

mesure que le nombre de « J'aime » augmente sur un type de récit spécifique, la probabilité d'avoir un réseau social virtuel composé uniquement d'« amis » ayant le même profil augmente aussi, de façon proportionnelle. Ainsi, plus un internaute polarisé est actif, plus ses amis ont le même profil.

Cette division des réseaux sociaux en groupes homogènes selon le type de contenu consommé aide ainsi à comprendre la viralité des phénomènes sur la Toile. Chaque groupe a tendance à exclure tout ce qui n'est pas en cohérence avec sa vision du monde. Il s'agit ainsi d'une structure collective qui amplifie la sélection des contenus par le biais de confirmation à l'œuvre à l'échelle de chaque individu.

Nous avons ensuite poursuivi nos travaux en examinant les effets des campagnes de désintoxication visant à corriger la diffusion des fausses informations sur les médias sociaux. Nous avons ainsi comparé, parmi les utilisateurs généralement exposés à des sources conspirationnistes, ceux qui ont été exposés à des posts de démystification avec ceux qui ne l'ont pas été. Nous avons notamment mesuré la persistance, c'est-à-dire la probabilité de continuer à accorder des « J'aime » à un type de contenu spécifique dans le temps, pour les utilisateurs exposés ou pas à des campagnes de décodage des fausses informations.

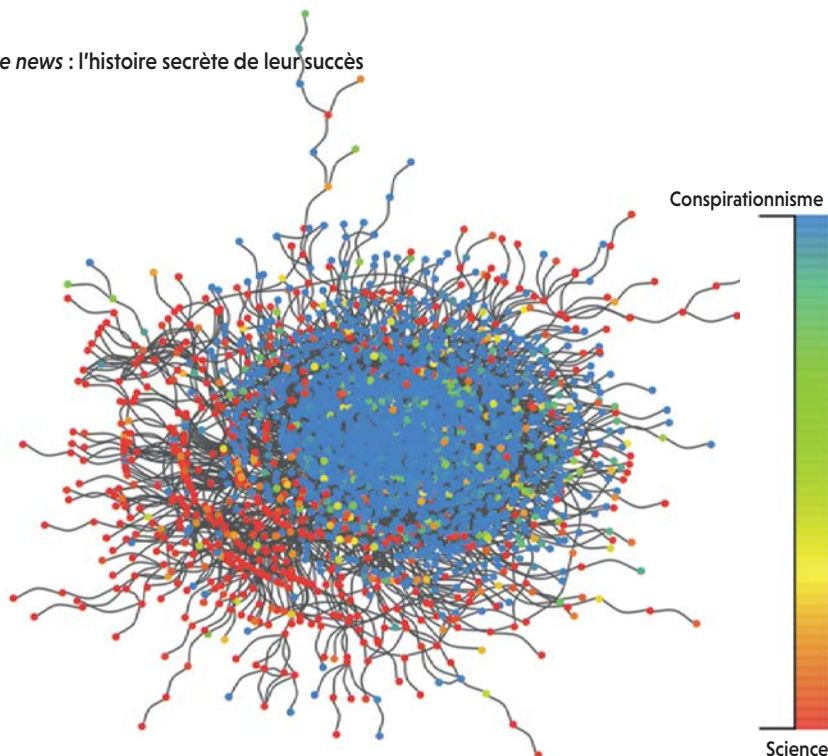
Pour ces utilisateurs exposés à la démystification, la probabilité de continuer à interagir avec des informations conspirationnistes est supérieure d'environ 30 % que pour les autres ! Essayer de convaincre un fervent défenseur de la théorie des chemtrails de son erreur renforce ses croyances et augmente ses interactions avec les sources d'informations de type complotiste.

Pour écarter toute particularité liée au contexte italien, nous avons mené une analyse similaire avec le Facebook américain. On retrouve les mêmes tendances...

DES DISCUSSIONS QUI DÉGÈNÈRENT

Un autre résultat intéressant a été obtenu au moyen d'algorithmes qui, bien entraînés, sont en mesure de fournir avec une bonne approximation le sentiment exprimé par les utilisateurs dans les commentaires des posts. Plus la discussion entre les internautes est longue, plus on se dirige vers un sentiment négatif. Et cela vaut tant pour les informations conspirationnistes que pour les informations scientifiques, même si les internautes conspirationnistes ont tendance à être plus négatifs. Dans tous les cas, une discussion prolongée dégénère.

Les dynamiques sociales qui émergent de nos études révèlent les problématiques relatives à la formation et à la propagation massive, sur les réseaux sociaux d'Internet, de récits potentiellement erronés.



Les internautes qui suivent habituellement les sources d'informations scientifiques ont relativement peu de liens d'« amitié » (au sens de Facebook) avec ceux qui suivent les sources conspirationnistes, comme le montre ce réseau de liens d'amitié reconstitué par l'auteur et ses collègues (en rouge, les utilisateurs plutôt exposés à des sources scientifiques, en bleu, les utilisateurs exposés à des sources conspirationnistes). On visualise ainsi l'effet « chambre d'écho » : les communautés d'internautes ont peu de contacts entre elles.

BIBLIOGRAPHIE

M. DEL VICARIO ET AL., The spreading of misinformation online, *PNAS*, vol. 113(3), pp. 554-559, 2016.

D. MOCANU ET AL., Collective attention in the age of (mis)information, *Computers in Human Behavior*, vol. 51, pp. 1198-1204, 2015.

A. BESSI ET AL., Science vs conspiracy: Collective narratives in the age of misinformation, *PLoS One*, vol. 10(2), e0118093, 2015.

A. BESSI ET AL., Trend of narratives in the age of misinformation, *PLoS One*, vol. 10(8), e0134641, 2015.

A. BESSI ET AL., Social determinants of content selection in the age of (mis)information, *Social Informatics, Lecture Notes in Computer Science*, vol. 8851, pp. 259-268, 2014.

La sélection des contenus par l'internaute s'effectue sous l'influence du biais de confirmation, ce qui conduit à la formation de groupes relativement homogènes de personnes qui s'intéressent aux mêmes thèmes et aux mêmes récits. Parallèlement au renforcement de cette focalisation, les membres de chaque groupe tendent à ignorer tout ce qui ne conforte pas leurs préjugés. Et quand discussion il y a avec des personnes ayant une opinion différente, elle dégénère souvent, ce qui encourage la polarisation.

Dans ce contexte, il devient difficile d'informer correctement, et arrêter la propagation d'une nouvelle infondée ou fausse devient impossible (voir la figure page 86). Le problème de la désinformation sur les réseaux sociaux est devenu si prégnant que les opérateurs de ces réseaux sont maintenant souvent accusés de ne pas prendre leurs responsabilités et sont incités à réagir de façon à limiter la propagation des fausses rumeurs et théories. Mais comment ? La question est complexe et les réponses ne font pas encore consensus.

Une partie des solutions pourrait être d'adapter les algorithmes qui gèrent les réseaux sociaux. Par exemple, Facebook a introduit pour les utilisateurs la possibilité de signaler les fausses informations, tandis que Google étudie le moyen de tenir compte de la fiabilité des pages dans le classement des résultats affichés à la suite d'une requête par mots-clés. Mais nos études ne nous invitent guère à l'optimisme.

Nous entendrons probablement pour quelque temps encore parler du complot mondial orchestré par les reptiliens, des extraterrestres d'apparence humaine qui seraient dissimulés parmi nous (et auxquels croient des millions d'Américains). Nous ne sommes pas dans l'ère de l'information mais bien dans celle de la crédulité ! ■

L'ESSENTIEL

● Préoccupés par notre santé, nous sommes prêts à subir de nombreux tests médicaux... qui n'apportent aucune certitude.

● Pour bien interpréter les résultats, médecins, malades et non-malades devraient se familiariser avec les statistiques.

● En apprenant à distinguer risque absolu et risque relatif, à utiliser la fréquence naturelle d'une maladie, les taux de mortalité plutôt que la survie à cinq ans, nous ne nous inquiéterons plus inutilement.

LES AUTEURS



GERD GIGERENZER travaille à l'institut Max-Planck de Berlin. WOLFGANG GAISSMAIER est professeur à l'université de Constance. ELKE KURZ-MILCKE travaille à l'Institut de mathématiques et d'informatique, à Ludwigsbourg. LISA SCHWARTZ et STEVEN WOLOSHIN sont professeurs à l'école de médecine de Dartmouth, à Hanovre, aux États-Unis.

SANTÉ : halte à la manipulation

Interpréter les résultats d'examens médicaux et d'études de risques sanitaires est un art difficile auquel peu de personnes sont initiées. Quelques conseils pour mieux comprendre leur signification ne sont donc pas inutiles.

E

n 1938, dans son essai *World Brain*, l'écrivain britannique H. G. Wells, auteur de *L'Homme invisible* et de *La Guerre des mondes*, prédisait que penser en termes statistiques serait aussi indispensable aux citoyens éduqués d'une démocratie moderne que lire et écrire. Qu'en est-il presque un siècle plus tard, en ce début du XXI^e siècle? Presque tous ceux qui vivent

dans les sociétés industrielles savent lire et écrire, mais peu savent interpréter correctement les statistiques et comprendre les notions de risque et d'incertitude. C'est aussi le lot de nombreux médecins, journalistes et hommes politiques qui, en conséquence, répandent de fausses idées dans le public.

L'inculture statistique n'est pas due à des déficits intellectuels particuliers – le «gène des statistiques» n'existe pas! – mais à divers facteurs sociaux et psychologiques: dans le domaine de la médecine, la nature paternaliste de la relation médecin-malade, l'illusion que la médecine offre des certitudes, que les interventions médicales sont toujours bénéfiques. L'anxiété et les espoirs des citoyens peuvent être facilement manipulés



Les statistiques médicales sont sujettes à diverses interprétations. Bien les comprendre éviterait des frayeurs inutiles.

pour des raisons politiques et commerciales. Avec des conséquences médicales et psychiques parfois redoutables.

Bonne nouvelle, on peut éviter certaines manipulations statistiques en médecine, donner du sens à des données chiffrées parfois peu claires, et utiliser cette information pour prendre les bonnes décisions. Et surtout, on peut tous sensibiliser les enfants aux statistiques de façon à les aider à résoudre des problèmes concrets.

Cela fait longtemps que la médecine se méfie des statistiques. Pendant des siècles, les thérapies se fondaient sur une confiance mutuelle plutôt que sur des données chiffrées, auxquelles on reprochait d'être impersonnelles ou peu pertinentes. Aujourd'hui encore, certains

médecins se fient davantage à leur intuition et à leur propre jugement qu'aux statistiques. De leur côté, nombre de patients préfèrent faire confiance à leur médecin plutôt que d'analyser eux-mêmes les résultats qui les concernent.

Les gens n'aiment pas les statistiques parce qu'ils ont besoin de certitude face à la maladie, tandis que les statistiques obligent à prendre des décisions sans certitude. Ainsi, une étude menée auprès de 1000 Allemands majeurs, en 2006, suggère que la plupart des gens considèrent que les tests de dépistage du VIH et les tests génétiques sont fiables à 100%, ce qui est faux.

De même, alors que la mammographie a réduit le risque de décès par cancer du sein des ➤

► femmes cinquantenaires d'environ 5 à 4 pour 1 000 en treize ans, 60% d'un échantillon aléatoire de femmes américaines pensaient que le bénéfice était 80 fois plus élevé.

Les citoyens des sociétés où la technologie est omniprésente sont confrontés à de nombreux dilemmes médicaux. Une femme enceinte âgée de 35 ans doit-elle subir une amniocentèse pour dépister une éventuelle anomalie chromosomique du fœtus, alors que cette procédure présente un risque (de l'ordre de 1%) d'entraîner une fausse couche? Doit-on vacciner les filles contre les papillomavirus humains afin de les protéger contre le cancer du col de l'utérus, alors que quelques complications ont été signalées, notamment un risque potentiel de paralysie?

Pour prendre des décisions éclairées, nous devons comprendre les statistiques médicales. En particulier, nous devons distinguer un risque absolu d'un risque relatif, et interpréter correctement la fréquence naturelle d'une maladie pour en déduire la probabilité d'en être atteints en cas de test positif. Nous devons aussi davantage nous fier aux taux de mortalité plutôt qu'aux statistiques, trompeuses, de survie à cinq ans.

RISQUES ABSOLU ET RELATIF

En octobre 1995, l'Agence de sécurité sanitaire du Royaume-Uni émit un avis selon lequel les pilules contraceptives de troisième génération doubleraient le risque de phlébite potentiellement mortelle dans les jambes ou les poumons (un caillot sanguin obstrue une veine); ce risque augmentait donc de 100%. Cette information fut transmise par courrier à 190 000 médecins généralistes, pharmaciens et directeurs de services médicaux, et sous forme de messages d'alerte dans les médias. La nouvelle émut tout le pays, et beaucoup de femmes cessèrent de prendre la pilule; l'année suivante, on compta 13 000 avortements supplémentaires en Angleterre et au pays de Galles; quelque 800 jeunes filles de moins de 16 ans eurent un enfant. Pourtant, les avortements et les grossesses augmentent le risque de thrombose dans des proportions bien supérieures à celle possiblement liée à la pilule de troisième génération.

Une telle panique aurait pu être évitée avec une meilleure information. En réalité, les données montraient qu'environ 1 femme sur 7 000 prenant une pilule de deuxième génération avait une thrombose; ce chiffre passait à 2 pour 7 000 avec la pilule de troisième génération. Ainsi, l'augmentation du risque absolu (voir les Repères, page 6) n'était que de 1 pour 7 000, alors que celle du risque relatif était, effectivement, de 100%.

Annoncer des risques relatifs peut provoquer des espoirs infondés, aussi bien que des inquiétudes inutiles. Nombre de patients et de médecins évaluent plus favorablement un traitement ou un test si les bénéfices correspondent à une diminution du risque relatif. En 2007, les travaux

de Judith Covey, de l'université de Durham, en Angleterre, le montraient: lorsque le bénéfice d'un médicament était présenté sous forme d'une réduction du risque relatif, 91% des généralistes danois le recommandaient à leurs patients. Mais lorsque l'information était présentée sous forme de réduction du risque absolu, seuls 63% recommandaient ce même médicament.

Les brochures d'information, les médecins, les revues médicales et les médias continuent à informer le public en termes de changements relatifs, en partie parce que les chiffres élevés attirent davantage l'attention. La confusion est encore plus grande quand on conjugue bénéfices et risques. Ainsi une publicité a affirmé que le traitement hormonal substitutif recommandé pour compenser le déficit en œstrogènes chez les femmes ménopausées «protège les femmes contre le cancer colorectal (jusqu'à plus de 50%)», tandis que le risque de cancer du sein «pourrait augmenter de 0,6%». En fait, le bénéfice relatif de 50% correspond à un nombre absolu inférieur à 6 pour 1 000. En d'autres termes, moins de 6 femmes sur 1 000 sont protégées du cancer colorectal par le traitement. Cela signifie que cette thérapie engendre au total plus de cancers qu'elle n'en prévient. Néanmoins, selon une étude de 2003 dans laquelle on distribuait cette brochure à 80 femmes âgées de 41 à 69 ans, 60 en avaient conclu le contraire.

Le risque absolu est plus informatif parce qu'il intègre l'information sur les proportions réelles à partir desquelles sont effectués les

LA CERTITUDE N'EXISTE PAS

Les questions à se poser quand on parle d'un risque:

De quel risque s'agit-il?

À quoi se réfère le risque: est-ce le risque de mourir d'une maladie, d'attraper une maladie ou d'en présenter les symptômes? Parle-t-on de risque absolu (la probabilité de développer une maladie durant une période de temps donnée) ou de risque relatif (le rapport entre le risque dans le groupe concerné et le risque dans un groupe témoin).

Quelle est l'échelle de temps?

On se représente plus concrètement les risques à l'horizon d'une dizaine d'années plutôt que de la vie entière.

Quelle est l'importance du risque?

Les nombres devraient être exprimés en termes absolus, ou bien sous une forme comparative, liant le risque à d'autres.

Ce risque me concerne-t-il?

Le calcul du risque repose-t-il sur l'étude d'individus au profil similaire (âge, sexe, santé...)?

L'annonce de risques relatifs peut susciter des espoirs infondés ou des inquiétudes inutiles

calculs. Du risque absolu on peut déduire le risque relatif, mais l'inverse n'est pas vrai. Une réduction de 50% du risque relatif peut décrire une diminution importante de 200 à 100 pour 10 000, ou bien minime, de 2 à 1 pour 100 000. En médecine, les résultats apportés



Une mammographie positive peut provoquer une angoisse considérable. Elle serait pourtant bien moins alarmante si l'on annonçait qu'avec ce résultat, la probabilité d'avoir un cancer du sein est bien inférieure à 100 %.

par les essais cliniques sont particulièrement fiables, mais s'ils sont exprimés de façon inadéquate, le public n'a aucune chance de les interpréter correctement.

DÉMÊLER LE VRAI DU FAUX POSITIF

Prenons le cas d'une femme qui, face au résultat positif d'une mammographie, demande à son médecin si elle a vraiment un cancer du sein, ou quelle est la probabilité qu'elle en soit vraiment atteinte. En 2007, l'un de nous (G. Gigerenzer) a demandé à 160 spécialistes de répondre à une telle question, en tenant compte des données suivantes: la prévalence de cancer du sein dans la région de la patiente est de 1%; si une femme a un cancer du sein, la probabilité que le test soit positif (sensibilité) est de 90%; si une femme n'a pas de cancer du sein, la probabilité qu'un test soit quand même positif (faux positif) est de 9%.

Quelle est, parmi les propositions suivantes, la meilleure réponse à donner à la patiente? (a) la probabilité qu'elle ait un cancer du sein est d'environ 81%; (b) sur 10 femmes ayant une mammographie positive, environ 9 ont un cancer du sein; (c) sur 10 femmes ayant une mammographie positive, environ 1 a un cancer du sein; (d) la probabilité qu'elle ait un cancer du sein est d'environ 1%.

La meilleure réponse est la troisième: en moyenne, sur 10 femmes dont les résultats sont positifs au dépistage mammographique, une seule environ a effectivement un cancer du sein. Les 9 autres sont alarmées inutilement. Seuls 21% des spécialistes interrogés avaient choisi la bonne réponse...

Nombre de médecins ne connaissent pas la probabilité qu'une personne soit effectivement malade en cas de test de dépistage positif, c'est-à-dire la valeur prédictive positive de ce test. Ils sont aussi incapables de l'estimer à partir de probabilités dites conditionnelles telles que la sensibilité du test (la probabilité d'un test positif en présence de la maladie) et la spécificité du test (le taux de faux positifs). De telles lacunes risquent d'entretenir des frayeurs inutiles. Or plusieurs mois après avoir reçu un résultat faux positif de mammographie, 1 femme sur 2 signale une anxiété importante liée à ce résultat, et 1 sur 4 rapporte que cette anxiété a affecté son humeur et sa vie quotidienne.

Les médecins seraient plus à même de déduire les probabilités correctes avec des statistiques relatives aux tests présentées sous forme de fréquences naturelles. Par exemple, avec les données de la mammographie évoquées plus haut: 10 femmes sur 1000 ont un cancer du sein; sur ces 10 femmes, 9 ont un test positif; sur les 990 femmes non atteintes, environ 89 ont quand même un résultat positif. Ainsi, 98 patientes (89 + 9) ont un test positif, mais seulement 9 ont un cancer.

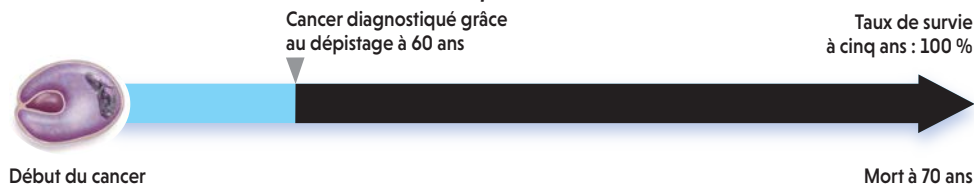
Les médecins devraient toujours informer leurs patients qu'aucun test n'est parfait, et que les résultats doivent être interprétés avec prudence, ou qu'ils doivent être réitérés pour voir si l'on obtient à nouveau le même résultat. Toutes les femmes qui passent une mammographie devraient savoir que les résultats indiquant une suspicion peuvent être de fausses alertes.

Une incertitude similaire existe pour tous les tests de dépistage, y compris celui du VIH. Bien que le test détecte effectivement 99,9% des infections réelles, et que 99,99% des résultats négatifs soient effectivement négatifs, le très faible risque chez les hommes hétérosexuels implique que le risque qu'ils ont d'être infectés n'excède pas 50% même quand un test est positif (voir l'encadré page 97). Cependant, lorsque le risque intrinsèque est supérieur, comme chez les homosexuels qui ont des rapports sexuels non protégés ou les toxicomanes qui partagent leurs seringues, la probabilité que le sujet soit infecté par le virus en cas de test positif est presque de 100%. Ainsi, le risque intrinsèque à un groupe donné détermine la signification d'un test positif.

UN INDICE TROMPEUR

Dans un spot télévisé de campagne électorale, en 2007, l'ancien maire de New York Rudy Giuliani annonçait: «J'ai eu un cancer de la prostate, il y a cinq ans. Mes chances de survie étaient de 82% aux États-Unis. En Angleterre, où le système de santé est socialisé, elles n'auraient été que de 42%. Dieu merci, j'ai guéri.» L'ex-édile sous-entendait qu'il avait eu de la chance de vivre à New York. Cette déclaration a fait les gros titres de la presse américaine. Pourtant, les >

AVEC UN DOSAGE DE L'ANTIGÈNE PROSTATIQUE, PSA



SANS DOSAGE DU PSA



Les statistiques de survie dépendent du moment du diagnostic, ce qui rend trompeuses les données statistiques. Dans le cas fictif présenté ici, un diagnostic de cancer de la prostate à l'âge de 60 ans (*en haut*) peut faire augmenter le taux de survie à cinq ans par rapport à un diagnostic posé sept ans plus tard (*en bas*). Pourtant, dans les deux cas, l'âge du décès est le même : 70 ans.

➤ chiffres qu'il avait donnés trahissaient une grossière erreur d'interprétation!

Selon les données de l'année 2000 qu'il a apparemment utilisées, sur les 49 Britanniques sur 100 000 chez qui un cancer de la prostate avait été diagnostiqué, 28 étaient décédés au bout de cinq ans, ce qui correspondait à un taux de survie à cinq ans d'environ 43 %. Le taux correspondant aux États-Unis était de 82 %, ce qui suggérerait que les Américains avaient deux fois plus de chances de survivre que les Britanniques à un cancer de la prostate. Mais cette déduction était fautive, parce que les statistiques de survie reflètent davantage des différences de diagnostic entre les deux pays que des traitements de meilleure qualité.

ÊTRE MALADE QUELQUE PART

Pour le comprendre, imaginons un groupe fictif de patients souffrant d'un cancer de la prostate diagnostiqué, d'après leurs symptômes, à l'âge de 67 ans au Royaume-Uni, et qui meurent tous à 70 ans. Chaque malade n'a survécu que trois ans, et en conséquence le taux de survie à cinq ans de ce groupe est égal à zéro. Prenons un groupe comparable aux États-Unis, où les médecins détectent la plupart des cancers de la prostate en dosant un antigène spécifique (le PSA), alors que ce dosage n'est pas réalisé en routine au Royaume-Uni. Ces patients sont diagnostiqués plus tôt, vers 60 ans, mais ils meurent quand même à 70 ans. Ils ont tous survécu dix ans, et en conséquence le taux de survie à cinq ans est de 100 %. Bien que les taux de survie soient radicalement différents, l'âge du décès est le même dans les deux groupes. Ainsi, en fixant le moment du diagnostic plus tôt, on augmente les taux de survie (biais d'avance au diagnostic) bien qu'aucune vie n'ait été prolongée ou sauvée (*voir la figure ci-dessus*).

Des taux de survie artificiellement élevés peuvent aussi résulter d'un « surdiagnostic », par exemple la détection d'anomalies qui sont des cancers, mais qui n'évolueront jamais assez pour menacer la vie du patient qui décèdera d'une

autre cause. Supposons 1000 hommes atteints d'un cancer de la prostate évolutif ne bénéficiant pas d'un test de dépistage. Au bout de cinq ans, 440 seront toujours en vie, ce qui représente un taux de survie à cinq ans de 44 %. Dans un autre groupe de 3000 hommes, le dosage du PSA révèle que 1000 ont un cancer évolutif et 2000 un cancer non évolutif (ils ne mourront pas de ce cancer dans les cinq ans). En ajoutant ces 2000 cas aux 440 qui ont survécu au cancer évolutif, on aboutit à un taux de survie à cinq ans gonflé, de 81 %, alors que la mortalité n'a pas été réduite.

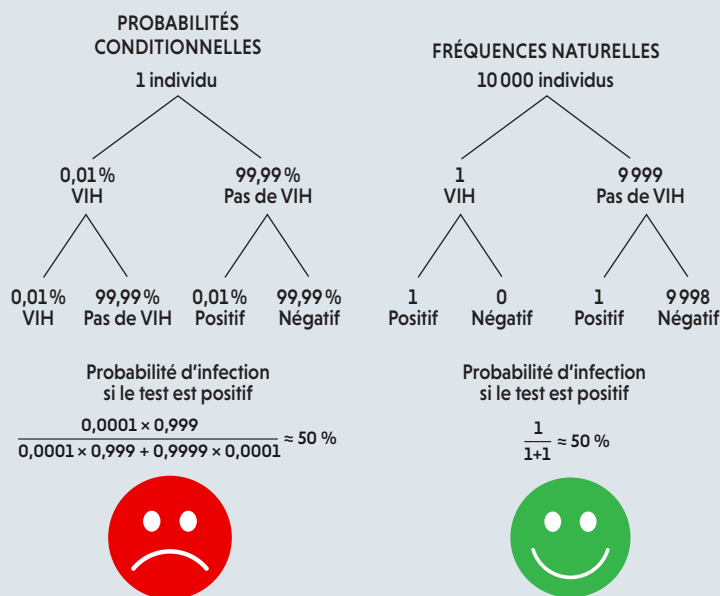
Aux États-Unis, le dépistage du cancer de la prostate par dosage du PSA a conduit à la fin des années 1980 à une explosion du nombre de nouveaux cancers diagnostiqués. Au Royaume-Uni, l'effet a été bien plus faible à cause d'une utilisation moins systématique du dosage de cet antigène. Cette disparité dans les diagnostics explique en grande partie pourquoi le taux de survie à cinq ans pour les cancers de la prostate est plus élevé outre-Atlantique.

Lorsque les pratiques diagnostiques diffèrent d'un pays à l'autre, la différence des taux de mortalité à cinq ans ne reflète pas de façon fiable la différence des taux de mortalité. Et pourtant, nombre d'agences officielles continuent de faire état de taux de survie à cinq ans. Un rapport du Bureau des statistiques du Royaume-Uni notait que le taux de survie à cinq ans du cancer du côlon était de 60 % aux États-Unis contre 35 % au Royaume-Uni. Les experts qualifiaient ce résultat de « scandaleux » et appelaient à un doublement des dépenses gouvernementales pour le traitement du cancer du côlon. En fait, les taux de mortalité dus à ce cancer sont à peu près les mêmes dans les deux pays.

Plus curieux encore, en 2001, une brochure publicitaire du Centre Anderson, au Texas, mélangeait les taux de survie avec les taux de mortalité : « Tandis que le taux de mortalité du cancer de la prostate a fluctué entre 1960 et 1990, le taux de survie au cancer de la prostate des patients du Centre Anderson a continué de progresser. »

PRÉVOIR UNE INFECTION

Si votre test de dépistage du VIH est positif et que vous êtes un homme à faible risque d'infection (hétérosexuel, non drogué...), quel est le risque que vous soyez effectivement porteur du virus ? Les probabilités conditionnelles (à gauche) proposent un calcul compliqué. En revanche, en se fondant sur les fréquences naturelles (à droite), on obtient facilement la réponse : sur 10 000 hommes, on s'attend à ce qu'un seul soit infecté et que son test soit donc positif ; parmi les 9 999 qui ne sont pas infectés, 1 devrait aussi avoir un résultat positif. En conséquence, on a deux tests positifs, mais un seul individu infecté. Donc la probabilité d'infection donnée par un test positif n'est pas de 100 %, mais de 50 %.



Les taux de mortalité sont des indicateurs plus fiables de la valeur des programmes de dépistage que les taux de survie à cinq ans. Si l'on se fie à ces taux, un homme doit-il faire systématiquement un dosage de l'antigène PSA ? Un fumeur doit-il passer systématiquement un scanner des poumons ? Il est vrai que ces deux examens détectent plus de cancers à un stade précoce ; mais aucun des deux ne permet de réduire la mortalité.

LES ÉPIDÉMIES DE DIAGNOSTICS

Les gens considèrent souvent les tests de dépistage comme des garants de leur santé. Toutefois, des examens supplémentaires peuvent conduire à des interventions médicales inutiles dont les effets sont parfois délétères. Et pour les nombreux patients inutilement diagnostiqués, le traitement a forcément des conséquences indésirables. Une épidémie de diagnostics peut être aussi dangereuse pour la santé que la maladie.

Les erreurs d'interprétation des statistiques seraient moins fréquentes si les chercheurs, les médecins et les médias utilisaient des données chiffrées directes au lieu de nombres qui prêtent à confusion : le risque absolu au lieu du risque relatif, les fréquences naturelles à la place des probabilités conditionnelles, et les taux de mortalité plutôt que les taux de survie à cinq ans. En outre, nous devons mieux éduquer les jeunes à la science du risque et de l'incertitude.

Comme le suggérait H. G. Wells, les statistiques devraient être enseignées en même temps que la lecture et l'écriture. De fait, aux États-Unis, l'Association nationale des enseignants de mathématiques insiste pour que l'enseignement des statistiques et des probabilités commence à l'école primaire. Si les

enfants apprenaient que le monde n'est pas fait de certitudes et ce d'une façon ludique, les statistiques seraient mieux comprises.

Au lieu d'apprendre aux étudiants comment appliquer des formules de probabilité pour résoudre des problèmes virtuels, les professeurs devraient leur montrer comment utiliser les statistiques pour résoudre des problèmes concrets. Par exemple, ils pourraient leur enseigner l'usage des statistiques et des probabilités quand il s'agit de décider comment se comporter face aux drogues, à la consommation d'alcool, à la conduite automobile, aux biotechnologies et à d'autres questions importantes pour la vie quotidienne.

Un livre scolaire américain du secondaire raconte l'histoire vraie d'une mère célibataire de 26 ans. À la suite d'un test positif de dépistage du VIH, elle perd son travail, déménage dans un foyer hébergeant d'autres personnes séropositives, a des relations sexuelles non protégées avec l'un d'entre eux, et attrape une bronchite. Son médecin lui prescrit alors un nouveau test de dépistage. Le résultat est négatif, tout comme celui de son échantillon sanguin précédent, qui a été réanalysé. Cette femme a vécu un cauchemar parce que ses médecins n'ont pas compris qu'un résultat positif à ce test n'était pas un verdict définitif, mais qu'il signifiait que cette femme avait une probabilité d'être infectée de 50 %, étant donné son appartenance à un groupe à faible risque.

L'éducation statistique peut changer des vies, aider les gens à prendre de meilleures décisions personnelles, à reconnaître les messages trompeurs et à développer une attitude plus sereine envers leur santé. Comme le recommandait le philosophe Emmanuel Kant : « Osez savoir ! » ■

BIBLIOGRAPHIE

WOLOSHIN ET AL., *Know your chances: Understanding health statistics*, University of California Press, 2008.

A. PLEASANT, *Communiquer sur les statistiques et le risque*, *SciDev Net*, 15 décembre 2008.

B. FAGERLIN ET AL., *Making numbers matter: present and future research in risk communication*, *Am. J. of Health and Behav.*, vol. 31, pp. S47-S51, 2007.

G. GIGERENZER, *Calculated risk: how to know when numbers deceive you*, Simon & Schuster, 2003.

B. FALISSARD, *Comprendre et utiliser les statistiques dans les sciences de la vie*, Masson, Abrégés, 3^e éd., 2005.

L'ESSENTIEL

- Cachés derrière un pseudonyme, nous laissons de nombreuses données personnelles sur Internet ou à divers organismes.
- C'est insuffisant pour garantir l'anonymat de ces données!
- Diverses techniques sont développées pour protéger les données sensibles sans empêcher leur exploitation.

LES AUTEURS



TRISTAN ALLARD est postdoctorant à l'Inria (Institut national de recherche en informatique et en automatique).



BENJAMIN NGUYEN est chercheur à l'Inria et maître de conférences à l'Université de Versailles-Saint-Quentin-en-Yvelines.



PHILIPPE PUCHERAL est chercheur à l'Inria et professeur à l'Université de Versailles-Saint-Quentin-en-Yvelines.

L'art de préserver L'ANONYMAT

Presque tous nos comportements laissent des empreintes numériques. Les informaticiens conçoivent et développent des techniques pour que l'analyse de ces gisements de données ne compromette pas la vie privée.

S

anté, déplacements, achats, appels téléphoniques, réseaux sociaux, recherches d'information: toutes ces facettes de nos vies laissent des empreintes numériques qui sont stockées et classées dans des bases de données gigantesques, telles celles des moteurs de recherche, qui gardent l'historique des requêtes associées à une adresse IP (la carte d'identité d'un appareil connecté) pendant des années. Ces masses d'informations constituent une manne sans

précédent pour les sociétés humaines. Dans le domaine de la santé, par exemple, on peut analyser les caractéristiques individuelles de chaque patient, afin de mieux le prendre en charge, ou de mener des études de santé publique. Si l'analyse de données sur les individus est pratiquée depuis des millénaires, notamment lors des recensements, la quantité et la diversité des informations aujourd'hui disponibles en multiplient l'intérêt potentiel... et les dangers.

En effet, ces myriades d'informations font planer une menace, elle aussi sans précédent, sur la vie privée des individus. La plupart du temps, les données ne sont pas publiques, mais elles circulent néanmoins entre différents acteurs: des ensembles de données médicales sont transmis à des organismes de recherche, les sites en ligne peuvent vendre une partie des données personnelles de leurs utilisateurs... Certaines données sont même accessibles à tous sur Internet: par exemple, le site Netflix a publié des données de ses utilisateurs à l'occasion d'un concours visant à optimiser les algorithmes qui déterminent les films à recommander.