

Evaluation d'un signalement d'agrégats de maladies dont une origine environnementale est suspectée

AIDE-MÉMOIRE

Janvier 2018

Sommaire

Mise en contexte	1
Objectif	2
Quelques définitions et contexte légal	2
Démarche d'évaluation	3
Limites	7
Conclusion	8
Bibliographie	9
Annexe	11

1 Mise en contexte

Les directions de santé publique (DSP) régionales et parfois le ministère de la Santé et des Services sociaux (MSSS) reçoivent de plus en plus de signalements d'agrégats spatio-temporels de maladies. La même situation est observée partout en Amérique du Nord. L'Institut national de santé publique du Québec (INSPQ) est à cet effet fréquemment interpellé afin de soutenir les DSP dans l'évaluation des agrégats de maladies.

Les objectifs épidémiologiques d'une démarche d'évaluation d'un agrégat spatio-temporel sont de déterminer s'il existe un excès statistiquement significatif de cas dans la population observée, et de déterminer s'il existe une ou plusieurs causes au regroupement des cas, autres que le hasard.

C'est dans ce contexte que le Comité méthodologique en analyses statistiques et épidémiologiques de l'INSPQ a mandaté les auteurs afin de formuler cet aide-mémoire concernant les démarches d'évaluation des agrégats de maladies d'origine environnementale suspectée.

La détection d'agrégats peut s'appliquer à différents contextes et pour différents objectifs (p. ex. : agrégats d'accidents routiers). Toutefois, ce document s'intéresse plus particulièrement à l'évaluation de signalements d'agrégats de maladies chroniques (incluant le cancer) dont une origine ou une source d'exposition environnementale est suspectée.

L'évaluation et la prise en charge des agrégats sont complexes et requièrent généralement une approche multidisciplinaire (médecine, épidémiologie, toxicologie, statistique, géographie, géomatique, démographie, sciences sociales, économie...).

L'évaluation d'agrégats spatio-temporels est un vaste champ d'études, lui-même niché dans le plus vaste domaine de l'épidémiologie spatiale (Carpenter, 2011).

La démarche d'évaluation d'agrégats a tout avantage à suivre une série d'étapes standardisées. Il s'agit ainsi d'une approche scientifique, nécessairement couplée à la gestion du risque (Cortin et collab., 2016) et d'un plan de communication.

Dans un premier temps, le présent aide-mémoire résume les étapes d'une évaluation d'agrégats, telles que définies dans le [Guide méthodologique pour l'évaluation et la prise en charge des agrégats spatio-temporels de maladies non infectieuses](#) de l'Institut de veille sanitaire (InVS) (Germonneau, Tillaut et Gomes Do Espirito Santo, 2005). Les mises à jour des lignes directrices des Centers for Disease Control and Prevention (CDC) aux États-Unis (CDC, 2013; CDC et The National Public Health Information Coalition, 2013) ainsi qu'en Australie (Department of Health, Western Australia, 2016), qui sont spécifiques aux agrégats de cancer, ont aussi été utilisées.

Dans un deuxième temps, sans pour autant constituer une revue de la littérature, l'évaluation d'un signalement d'agrégats spatio-temporels est complétée par la description de quelques méthodologies statistiques utiles à leur détection.

Finalement, plusieurs références bibliographiques sont fournies et permettront aux lecteurs d'approfondir certains éléments techniques qu'il serait trop long de présenter en détail dans cet aide-mémoire.

2 Objectif

Cet aide-mémoire est principalement destiné aux intervenants de santé environnementale et de surveillance des DSP, de l'INSPQ et du MSSS. Son objectif est de déterminer les meilleures pratiques épidémiologiques et statistiques pour l'évaluation d'un signalement d'agrégats spatio-temporels impliquant des maladies dont une origine environnementale est suspectée.

3 Quelques définitions et contexte légal

Un agrégat est un regroupement inhabituel de cas d'une même maladie dans une population, pour une zone géographique et une période de temps donnée. Il est suspecté dans une collectivité lorsque le nombre de personnes atteintes est perçu par le déclarant comme inhabituellement élevé. Dans le contexte d'une investigation d'un agrégat, l'analyse épidémiologique permet de confirmer son existence. Ces regroupements de cas sont appelés « agrégats spatio-temporels » ou « clusters », en anglais. Il s'agit souvent de regroupements de cas de cancers (Germonneau et collab., 2005).

Les causes des maladies, en particulier des cancers, sont multifactorielles : elles peuvent être génétiques, environnementales, professionnelles ou liées aux modes de vie. Le signalement d'un agrégat pose la question de l'existence d'une cause commune suspectée, à l'origine de la survenue de ces cas. Un tel regroupement peut aussi être lié uniquement au hasard.

Au Québec, l'origine légale des signalements d'agrégats réfère aux articles 92 à 94 de la [Loi sur la santé publique](#) (2001). Les ministères, les organismes gouvernementaux, les municipalités, les médecins, les établissements de santé et d'enseignement, les entreprises et les centres de la petite enfance doivent signaler au directeur de santé publique du territoire concerné ou au directeur national de santé publique les menaces à la santé de la population dont ils ont connaissance ou les situations qui leur donnent des motifs sérieux de croire que la santé de la population est menacée.

Il arrive également que des citoyens ou des regroupements de citoyens fassent de tels signalements de manière spontanée. Dans de tels cas, les articles 96 à 115 de la même loi confèrent au directeur de santé publique des pouvoirs d'enquête lui permettant de recueillir tous les renseignements qu'il juge pertinents afin de mener son enquête épidémiologique.

L'investigation est généralement menée par la DSP du lieu de résidence des individus visés par le signalement ou du lieu présumé de l'exposition.

4 Démarche d'évaluation

En 2005, l'InVS a publié un [guide](#) pour l'évaluation et la prise en charge des cas groupés de pathologies chroniques (Germonneau et collab., 2005). La démarche d'évaluation proposée consiste en un protocole standardisé de collecte et d'analyse de données. Elle a pour but de déterminer si une exposition commune à un facteur causal de la maladie est à l'origine du regroupement de cas, afin de guider les actions de protection de la santé publique.

Les objectifs épidémiologiques de la démarche d'évaluation d'un agrégat spatio-temporel sont de déterminer s'il existe un excès statistiquement significatif de cas dans la population observée, et de déterminer s'il existe une ou plusieurs causes au regroupement des cas, autres que le hasard.

Il serait inutile de reproduire intégralement les algorithmes d'évaluation des agrégats de l'InVS. Un résumé des quatre grandes étapes est proposé dans les sections 4.1 à 4.4. Le lecteur est invité à consulter la publication de l'InVS pour plus de détails (Germonneau et collab., 2005).

MISE EN GARDE

+ À la fin de chaque étape, la décision de poursuivre ou non l'évaluation de l'agrégat suspecté doit être prise. Qu'il y ait continuation ou non, la communication avec la population et le déclarant est une étape cruciale (CDC, 2013). L'élaboration et la mise en place d'un bon plan de communication sont des conditions de la réussite d'une évaluation d'agrégats. À ce propos, le lecteur est invité à prendre connaissance de la boîte à outils développée par les CDC qui présente plusieurs recommandations de la communication dans le cas d'une investigation d'agrégats (CDC et The National Public Health Information Coalition, 2013).

4.1 Analyse préliminaire du signalement d'agrégat (1^{er} point décisionnel)

La première étape consiste à prendre contact avec le déclarant afin de recueillir les informations factuelles sur les cas et l'exposition suspectés. Une description de l'agrégat doit être effectuée à partir des propos du déclarant en précisant les paramètres épidémiologiques classiques de temps, de lieux et de personnes. Une recherche bibliographique sommaire (en utilisant, si disponibles, les analyses critiques et les revues systématiques) concernant l'épidémiologie et les facteurs de risque connus et suspectés des maladies en cause devrait aussi être effectuée.

Plus spécifiquement pour le cancer, il est important de rappeler trois considérations générales (CDC, 2013) :

- Les facteurs étiologiques et prédisposants, les taux d'incidence et la prévalence varient selon le siège (l'organe cible) et le type (morphologie) de cancer;
- L'origine des cancers est souvent multifactorielle, et tous les mécanismes d'interaction ne sont pas connus;
- Pour la majorité des cancers, le temps de latence est long (délai entre l'exposition et le diagnostic ou les premiers symptômes) et complique l'investigation des agrégats dont une origine environnementale est suspectée.

Une investigation sommaire de la qualité de l'environnement, ainsi que des sources de contamination ou d'exposition suspectées, permettra de juger de la plausibilité de l'agrégat.

Au terme de cette étape préliminaire, une décision devra être prise en vue de la poursuite de l'investigation, en considérant notamment : le caractère rare et inhabituel des problèmes de santé; le nombre de cas et la probabilité que l'excès soit réel; la plausibilité que l'exposition suspectée génère les problèmes de santé rapportés; le potentiel d'extension. Cette première étape ne nécessite généralement pas de calcul statistique, mais plutôt un jugement épidémiologique et qualitatif sur la plausibilité de l'exposition à partir des informations recueillies.

4.2 Validation des cas et de l'exposition (2^e point décisionnel)

Cette étape est celle de la collecte des données concernant les cas et l'exposition. En ce qui concerne les données sanitaires, il s'agit dans un premier temps de confirmer les cas rapportés. Il importe d'identifier les sources de données permettant de confirmer les cas et les diagnostics. Les sources de données sont nombreuses, par exemple : le Registre des événements démographiques (RED; décès, naissances), le registre MED-ÉCHO (hospitalisations), le Système intégré de surveillance des maladies chroniques du Québec (SISMACQ) (Blais et collab., 2014; Gagnon, 2017) et le Registre québécois du cancer (RQC).

Le recours aux dossiers médicaux hospitaliers est une option qui permet de documenter de manière spécifique les cas. Lorsque le nombre de cas est restreint, les cas peuvent aussi être complétés à l'aide des dossiers médicaux obtenus auprès des professionnels de la santé (p. ex. : cliniques privées, clinique de médecine familiale, CLSC, etc.). Les objectifs sont de valider les diagnostics, d'établir une définition de cas précise et d'être le plus exhaustif possible concernant les cas potentiellement liés à l'agrégat.

Une collecte bien structurée des différentes informations personnelles et environnementales (voir les annexes 1 et 2 du [Guide de l'InVS](#)) peut être facilitée par l'utilisation de logiciels gratuits, comme [Epi Info](#). Rappelons l'importance de recueillir l'adresse de résidence (ou du lieu d'exposition) la plus précise possible. Autant que possible, il ne faut pas se contenter du code postal à six caractères, mais tenter d'obtenir l'adresse complète. De même, en fonction du contexte de l'agrégat, l'historique résidentiel constitue un incontournable.

C'est également à cette étape que la recherche de sources environnementales potentielles est effectuée. La validation de l'exposition environnementale suspectée peut être effectuée à partir, par exemple, de campagnes de mesure ou de données de surveillance (lorsque disponibles) : types de contaminants impliqués,

populations exposées, principales sources d'émissions, voies d'exposition, milieux contaminés, etc. Une revue de la littérature concernant la toxicologie des contaminants suspectés peut être envisagée, de même qu'une évaluation de la possibilité de contact des individus avec la source suspectée.

Parmi les sources de données environnementales possibles, l'[Inventaire national des rejets de polluants](#) (INRP) est disponible gratuitement et en ligne. Il peut être utile s'il s'agit d'émissions atmosphériques industrielles, dans l'eau ou le sol. Les données de surveillance de la qualité de l'air sont aussi disponibles à l'Infocentre de santé publique ou par le [Service Info-Air](#) du ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques (MDDELCC). Le [registre des terrains contaminés](#) du MDDELCC constitue une source à considérer. Enfin, la direction régionale du MDDELCC peut aussi constituer une source intéressante d'information (eau, air, sol).

Après avoir défini et validé les cas, si un surplus de cas semble encore présent dans la zone d'étude, il s'agit de déterminer si cette collectivité est plus exposée qu'une population de référence à un ou plusieurs facteurs de risque connus ou suspectés de la maladie.

L'analyse des données concernant le type de pathologies, la validation des cas, la présence d'une exposition suspecte et la plausibilité d'une relation entre l'exposition et la maladie devrait permettre de décider de la suite du processus. Si le résultat de l'évaluation conduit à un arrêt de l'investigation, un rapport est produit et diffusé aux populations concernées.

La mise en évidence d'un excès de cas et d'une surexposition conduit à mener une étude épidémiologique plus approfondie (étape 4.3). Cette dernière est destinée à rechercher la ou les causes principales des cas et à établir si l'excès de cas suspecté est statistiquement significatif. Si aucune source environnementale n'est identifiée, une campagne de mesure peut être envisagée, de même, la description des expositions et des facteurs de risque individuels est à considérer (voir section 4.3.2).

4.3 Étude épidémiologique descriptive approfondie

L'analyse épidémiologique descriptive approfondie consiste à appliquer des méthodes statistiques afin d'évaluer l'impact de la contamination environnementale sur la santé. Il est important de bien délimiter la zone et la période à l'étude, puis de réviser la définition de cas (voir étape précédente) et de les recenser de la manière la plus exhaustive possible.

Il faut définir quelles études devront être utilisées et décider de la séquence de réalisation de celles-ci. L'analyse de la puissance statistique des différentes possibilités d'études en guidera la priorité. Une équipe d'investigation multidisciplinaire est mise sur pied à cet effet et un protocole complet est rédigé. De même, s'il n'a pas été mis en place précédemment, un comité de suivi devrait être mis sur pied afin de bien encadrer les échanges avec la population et de répondre à leurs questions et préoccupations. Ce comité est distinct de l'équipe d'investigation. Sa composition devrait inclure des représentants de la communauté, des élus (p. ex. municipaux) et des scientifiques.

4.3.1 CONFIRMATION DE L'EXCÈS DE CAS

Afin de confirmer l'excès de cas, la fréquence des cas peut être comparée à des données de référence (p. ex. RQC, registre MED-ÉCHO, RED, etc.). L'objectif est de vérifier si le nombre de cas observé est supérieur à celui attendu dans l'hypothèse d'une absence d'exposition environnementale spécifique dans la collectivité où les cas sont survenus. Il importe, avant tout, de bien définir les cas (réviser la définition) et de les rechercher de manière exhaustive.

Afin de déterminer si la fréquence des cas est anormalement élevée, l'utilisation de ratios standardisés de mortalité (*Standardized Mortality Ratio* [SMR]) ou d'incidence (*Standardized Incidence Ratio* [SIR]) est à privilégier s'il s'agit d'un problème peu fréquent (maladie rare), qui plus est dans un territoire peu peuplé. Il s'agit en fait de la méthode de standardisation indirecte largement connue en épidémiologie (Bernard et Lapointe, 1998). Le calcul d'un intervalle de confiance permettra de juger d'une différence significative ou non entre la fréquence relative des cas observés et attendus (Bouyer et collab., 2009; Breslow et Day, 1987).

La standardisation indirecte exige que le nombre de cas observé ainsi que les effectifs de population dans le territoire à l'étude soient disponibles par groupe d'âge (et sexe). Il faut également que les taux spécifiques de la population de référence soient disponibles.

L'utilisation de la standardisation directe pourra aussi être envisagée si les taux spécifiques par groupe d'âge (et sexe) dans la population étudiée sont fiables (stabilité statistique). Cette méthode est à privilégier pour des comparaisons de la fréquence de la maladie dans de plus vastes territoires ou pour des problèmes de santé fréquents. Un exemple d'utilisation de la standardisation directe pour l'analyse d'un agrégat de cancer suspecté et associé à une contamination environnementale de l'eau potable aux États-Unis a récemment été publié (Press, McKinley, Deapen, Clarke et Gomez, 2016).

MISE EN GARDE

Sur le plan théorique, il n'est pas possible de comparer deux SMR ou deux SIR. Une telle comparaison doit être envisagée avec un œil critique (Bouyer et collab., 2009; Rothman, Greenland et Lash, 2012). En effet, la standardisation indirecte ne devrait pas être utilisée pour comparer la fréquence d'une maladie dans plusieurs unités géographiques, et ce, même si une même population de référence est utilisée. De même, les SMR et SIR ne devraient pas être utilisés dans des représentations cartographiques, et la standardisation directe devrait être priorisée.

Dans certains contextes, il est parfois requis de déterminer si les cas sont localisés en grappes sur le territoire. Dans un tel cas, l'objectif est de déterminer s'il existe des regroupements statistiquement significatifs de cas dans l'espace. L'annexe 1 présente certains des outils et des méthodes statistiques disponibles afin de déterminer la présence d'agrégats spatiaux.

Il est aussi possible de documenter l'évolution temporelle des cas. Il s'agit alors de déterminer si, pendant une période donnée, la fréquence des cas correspond à ce qui est attendu en considérant une distribution type. La fréquence des cas dans le temps pendant la période d'étude est alors comparée à une période de référence. Le logiciel *SatScan* permet de vérifier de telles hypothèses et de repérer des périodes où la fréquence est anormale. De multiples options sont

possibles en fonction de la disponibilité des données. Les résultats de ces analyses sont fortement influencés par la qualité et la stabilité de la série de données de comparaison.

L'analyse temporelle peut aussi vérifier la présence d'une tendance à la hausse ou à la baisse sur une longue période. Selon la fréquence de la maladie, il sera possible d'effectuer les tests sur une base hebdomadaire, mensuelle ou annuelle. Un outil possible pour réaliser ces tests de tendance est *JoinPoint* (Joinpoint Regression Program - Surveillance Research Program, s.d.). Ce programme permet de modéliser des taux bruts ou standardisés avec un modèle de régression par période, et donc d'identifier des points de rupture temporels.

En fonction des résultats précédents, il peut être nécessaire de déterminer si les cas sont statistiquement regroupés à la fois dans l'espace et dans le temps. Encore une fois, cette analyse nécessite de bonnes données de comparaison et une localisation géographique précise des cas. L'analyse statistique peut être menée à l'aide de *SatScan*. Plusieurs paramètres relatifs, notamment à la distance et au délai temporel entre les cas et aux modèles statistiques, doivent être déterminés.

Le lecteur intéressé à explorer les concepts et méthodes de détection d'agrégats spatio-temporels peut se référer à d'autres publications, dont : Gaudart et collab., 2007; Gorla, Stempfelet et De Crouy-Chanel, 2011; Jacquez, 2008; Waller et Gotway, 2004.

4.3.2 DESCRIPTION DES CAS ET DE L'EXPOSITION

Des entrevues auprès des cas ou de leur famille peuvent être envisagées afin de documenter une exposition partagée, ainsi que les facteurs de risque. Un exemple de questionnaire est fourni en annexe 2 du [Guide de l'InVS](#). En fonction du contexte de l'investigation, la décision de mener de telles entrevues est délicate (biais de désirabilité).

4.3.3 CARACTÉRISATION DE L'EXPOSITION

L'exposition peut être quantifiée de façon directe (marqueurs d'exposition) ou indirecte (marqueur de proxy de l'exposition). L'exposition de la population étudiée peut être comparée à celle de la population de référence ou au bruit de fond (présence naturelle ou généralement acceptée d'un contaminant dans un milieu).

C'est aussi à cette étape où l'on tentera de quantifier les risques à la santé. Cette quantification des risques est sujette à la capacité de quantifier l'intensité et la durée de l'exposition de la population à l'étude et à l'existence de valeurs toxicologiques de référence.

4.3.4 ÉVALUATION DE LA PLAUSIBILITÉ DU LIEN ENTRE L'EXPOSITION ET LA MALADIE (3^e POINT DÉCISIONNEL)

L'équipe d'investigation analyse les données colligées et décide de la mise en œuvre des mesures de gestion ou de la mise en place d'une étude épidémiologique de type analytique (voir l'étape 4.4). Les mesures de gestion du risque visent à réduire les risques à la santé, notamment par le retrait de l'exposition ou par la mise en place d'une surveillance active des cas potentiels. Le cadre de référence de gestion des risques en santé publique (Cortin et collab., 2016) peut également servir de guide à ce propos.

En fonction de la présence d'un agrégat statistiquement significatif et d'une exposition jugée significative, un tableau d'aide à la décision est proposé dans le [document de l'InVS](#) (p. 40). En résumé :

- en l'absence d'un agrégat et d'exposition significative, l'investigation est terminée;
- en l'absence d'un agrégat, mais en présence d'une exposition significative, des recommandations pour la gestion du risque sont effectuées;
- en l'absence d'exposition significative et en présence d'un agrégat, une surveillance épidémiologique doit être envisagée;
- en présence d'une exposition significative et d'un agrégat, la gestion du risque doit être effectuée et la possibilité de réaliser une étude épidémiologique de type analytique devra être évaluée (voir l'étape 4.4).

D'autres éléments pertinents à la prise de décision sont aussi présentés dans les lignes directrices des CDC (2013).

4.4 Compléments épidémiologiques en présence d'une hypothèse étiologique

L'objectif général de cette étape est de rechercher une relation entre le facteur de risque suspecté et la pathologie en cause. Ce type d'analyse ne devrait être effectué qu'après avoir confirmé objectivement l'excès de cas, l'existence d'une exposition significative et la plausibilité biologique d'une association entre la maladie et l'exposition. Les critères de causalité devraient être considérés dans cette analyse.

Une étude épidémiologique de type analytique en bonne et due forme devrait être mise en place, incluant la rédaction d'un protocole. L'hypothèse de recherche devrait être clairement identifiée. L'équipe multidisciplinaire déterminera la faisabilité, les coûts, le design, la puissance statistique, ainsi que les mesures d'association applicables selon le type d'étude épidémiologique et les données disponibles. Les difficultés méthodologiques sont nombreuses. L'étude de faisabilité devrait inclure des estimations de la puissance statistique.

Précisons qu'en présence d'un excès de cas, des mesures de surveillance et de suivi épidémiologique méritent d'être envisagées, même en l'absence d'une association biologique plausible avec l'exposition suspectée ou lorsque les conditions minimales pour entreprendre une étude (p. ex. un nombre de cas trop restreint) ne sont pas satisfaites.

5. Limites

Les principales difficultés d'ordre méthodologique rencontrées lors d'une évaluation d'agrégats se rapportent, d'une part, à la qualité et à l'exhaustivité des cas et des données de comparaison, et, d'autre part, à la disponibilité de bonnes données d'exposition.

Il est parfois difficile d'obtenir de bonnes estimations de la population exposée. La disponibilité des coordonnées géographiques les plus précises possible du lieu de résidence ou du lieu de l'exposition est le talon d'Achille des évaluations d'agrégats, parce qu'elles sont, en règle générale, difficiles à obtenir. Le choix du groupe de comparaison est important et dépend du contexte de l'investigation. Il faut chercher à établir les comparaisons les plus stables possible sur le plan statistique (p. ex. comparativement à la région sociosanitaire ou à la province ou encore comparer les données hebdomadaires par rapport aux mêmes semaines des cinq années antérieures, etc.).

De plus, la quantification de l'exposition est souvent rudimentaire, surtout lorsqu'il s'agit d'une exposition passée, qui plus est n'a pas été mesurée objectivement. Dans le cas des cancers, l'importance de la période de latence, qui est souvent de plusieurs dizaines d'années, constitue un enjeu incontournable. L'historique résidentiel est rarement disponible, ce qui limite aussi l'identification des individus ou des groupes exposés et non exposés. Il est aussi possible que l'exposition ne soit pas liée au lieu de résidence, mais au lieu de travail par exemple.

Différentes approches méthodologiques ont été expérimentées afin de tenter de contourner certains de ces problèmes (Goodman et collab., 2014). On a déjà utilisé le Fichier d'identification des personnes assurées (FIPA) de la Régie de l'assurance maladie du Québec (RAMQ) afin de contourner l'absence d'historique résidentiel (Nourissat, Duchesnes, Riffon et Levaque, 2016). La modélisation, l'interpolation ou le lissage (comme le krigeage [Iglesias, 2017]) sont parfois utilisés afin d'obtenir des estimations de la répartition de l'exposition environnementale. La qualité des interpolations est dépendante de la qualité (et de la quantité) des données de base utilisées pour faire la modélisation. Enfin, l'investigation à l'intérieur de petites unités géographiques pose le problème des petits

effectifs et du biais écologique¹, ce qui limite les analyses statistiques et épidémiologiques possibles.

Parmi la multitude de tests et de méthodes statistiques disponibles, il peut être difficile de sélectionner la méthode la mieux adaptée au contexte spécifique de l'investigation. C'est parfois la convergence des résultats de quelques approches différentes qui permettra d'infirmer ou de confirmer la présence d'un agrégat.

L'analyse de la répartition géographique des maladies et des associations avec des facteurs de risque (exposition) est souvent basée sur des données agrégées. Les biais écologiques sont propres à ce type d'études épidémiologiques. La présence de tels biais doit être considérée lors de l'évaluation d'un agrégat suspecté, notamment parce qu'ils peuvent entraîner une erreur d'estimation du degré d'association entre l'exposition et l'effet.

6. Conclusion

Cet aide-mémoire se veut un outil de référence aux intervenants du réseau de la santé publique qui font face à un signalement d'agrégat. Il ne constitue qu'une courte introduction à une approche systématique. La lecture des documents de base de l'InVS, des CDC et du Département de santé australien est fortement conseillée.

La démarche d'évaluation d'un agrégat dont une origine environnementale est suspectée est complexe et présente plusieurs défis pour les autorités de santé publique. Elle implique une série d'étapes successives de même que le recours à une équipe multidisciplinaire et à des données parfois difficiles à recueillir.

La démarche d'investigation d'un agrégat n'ayant pas évolué de manière substantielle dans la dernière décennie, le [guide de l'InVS](#) demeure une référence standard pour planifier une telle intervention. Par contre, comme les outils et méthodes ont évolué, il est souhaitable de compléter le guide de l'InVS par d'autres documents plus récents, dont tout particulièrement les guides des CDC (2013) et du Département de santé

australien (2016). L'adaptation aux sources de données disponibles au Québec doit être effectuée.

Le [cadre de référence pour la gestion des risques en santé publique au Québec](#) (Cortin et collab., 2016) est complémentaire à l'évaluation des agrégats suspectés. Ce cadre vise le développement d'une compréhension commune de l'évaluation et de la gestion des risques en santé publique au Québec. Il propose un processus systématique pour évaluer et gérer les risques à la santé de la population, souvent marqués par l'incertitude, la complexité ou l'ambiguïté. Ce processus vise à accompagner les intervenants de santé publique pour structurer les actions de prévention et de protection. De plus, l'élaboration d'un plan de communication, dès le début de l'évaluation, constitue un autre élément menant au succès d'une évaluation d'agrégats.

Il est rare que les quatre étapes de l'évaluation d'un agrégat soient complétées. Seulement un petit nombre de signalements d'agrégats satisfait les critères statistiques et étiologiques nécessitant une étude épidémiologique de type analytique.

¹ Le biais écologique est un biais propre aux études réalisées à partir de données agrégées. Celui-ci résulte : du biais d'agrégation (les individus qui ont présenté un effet ne sont pas forcément ceux qui ont été exposés); du biais de spécification, ou du biais de confusion écologique. Pour en savoir plus : http://invs.santepubliquefrance.fr/publications/pol_atmo1/page3.html

7. Bibliographie

- Banerjee, S. (2016). Spatial Data Analysis. *Annual Review of Public Health*, 37(1), 47- 60. doi:10.1146/annurev-publhealth-032315-021711
- Bernard, P.-M. et Lapointe, C. (1998). Mesures statistiques en épidémiologie. Québec : Presses de l'Université du Québec.
- Blais, C., Jean, S., Sirois, C., Rochette, L., Plante, C., Larocque, I., et collab. (2014). Le Système intégré de surveillance des maladies chroniques du Québec (SISMACQ), une approche novatrice. *Maladies chroniques et blessures au Canada*, 34(4), 247-56.
- Bouyer, J., Hémon, D., Cordier, S., Derriennic, F., Stücker, I., Stengel, B. et Claver, J. (2009). *Épidémiologie, Principes et méthodes quantitatives*. Paris.
- Breslow, N. E. et Day, N. E. (1987). *Statistical methods in cancer research, Vol II The design and analysis of cohort studies (vol. 1-v)*. Lyon. : International Agency for Research on Cancer, World Health Organization.
- Bull, M., Hall, I. M., Leach, S. et Robesyn, E. (2012). The application of geographic information systems and spatial data during Legionnaires disease outbreak responses. *Euro Surveill*, 17, 20331.
- Carpenter, T. E. (2011). The spatial epidemiologic (r)evolution: A look back in time and forward to the future. *Spatial and Spatio-temporal Epidemiology*, 2(3), 119-124. doi:10.1016/j.sste.2011.07.002
- Centers for Disease Control and Prevention et The National Public Health Information Coalition. (2013, septembre). Cancer Clusters : A toolkit for communicators. Repéré à <https://www.nphic.org/toolkits/cancer-cluster>
- Centers for Disease Control and Prevention. (2013). Investigating Suspected Cancer Clusters and Responding to Community Concerns: Guidelines from CDC and the Council of State and Territorial Epidemiologists. *MMWR*, 62(8), 1-24.
- Cortin, V., Laplante, L., Dionne, M., Filiatrault, F., Laliberté, C., Lessard, P., ... Pouliot, B. (2016). La gestion des risques en santé publique au Québec : cadre de référence (p. 87). Institut national de santé publique du Québec. Repéré à <http://collections.banq.qc.ca/ark:/52327/2563951>
- Department of Health, Western Australia. (2016). Guidelines for the investigation of cancer clusters in Western Australia. Repéré à <http://ww2.health.wa.gov.au/~-/media/Files/Corporate/general%20documents/Epidemiology/Cancer%20Cluster%20Guidelines%20PDF.ashx>
- Gagnon, R. (2017). Cadre de qualité des données du Système intégré de surveillance des maladies chroniques du Québec : rapport méthodologique. Montréal : Institut national de santé publique du Québec.
- Gaudart, J., Giorgi, R., Poudiougou, B., Toure, O., Ranque, S., Doumbo, O. et Demongeot, J. (2007). Détection de clusters spatiaux sans point source prédéfini : utilisation de cinq méthodes et comparaison de leurs résultats. *Rev Epidemiol Sante Publique*, 55, 297-306.
- Germonneau, P., Tillaut, H. et Gomes Do Espirito Santo, E. (2005). Guide méthodologique pour l'évaluation et la prise en charge des agrégats spatio-temporels de maladies non infectieuses (p. 77). Paris. Repéré à http://invs.santepubliquefrance.fr//publications/2005/guide_ast/guide.pdf
- Goodman, M., LaKind, J. S., Fagliano, J. A., Lash, T. L., Wiemels, J. L., Winn, D. M., ... Mattison, D. R. (2014). Cancer Cluster Investigations : Review of the Past and Proposals for the Future. *International Journal of Environmental Research and Public Health*, 11(2), 1479-1499.
- Goria, S. et Le Tertre, A. (2010). Les études locales autour d'un point source - Les différentes méthodes statistiques, leurs avantages et leurs inconvénients. Note méthodologique. (p. 8). Institut de Veille Sanitaire. Repéré à http://opac.invs.sante.fr/doc_num.php?explnum_id=241
- Goria, S., Stempfelet, M. et De Crouy-Chanel, P. (2011). Introduction aux statistiques spatiales et aux systèmes d'information géographique en santé environnement, Application aux études écologiques - Résultats 2010. Institut de veille sanitaire (InVS). Repéré à http://opac.invs.sante.fr/doc_num.php?explnum_id=6974

Iglesias, M. A. C. (2017, 4 février). Kriging Interpolation - The Prediction Is Strong in this One. GIS Geography. Repéré à <http://gisgeography.com/kriging-interpolation-prediction/>

Jacquez, G. M. (2008). Spatial cluster analysis. Dans *The handbook of geographic information science* (p. 395-416). S. Fotheringham and J. Wilson (Eds.).

Joinpoint Regression Program - Surveillance Research Program. (s.d.). Repéré à <https://surveillance.cancer.gov/joinpoint/>

Lawson, A. B. (2012). Bayesian point event modeling in spatial and environmental epidemiology. *Statistical Methods in Medical Research*, 21(5), 509-529. doi:10.1177/0962280212446328

Lee, J., Gangnon, R. E. et Zhu, J. (2017). Cluster detection of spatial regression coefficients. *Statistics in Medicine*, 36(7), 1118-1133. doi:10.1002/sim.7172

Loi sur la santé publique. (2001). LRQ c S-2.2. Repéré à <http://legisquebec.gouv.qc.ca/fr/ShowDoc/cs/S-2.2>

Nourissat, A., Duchesnes, J.-F., Riffon, J. et Levaque, R. (2016). Étude de l'incidence des cas de cancer du cerveau, du rein, du foie et de lymphome non hodgkinien chez les personnes ayant habité la municipalité de Shannon (Québec, Canada) entre 1987 et 2001. Direction de la santé publique du Centre intégré de santé et de services sociaux de la Capitale-Nationale. Repéré à <https://www.ciuss-capitalenationale.gouv.qc.ca/etude-de-lincidence-des-cas-de-cancer-du-cerveau-du-rein-du-foie-et-de-lymphome-non-hodgkinien-chez>

Press, D. J., McKinley, M., Deapen, D., Clarke, C. A. et Gomez, S. L. (2016). Residential cancer cluster investigation nearby a Superfund Study Area with trichloroethylene contamination. *Cancer Causes & Control*, 27(5), 607-613. doi:10.1007/s10552-016-0734-5

Rothman, K. J., Greenland, S. et Lash, T. L. (2012). *Modern Epidemiology* (3 édition). Philadelphia Baltimore New York London Buenos Aires Hong Kong Sydney Tokyo : Wolters Kluwer.

Waller, L. A. et Gotway, C. A. (2004). *Applied spatial statistics for public health data*. Wiley. Repéré à <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471387711.html>

Membres actifs du Comité méthodologique en analyses statistiques et épidémiologiques de l'INSPQ

Gilles Paradis, président
Vice-présidence aux affaires scientifiques

Denis Hamel, coordonnateur
Bureau d'information et d'études en santé des populations

Louis Rochette, membre
Bureau d'information et d'études en santé des populations

Ernest Lo, membre
Bureau d'information et d'études en santé des populations

Sonia Jean, membre
Bureau d'information et d'études en santé des populations

Véronique Boiteau, membre
Bureau d'information et d'études en santé des populations

Christophe Garenc, membre
Direction des risques biologiques et de la santé au travail

Alfreda Krupoves, membre
Direction des risques biologiques et de la santé au travail

Nathalie Vandal, membre
Bureau d'information et d'études en santé des populations

Germain Lebel, membre
Direction de la santé environnementale et de la toxicologie

Benoît Lasnier, membre
Direction du développement des individus et des communautés

Marc-André Dubé, membre
Bureau d'information et d'études en santé des populations

Vanessa Fillion, secrétaire et soutien professionnel
Bureau d'information et d'études en santé des populations

Annexe Quelques outils et méthodes d'analyse spatiale

En faisant abstraction de la dimension temporelle des données, l'analyse typique d'un agrégat spatial consiste à comparer la distribution géographique des cas par rapport à une distribution aléatoire ou à une distribution typique des cas. Une première approche analyse globalement la distribution spatiale (ou temporelle) d'une maladie. On retrouve de nombreux tests dans cette première approche, qui permettent d'estimer la corrélation spatiale des données dans l'ensemble de la région étudiée (une valeur unique est obtenue). La deuxième approche consiste à déterminer la localisation géographique d'agrégats. Les tests d'agrégats locaux sont de deux types : (i) sans a priori quant à la présence d'une source et (ii) avec la présence d'une source. Dans ce dernier cas, on parle des tests focalisés qui visent à estimer le risque d'une maladie par rapport à une source ponctuelle prédéterminée.

Les méthodes globales : détection d'agrégats spatio-temporels

L'analyse de la localisation d'agrégats spatio-temporels peut être réalisée à l'aide de logiciels spécialisés gratuits, par exemple *SatScan*, développé pour les Centers for Disease Control and Prevention (CDC) aux États-Unis. Il existe également plusieurs produits commerciaux (p. ex. *Cluster Seer*). *SatScan* est très polyvalent; il peut être utilisé à la fois pour des distributions de données discrètes et continues ainsi que pour l'analyse globale ou locale. Selon le type de données et le type d'étude envisagé, il permet d'utiliser différentes distributions statistiques (p. ex. loi de Poisson [données agrégées ou petit nombre de cas par rapport à la population], loi binomiale [données individuelles des cas et des témoins], loi de Bernoulli, loi multinomiale, etc.). Dans le cas de données unitaires, il va sans dire que l'utilisation de *SatScan* requiert les coordonnées géographiques des cas. Dans le cas des données agrégées, les coordonnées des centroïdes des polygones sont requises. Selon le type d'analyse effectuée, les effectifs de la population qui réside à l'intérieur des polygones sont aussi nécessaires. De même, des covariables (p. ex. sexe, groupe d'âge, facteurs socioéconomiques, densité de la population, etc.) peuvent être considérées dans les analyses.

Une utilisation type de *SatScan* consiste à détecter la présence d'agrégats à partir de données unitaires. L'analyse consiste à tracer des cercles de rayons fixes prédéterminés (ou de rayons variables), en utilisant les coordonnées géographiques de chaque cas, puis de comparer le nombre de cas observés (à l'intérieur des cercles) au nombre de cas attendus. Des simulations de Monte-Carlo permettent de calculer la valeur-*p* de chaque agrégat identifié, en testant l'hypothèse nulle de l'homogénéité spatiale des risques (Goria et collab., 2011; Lee, Gangnon et Zhu, 2017). Les résultats des agrégats statistiquement significatifs obtenus de *SatScan* peuvent être visualisés avec Google Earth, ou exportés dans un logiciel système d'information géographique comme QGIS ou ARCGIS : <https://www.satscan.org/tutorials/nysbirthdefect/SaTScanTutorialNYSBirthDefect.pdf>.

D'autres techniques permettent aussi de tester la présence d'agrégats en utilisant des formes géométriques différentes d'un cercle (p. ex. des ellipses). De même, des modèles bayésiens peuvent être utilisés. Ils permettent notamment de tester une répartition différentielle des risques sur le territoire (Banerjee, 2016; Lawson, 2012; Lee et collab., 2017).

Plus rarement applicable au contexte d'une évaluation d'agrégats de maladies, l'analyse peut être effectuée afin d'estimer l'association géographique entre des taux d'incidence ou de mortalité et la répartition géographique d'un facteur de risque (Lee et collab., 2017). La méthode de régression spatiale requiert l'analyse des données agrégées de la maladie et du facteur de risque sur la base d'unités géographiques les plus fines possible. Des analyses de régression spatiale (régression de Poisson) sont aussi possibles afin de considérer simultanément les contributions de plusieurs variables (p. ex. exposition environnementale et facteurs socioéconomiques) à la variance de la distribution géographique d'une maladie ou de la partie résiduelle (non expliquée). La mesure et l'ajustement de l'autocorrélation spatiale (dans chacune des séries de données) doivent être effectués. L'autocorrélation spatiale est l'absence d'indépendance entre les valeurs observées dans les unités géographiques. Il s'agit en fait de la propension naturelle des unités géographiques adjacentes à présenter des valeurs qui sont fortement

corrélées. En présence d'autocorrélation spatiale, les postulats de base de l'analyse de corrélation ou de régression ne sont pas respectés et des ajustements doivent être appliqués. Différentes mesures de l'autocorrélation existent, les plus courantes étant le I de Moran et le C de Geary. Le I de Moran est plus répandu parce que plus stable. La mesure de l'autocorrélation spatiale exige l'élaboration d'une matrice de contiguïté (voisinage) des unités géographiques. La mesure de l'autocorrélation spatiale est disponible dans Cluster Seer, ainsi que dans les systèmes d'information géographique, notamment dans QGIS. Il existe également des modèles de régression qui introduisent l'autocorrélation spatiale et l'hétérogénéité spatiale (variation de la relation entre les variables sur le territoire). Par exemple, le modèle de régression géographiquement pondéré (variante du modèle classique de régression linéaire) génère une équation distincte pour chaque entité dans le jeu de données qui incorpore les variables dépendantes et explicatives. Cela permet un meilleur ajustement local des données.

Ce type d'analyse est complexe; le lecteur intéressé trouvera toute la documentation utile sur le site Internet suivant :

<http://www.spatialanalysisonline.com/HTML/index.html>

Les méthodes locales : les tests locaux focalisés

Une autre possibilité d'analyse consiste à vérifier l'hypothèse d'un regroupement des cas à proximité d'une source connue ou suspectée de contamination. Il s'agit des tests focalisés ou de concentration. Ils servent à mettre en évidence une augmentation du taux d'incidence ou de mortalité de populations qui résident à proximité de zone contaminée. Dans certains cas, l'analyse vise à déterminer si les cas sont localisés de manière structurée (c.-à-d. plus près de la source suspectée), comparativement à des témoins ou à des non-malades. Le choix des témoins est d'une importance primordiale. L'analyse consiste à comparer la proximité des cas aux non-malades. La distance est alors considérée comme un indicateur de l'exposition. Ces techniques ont été utilisées dans l'analyse des cas de légionellose à proximité de tours de refroidissement (Bull, Hall, Leach et Robesyn, 2012).

Pour plus de détails sur les tests focalisés, dont les avantages et inconvénients, le lecteur est invité à prendre connaissance de la [note méthodologique de l'InVS](#) (Goria et Le Tertre, 2010).

Évaluation d'un signalement d'agrégats de maladies dont une origine environnementale est suspectée – Aide-mémoire

AUTEURS

Germain Lebel, M. A., M. Sc., conseiller scientifique
Direction de la santé environnementale et de la toxicologie

Marie-Claude Boivin, M. Sc., agente de recherche
Bureau d'information et d'études en santé des populations

Marjolaine Dubé, B. Sc., statisticienne
Direction de la santé environnementale et de la toxicologie

AVEC LA COLLABORATION DE

Linda Perron, médecin spécialiste
Bureau d'information et d'études en santé des populations

RÉVISEUR

Benoît Lévesque, M.D., M. Sc., FRCPC, médecin spécialiste
Direction de la santé environnementale et de la toxicologie

Comité méthodologique en analyses statistiques et épidémiologiques
de l'Institut national de santé publique du Québec

MISE EN PAGE

Nabila Haddouche
Bureau d'information et d'études en santé des populations

Ce document est disponible intégralement en format électronique (PDF) sur le site Web de l'Institut national de santé publique du Québec au : <http://www.inspq.qc.ca>.

Les reproductions à des fins d'étude privée ou de recherche sont autorisées en vertu de l'article 29 de la Loi sur le droit d'auteur. Toute autre utilisation doit faire l'objet d'une autorisation du gouvernement du Québec qui détient les droits exclusifs de propriété intellectuelle sur ce document. Cette autorisation peut être obtenue en formulant une demande au guichet central du Service de la gestion des droits d'auteur des Publications du Québec à l'aide d'un formulaire en ligne accessible à l'adresse suivante : <http://www.droitauteur.gouv.qc.ca/autorisation.php>, ou en écrivant un courriel à : droit.auteur@cspq.gouv.qc.ca.

Les données contenues dans le document peuvent être citées, à condition d'en mentionner la source.

Dépôt légal – 1^{er} trimestre 2018
Bibliothèque et Archives nationales du Québec
ISBN : 978-2-550-80636-3 (PDF)

© Gouvernement du Québec (2018)

N° de publication : 2357