
Brève. Détection statistique d'événements inhabituels à partir d'un système de surveillance

Short item. Statistical detection of unusual events from a surveillance system

Yann Le Strat (y.lestrat@invs.sante.fr)
Institut de veille sanitaire, Saint-Maurice, France

Mots-clés: événements inhabituels, détection, surveillance/Keywords: Unusual events, Detection, Surveillance

Un des objectifs de la surveillance épidémiologique est la détection de phénomènes inhabituels à des fins d'alerte sanitaire. Nous nous plaçons dans le cadre d'une surveillance continue, spécifique ou non, constituée d'un ensemble de déclarants (médecins, biologistes, vétérinaires, etc.) rapportant des événements de santé au système de surveillance. Pour simplifier, nous utiliserons le terme « cas » pour désigner ces événements de santé (infections, décès, etc.). Un événement inhabituel est l'observation d'un nombre de cas rapportés au système de surveillance, significativement plus élevé que le nombre de cas attendu. Cette définition est très proche de la définition classique d'une épidémie. Elle englobe des cas liés épidémiologiquement, mais aussi des cas sporadiques n'ayant pas de lien entre eux. La distribution des cas dans le temps et l'espace définissent trois types de clusters: les clusters temporels, spatiaux et spatio-temporels. Un cluster temporel est un intervalle de temps dans lequel le nombre de cas est anormalement élevé – ces cas étant dispersés géographiquement (par ex. une épidémie nationale de salmonellose). Un cluster spatial est une zone géographique dans laquelle le nombre de cas est anormalement élevé, l'occurrence des cas s'étalant dans le temps (sur plusieurs mois, voire même plusieurs années – c'est le cas de certains cancers par exemple). Enfin, un cluster spatio-temporel est un intervalle de temps et une zone géographique dans lesquels le nombre de cas est anormalement élevé – comme les toxi-infections alimentaires collectives par exemple. La mise en évidence le plus précocement possible de ces clusters permet, en cas de besoin, de mettre en œuvre des investigations et des mesures de contrôle (par ex. le retrait du marché d'un produit contaminé) pour limiter l'ampleur d'un phénomène épidémique. Devant l'augmentation du nombre de données à analyser et l'augmentation de la fréquence des analyses, la détection précoce s'appuie sur des méthodes statistiques produisant des signaux statistiques. Ces signaux rejoignent d'autres signaux épidémiologiques qui aident l'épidémiologiste à enclencher ou non des investigations.

Méthodes statistiques de détection

Nous distinguons les méthodes de détection purement temporelles des méthodes de détection spatiales ou spatio-temporelles.

Détection temporelle

La majorité des méthodes de détection temporelle repose d'abord sur l'estimation du nombre attendu de cas, puis sur un test statistique permettant de déterminer si le nombre de cas rapportés est significativement supérieur à ce nombre attendu.

L'estimation du nombre attendu est basée sur la prédiction d'un modèle, le plus souvent un modèle de régression, développé à partir des nombres de cas rapportés dans le passé. Les tendances, saisonnalités et surdispersions, si elles existent, doivent être prises en compte pour réaliser des estimations fiables du nombre attendu et de son intervalle de prédiction. Des régressions de Poisson, de quasi-Poisson ou binomiale négative peuvent être envisagées (Enki *et al.* 2013).

Le seuil épidémique est défini par la borne supérieure de l'intervalle de prédiction du nombre attendu. Si le nombre de cas rapportés est supérieur à cette borne supérieure, on peut statistiquement affirmer l'existence d'un cluster temporel. La valeur du seuil est estimée chaque fois que la méthode de détection est exécutée (le plus souvent chaque semaine ou chaque jour).

De nombreuses méthodes ont été développées depuis les années 1990 (Farrington *et al.* 2004; Unkel *et al.* 2012). Elles font maintenant partie des outils indispensables à une surveillance de qualité intégrant une détection précoce et automatisée d'événements inhabituels (Hulth *et al.* 2010). Le package surveillance dans le logiciel R propose de nombreuses méthodes de détection (Höhle 2007). Le choix d'une méthode statistique est souvent difficile en l'absence d'étude comparative. Une solution consiste à utiliser en parallèle plusieurs méthodes. En cas d'événement inhabituel, l'analyse de l'épidémiologiste sera ainsi facilitée si plusieurs méthodes produisent en même temps des alarmes.

Détection spatiale

Les méthodes de détection spatiales prennent en compte les coordonnées géographiques des cas. Cela signifie qu'il faut disposer du lieu de survenue ou d'exposition de chaque cas. De nombreuses méthodes ont été développées pour tester une tendance à l'agrégation de cas d'un événement de santé (Goria *et al.* 2012). L'une des plus utilisées est le balayage spatial. Son objectif est la surveillance d'une zone géographique afin d'identifier des zones pour lesquelles une incidence plus élevée d'une maladie est observée, dans un intervalle de temps défini, sans hypothèses supplémentaires *a priori*.

Les méthodes de balayage spatial appliquent des fenêtres (le souvent circulaires) sur la zone d'étude considérée et dénombrent les cas et les individus à risque à l'intérieur et à l'extérieur de chaque fenêtre. Différentes méthodes existent, se distinguant entre autres par la construction de la fenêtre qu'elles utilisent. Parmi les méthodes de détection des agrégats, la *statistique de scan spatiale* est devenue l'une des plus communément utilisées (Kulldorff 1997).

Détection spatio-temporelle

La détection spatio-temporelle est une généralisation de la détection spatiale. L'analyse spatio-temporelle ajoute aux deux dimensions spatiales, une troisième dimension représentant le temps. On ne balaie plus une zone géographique par un cercle mais par un cylindre, la hauteur de celui-ci représentant le temps. L'intérêt de la détection spatio-temporelle par rapport à la détection spatiale est de prendre en compte le temps de survenue des cas. L'analyse spatio-temporelle est donc généralement plus intéressante que la détection purement spatiale.

Le logiciel libre Satscan propose des méthodes temporelles, spatiales et spatio-temporelles (<http://www.satscan.org/>). Il est utilisé par de nombreuses équipes à travers le monde dans des domaines très variés de la surveillance.

Discussion

L'analyse d'un nombre croissant de données de surveillance, à un rythme journalier ou hebdomadaire, nécessite l'utilisation de méthodes automatisées de détection précoce. Plusieurs pays européens disposent de ces outils pouvant être utilisés en routine

pour la surveillance spécifique ou non-spécifique, chez l'Homme ou chez l'animal. Les défis des prochaines années pour les pays disposant de systèmes de surveillance, seront de deux ordres. D'une part il s'agira d'insérer au cœur de l'analyse des données de surveillance, des modalités de détection temporelles et spatio-temporelles à différents niveaux géographiques (national, régional, départemental) pour un ensemble de maladies et affections. D'autre part, il conviendra d'améliorer les performances des méthodes (en termes de sensibilité et de spécificité), en prenant en compte les caractéristiques propres aux systèmes de surveillance, notamment des délais de déclaration lorsqu'ils existent et sont non négligeables. Les progrès informatiques et statistiques devront ainsi permettre d'améliorer l'identification des événements de santé, aussi bien en termes de réactivité que de sensibilité et spécificité.

Références bibliographiques

- Enki, D.G., Noufaily, A., Garthwaite, P.H., Andrews, N.J., Charlett, A., Lane, C., *et al.*, 2013. Automated biosurveillance data from England and Wales, 1991-2011. *Emerg Infect Dis.* 19(1),35-42.
- Farrington, P., Andrews, N. Outbreak detection: Application to infectious disease surveillance. In: Brookmeyer R, Stroup D, editors. *Monitoring the health of populations.* Oxford University Press; 2004. 203-31.
- Goria, S., Le Strat, Y., 2012. Analyse statistique spatiale des données de surveillance. In: John Libey, editor. *Epidémiologie de terrain - Méthodes et applications.* 437-49.
- Höhle, M., 2007. Surveillance: An R package for the surveillance of infectious diseases. *Comput Stat.* 22, 571-82.
- Hulth A, Andrews N, Ethelberg S, Dreesman J, Faensen D, van PW, *et al.*, 2010. Practical usage of computer-supported outbreak detection in five European countries. *Euro Surveill.* 15(36).
- Kulldorff, M., 1997. A spatial scan statistic. *Commun Stat Theory Methods.* 26(6), 1481-96.
- Unkel, S., Farrington, C., Garthwaite, P., 2012. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Ser A Stat Soc.* 175(1), 49-82.