

INFORMATION

Une introduction aux Mathématiques et Big Data en cancérologie

MOTS-CLÉS : MÉGADONNÉES. BIG DATA. INFORMATIQUE. RECHERCHE

Bernard NORDLINGER*

L'analyse des mégas donnés (Big Data) ouvre d'importantes perspectives pour la recherche médicale et notamment la recherche sur le cancer. De nombreux acteurs publics mais aussi privés s'y sont investis. La France dispose d'atouts majeurs dans ce domaine avec une recherche en cancérologie au meilleur niveau international et des mathématiciens de très grande valeur.

L'Académie nationale de médecine et l'Académie des Sciences se sont associées pour créer un groupe de travail bi-académique intitulé « Mathématiques, Big Data et cancérologie ». Ce groupe de travail est composé de mathématiciens, informaticiens, statisticiens, cancérologues, chirurgiens, onco-généticiens, etc. Son but est de stimuler la réflexion, favoriser la coopération entre les différents acteurs et faire des propositions pour aider au développement de cette recherche en France par des partenariats et des initiatives publiques et privées.

Par Big Data on entend l'ensemble des nouvelles méthodes mathématiques et statistiques permettant la collecte, le stockage et l'analyse des données et leur utilisation pour améliorer les connaissances, la prévention et le traitement des cancers.

L'intérêt du sujet est lié à l'explosion de la quantité de données collectées notamment via internet et les objets connectés, aux grandes capacités de recueil et de stockages de ces données, aux hauts débits de transmission de l'information, enfin au perfectionnement des méthodes d'analyse grâce à des algorithmes de plus en plus perfectionnés.

Ces données sont les traces que nous laissons, volontairement ou non, sur internet, nos téléphones, mais aussi les banques et les assureurs, les objets connectés. Dans certains cas les propriétaires des données ont donné leur consentement à l'utilisation de leurs données personnelles, mais ce n'est pas toujours le cas. L'usage de données non consenties peut poser problème.

Les données non structurées (vidéos, enregistrements vocaux, textes non numérisés...) sont beaucoup plus abondantes que les données structurées et leur exploitation présente pour l'avenir une haute valeur ajoutée. Des algorithmes perfectionnés permettent même de retrouver des données manquantes.

Ainsi l'émergence du numérique pourrait modifier nos comportements: pour certains on passerait de l'ère de la possession des données à celle du partage...

Une autre question parmi celles qui se posent est de savoir si la quantité des données analysées que permettent les Big Data peut compenser une qualité imparfaite, autrement dit si l'exhaustivité permet d'être moins exigeant sur l'exactitude des informations.

* Membre de l'Académie nationale de médecine

En matière d'accès aux données publiques, la loi SNDS (décrets d'application décembre 2016) regroupe les données du PMSI, de la CNAM, du registre des causes de décès de l'INSERM et de la caisse de solidarité, en un seul registre, unique au monde. Elle ouvre en principe l'accès aux méga données publiques à des fins de recherche. Il restera à savoir si des conditions d'accès trop strictes, destinées à protéger la vie privée des individus, ne deviendront pas en pratique des obstacles empêchant leur utilisation pour la recherche.

Des registres privés (*The cancer genome atlas*, *Cancerlinq* de l'ASCO...) s'ouvrent également.

Pour certains l'ère des Big Data amènerait un changement de paradigme de la recherche médicale : on passerait ainsi d'une recherche fondée sur une hypothèse suivie d'une démonstration (études cliniques phases I, II, III) à une analyse directe des données (*data mining*), pour découvrir des corrélations, des tendances, mêmes à partir de signaux faibles méconnus auparavant. En laissant « parler les données » on découvrirait des relations dont on n'aurait pas soupçonné l'existence. Ensuite il resterait à expliquer. Il est plus probable que la plupart des recherches de corrélations seront plus efficaces en étant orientées par une hypothèse.

Les mathématiques et l'analyse des Big Data trouvent des applications dans de nombreux domaines de la cancérologie, notamment pour :

- Aider l'industrie pharmaceutique à la recherche de nouvelles molécules ou de plus anciennes dont l'efficacité n'avait pas été prouvée.
- Associer phénotype et génotype grâce au croisement des bases de données cliniques et génétiques stockées dans les plateformes. Des sociétés privées proposent déjà aux hôpitaux des services d'analyse et d'interprétation de profils génomiques de leurs patients atteints de cancer.
- Identifier des marqueurs épigénétiques qui contrôlent la prolifération tumorale dans les tumeurs.
- Partager des résultats expérimentaux entre différents laboratoires.
- Modéliser la progression des tumeurs et des métastases en fonction de la réponse au traitement grâce à des algorithmes perfectionnés qui prennent en compte les données d'imagerie et les caractéristiques cellulaires.
- Améliorer l'efficacité des méthodes actuelles d'épidémiologie pour mieux identifier les facteurs de risque qui influencent la survenue des cancers. Les réseaux sociaux peuvent aussi se révéler des sources d'information utiles pour détecter précocement des épidémies.
- Développer des algorithmes d'aide au diagnostic et au traitement avec le but avoué de permettre aux médecins d'économiser du temps, mais aussi, pour certains, pour que les patients puissent se passer de médecin !

Mathématiques et Big Data ouvrent des perspectives immenses dans le domaine de la médecine et de la cancérologie en particulier, mais, comme tout progrès, ils amènent de nouvelles questions, notamment celles qui concernent la protection des données et le respect de la vie privée. Les considérations éthiques doivent être traitées. Les données doivent être contrôlées, anonymisées et pseudonymisées.

Nous vivons une période exaltante de développement exponentiel des moyens de recherche en cancérologie. Il faut les connaître pour mieux les exploiter, mais ne pas oublier que sur internet « le service est gratuit, nous sommes le produit ».