaks of infectious epidemics. In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*; 2005 Oct 6 - Oct 8; Vancouver, Canada. Stroudsburg, PA: Association for Computational Linguistics; 2005. p. 22–3. <https://dl.acm.org/citation.cfm?id=1220575&picked=prox>
46. Zeng Z, Shi H, Wu Y, Hong Z. Survey of Natural Language Processing Techniques in Bioinformatics. *Comput Math Methods Med* 2015;2015:674296. [DOI PubMed](#)
47. Iroju OG, Olaleke JO. Information Technology and Computer Science. *Inf Technol Comput Sci*. 2015;08:44–50. [www.mecspress.org/DOI:10.5815/ijitcs.2015.08.07](http://www.mecspress.org/DOI:10.5815/ijitcs.2015.08.07)
48. Jordan S, Hovet S, Fung I, Liang H, Fu KW, Tsz Ho Tse Z. Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. *Data (Basel)* 2018;4(1):6. [DOI](#)
49. Liu F, Weng C, Yu H. *Clinical Research Informatics, Health Informatics*. pringer International Publishing; 2019. Chapter 17, Advancing Clinical Research Through Natural Language Processing on Electronic Health Records: Traditional Machine Learning Meets Deep Learning; p. 357–78. <https://link.springer.com/book/10.1007/978-3-319-98779-8>





50. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From linguistic signal to clinical reality; June 2014; Baltimore, MD: Association for Computational Linguistics; 2014. p. 51-60. DOI
51. Şerban O, Thapen N, Maginnis B, Hankin C, Foot V. Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Inf Process Manage* 2019;56(3):1166–84. DOI
52. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform* 2015 Feb;53:196–207. DOI PubMed
53. Sprugnoli R, Caselli T, Tonelli S, Moretti G. The Content Types Dataset: a New Resource to Explore Semantic and Functional Characteristics of Texts. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics; 2017 July 30 - Aug 4; Vancouver Canada. ACL; 2017. p. 260–6. <http://acl2017.org/>
54. Gritta M, Pilehvar MT, Collier N. A Pragmatic Guide to Geoparsing Evaluation: Toponyms, Named Entity Recognition and Pragmatics. 2018. Language Technology Lab, Department of Teoretical and Applied Linguistics, University of Cambridge, Cambridge UK. [www.academia.edu/38110706/A\\_Pragmatic\\_Guide\\_to\\_Geoparsing\\_Evaluation](http://www.academia.edu/38110706/A_Pragmatic_Guide_to_Geoparsing_Evaluation)
55. Ghosh S, Chakraborty P, Lewis BL, Majumder M, Cohn E, Brownstein JS, Marathe M, Ramakrishnan N. GELL: Automatic extraction of epidemiological line lists from open sources. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017 Aug 13- Aug 17; Halifax, Canada. New York USA: Association for Computing Machinery; 2017. p. 1477–86. DOI
56. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb;457(7232):1012–4. DOI PubMed
57. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) Google Flu Trends? *Am J Prev Med* 2014 Sep;47(3):341–7. DOI PubMed
58. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLOS Comput Biol* 2015 Oct;11(10):e1004513. DOI PubMed
59. Pollett S, Althouse BM, Forshey B, Rutherford GW, Jarman RG. Internet-based biosurveillance methods for vector-borne diseases: are they novel public health tools or just novelties? *PLoS Negl Trop Dis* 2017 Nov;11(11):e0005871. DOI PubMed
60. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. *PLoS Negl Trop Dis* 2014 Feb;8(2):e2713. DOI PubMed
61. Huang DC, Wang JF. Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. *Sci Total Environ* 2018 Jan;612:1293–9. DOI PubMed
62. Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, Merler S, Vespignani A. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect Dis* 2010 Jun;10:190. DOI PubMed
63. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJ. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *J Infect Dis* 2016 Dec;214 suppl\_4:S414–20. DOI PubMed
64. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA* 2009 Dec;106(51):21484–9. DOI PubMed
65. Arino J, Khan K. Analyzing and Modelling Spatial and Temporal Dynamics of Infectious Disease. Hoboken NJ: John Wiley & Sons; 2014. Chapter 5, Using Mathematical Modeling to Integrate Disease Surveillance and Global Air Transportation Data; p. 1–14.
66. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 2013 Dec;342(6164):1337–42. DOI PubMed
67. Bornman E. Information society and digital divide in South Africa: results of longitudinal surveys. *Inf Commun Soc* 2016;19. DOI
68. Nsoesie EO, Flor L, Hawkins J, Maharana A, Skotnes T, Marinho F, Brownstein JS. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? *PLoS Curr* 2016;8. DOI PubMed
69. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005 Nov;438(7066):355–9. DOI PubMed