

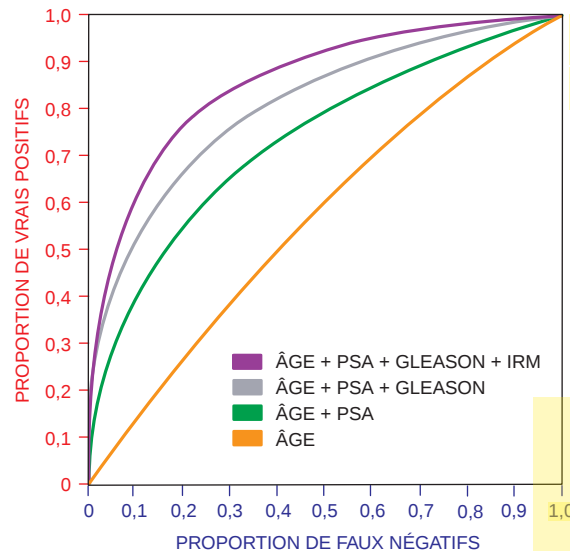
méthode ont une interprétation nettement plus pertinente, comparable en précision à celle des spécialistes.

Des méthodes statistiques prédictives sont également utilisées par les médecins pour le traitement du cancer de la prostate. Une de ces méthodes donne d'excellents résultats. Lorsqu'un homme est déclaré cliniquement atteint d'un cancer de la prostate (sur la base d'un examen médical, d'une biopsie simple et d'examens non invasifs), la question du traitement le plus adapté s'impose. Ni la prostatectomie ni la radiothérapie conformationnelle (irradiation concentrée sur la prostate) n'élimineront la tumeur si elle a envahi les tissus environnants, ou si elle a formé des métastases. Les médecins cherchent donc à déterminer le stade évolutif de la tumeur avant de choisir un traitement.

Malheureusement, des tumeurs qui semblent initialement confinées à la prostate se révèlent souvent être à des stades bien plus évolués. Pendant des années, les médecins avaient peu de moyens fiables pour prédire chez quels patients l'envahissement de la tumeur dépassait la glande. Aujourd'hui, ils utilisent un tableau admis par la communauté médicale, qui indique le stade d'une tumeur en fonction des caractéristiques de la tumeur humaine et de la présence de métastases.

Les chercheurs qui ont élaboré ces tableaux savaient que trois indices donnaient chacun une valeur prédictive indépendante : le «stade clinique de la tumeur» (déterminé par des examens non invasifs, de taille et d'étendue) ; la concentration sanguine en antigène prostatique spécifique (PSA, une protéine libérée par les cellules de la prostate) ; et le score de Gleason (un indicateur de l'agressivité de la tumeur, établi sur les analyses histologiques d'une biopsie). Les biostatisticiens ont élaboré une méthode statistique prédictive qui intègre la plupart des combinaisons de résultats pour ces trois variables et qui calcule la probabilité que le diagnostic initial de «cancer localisé» soit exact. Les probabilités sont finalement présentées sous la forme d'un tableau facile d'utilisation.

L'ensemble des exemples proposés a montré que la prise de décision repose sur une valeur de seuil critique. Le pro-



2. LES COURBES ROC comparent la pertinence de diverses mesures pour l'évaluation du stade évolutif d'un cancer de la prostate. Les variables étudiées sont : l'âge des patients, la concentration sanguine en PSA (antigène prostatique spécifique), l'agressivité de la tumeur (représentée par le score de Gleason) et l'aspect de la tumeur lors d'un examen d'imagerie par résonance magnétique (IRM). Le modèle prenant en compte les quatre variables s'est révélé être le meilleur (courbe du haut).

blème principal des analyses statistiques est la difficulté de choisir des seuils diagnostiques rationnels, comme pour les tests de dépistage du virus du SIDA (VIH).

Le dépistage du VIH commence avec un test simple, qui détecte la présence d'anticorps anti-VIH, molécules produites par l'organisme lorsque le système immunitaire commence à réagir à l'intrusion du virus. Quand le résultat est positif (à partir d'un certain taux d'anticorps), les laboratoires pratiquent un autre test plus complexe. On limite ainsi le nombre de faux positifs. Les tests sur les anticorps sont souvent problématiques, car ils ne reposent pas tous sur les mêmes valeurs seuils et n'ont pas tous la même pertinence. Des seuils variés seraient justifiés si chaque test était destiné à une population distincte, mais tel n'est pas le cas.

Les seuils sont également gênants d'un autre point de vue. Initialement mis au point pour distinguer les dons de sang sains des échantillons contaminés, on les a conservés pour l'identification des personnes infectées par le virus. Or, si l'on n'hésite pas à jeter 500 millilitres de sang sain déclaré – à tort – positif (faux positif), il est beaucoup plus grave d'inquiéter une personne saine en l'envoyant subir d'autres tests de dépistage du VIH parce que le premier test aura été faussement positif. Pis encore, les seuils initiaux

ont été utilisés sans distinction pour tous les donneurs de sang : ceux qui ont un faible risque d'infection, comme ceux qui ont un risque élevé. Lors de l'évaluation des individus à risque élevé, les seuils devraient être plus «laxistes» que pour les populations à risques faibles (afin d'augmenter la probabilité de découverte), même si ce choix augmente le nombre de faux positifs.

Ces dernières années, les tests de confirmation sont devenus plus précis, tandis que des traitements contre l'infection virale prolongent la vie et améliorent l'état de santé des malades. Les diagnostics de faux positifs sont plus rares aujourd'hui, et les personnes séropositives sont mieux dépistées qu'autrefois. Aujourd'hui, la question n'est plus celle du seuil de séropositivité, mais reste celle du dépistage des personnes contaminées et qui l'ignorent.

De la météorologie à l'aéronautique

Les médecins ne sont pas les seuls à utiliser les méthodes statistiques prédictives : les météorologues ont adopté cet outil pour leurs prévisions il y a plus de 25 ans. *Météo France*, par exemple, alimente régulièrement en données météorologiques des programmes statistiques qui estiment le risque de tornades, d'ouragans, de fortes pluies, etc. *Météo France* transmet ces prédictions objectives aux météorologues régionaux, qui les modifient en fonction de nouvelles informations ou de facteurs que les programmes informatiques n'auraient pas correctement pris en compte.

Enfin, d'autres professions ont également adopté les méthodes statistiques d'aide à la décision. Par exemple, les commissions universitaires américaines cherchent ainsi à mieux évaluer les candidats qui postulent pour des études en troisième cycle. Ils utilisent des méthodes prédictives pour prévoir les notes des étudiants à partir des notes obtenues au cours du cursus universitaire et des résultats aux tests d'aptitude pour l'entrée en troisième cycle. Les étudiants dont les notes sont supérieures à une valeur seuil préétablie sont généralement reçus, et ceux dont

les notes sont inférieures à une valeur seuil inférieure sont généralement refusés. Le comité évalue ensuite, plus subjectivement, les qualifications des postulants qui n'ont pas été admis par les méthodes statistiques prédictives. En France, les dossiers ne sont pas sélectionnés de cette façon statistique, mais par des évaluations... peut-être plus subjectives.

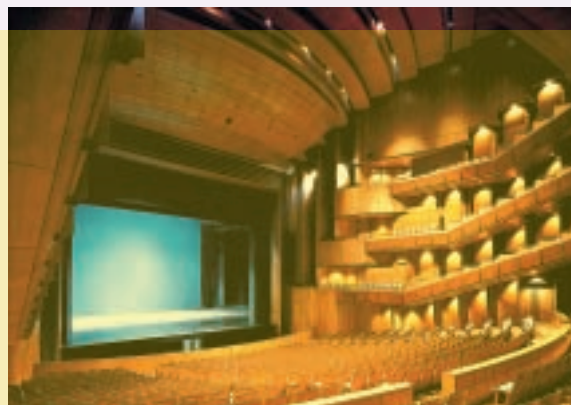
Les méthodes objectives d'établissement des seuils ont aussi été sous-utilisées dans l'industrie aéronautique. Lors de l'entretien des avions, on doit rechercher les défaillances graves, mais rares, telles que des fissures dans les ailes. Quand un diagnostic d'aile fissurée n'est pas fait, les conséquences sont catastrophiques : de nombreux passagers meurent si l'avion s'écrase. En revanche, une décision faussement positive met inutilement l'avion hors service, ce qui perturbe le service et coûte cher aux compagnies. À première vue, les bénéfices et les coûts sont en faveur d'un seuil «laxiste», les vies n'ayant pas de prix. Cependant, comme ces fissures sont rares, un seuil «laxiste» entraînerait un nombre incalculable de faux positifs. Aujourd'hui, aucune technique statistique ne résout le dilemme.

Les utilisateurs d'alarmes de cockpits (comme les compagnies aériennes ou l'armée) n'ont pas non plus réussi à choisir quels seuils de décision adopter. En vol, les alarmes se déclenchent en de multiples circonstances : lorsque les appareils de mesure enregistrent qu'un autre avion est trop proche, quand l'avion se rapproche trop du sol, quand un moteur est en panne ou quand un vent violent balaie la piste d'atterrissage. Toutefois, ces alarmes sonnent de façon intempestive, en raison du manque de précision des détecteurs ou du laxisme de leurs seuils. Aussi les pilotes hésitent-ils à tenir compte de ces alarmes trop fréquentes ; des interventions systématiques risqueraient d'ailleurs de perturber les vols. Toutefois, en raison du grand nombre de fausses alarmes, les pilotes risquent de finir par les ignorer ou par réagir trop lentement à une véritable urgence.

Les méthodes statistiques prédictives augmentent souvent la pertinence des diagnostics, et les formules pour établir des seuils de partage amélioreraient l'utilité de ces choix. Ces outils ont encore d'autres avantages : en standardisant les analyses sur lesquelles se fondent les diagnostics, les méthodes

Une question de goût

Les architectes et les amateurs de vin ont trouvé deux applications amusantes des méthodes statistiques prédictives. La «méthode architecturale», qui s'applique à des salles d'opéra, fut mise au point grâce à des chefs d'orchestre qui avaient noté la qualité sonore de 23 salles. Puis des ingénieurs acousticiens mesurèrent physiquement différentes propriétés acoustiques (telles que le délai entre les sons directs et les sons réfléchis, la diffusion des ondes sonores provoquée par les irrégularités du plafond et des murs) de chacune de ces 23 salles. Les analyses statistiques ont ensuite révélé quelles sont les caractéristiques des meilleures salles d'opéra. On a alors amélioré la conception de nouvelles salles.



Le nouvel opéra de Tokyo, dont les qualités acoustiques ont été validées par des méthodes statistiques de prévision du son.

En œnologie, une méthode statistique prévoit la qualité future des Bordeaux rouges (qualité mesurée par leur prix de vente aux enchères) à partir d'un test sur des échantillons de vins encore jeunes et imbuables. Auparavant, les experts prévoyaient la qualité future en sentant et en goûtant le nouveau produit. Il y a dix ans, les œnologues ont constaté que les années marquées par un hiver humide, des mois d'août et de septembre secs et une période de croissance chaude produisent d'excellents vins.

Ils ont alors formulé l'«équation bordelaise», une méthode statistique prédictive qui combine les conditions météorologiques et les années de vieillissement pour définir la probabilité qu'un vin soit excellent dans plusieurs années. Cette équation explique 83 pour cent des divergences de prix des vins rouges de Bordeaux à maturité, mis aux enchères. Pourtant cette équation, suscita une vive réaction des œnologues, qui la décrièrent violemment dans un article.

statistiques facilitent l'identification des paramètres importants. Elles fournissent aux décideurs des arguments faciles à communiquer et qui leur permettent de préciser de vagues impressions. De surcroît, elles facilitent la formation des débutants.

Ces outils sont parfois accueillis avec réticence, notamment quand les experts craignent qu'ils ne les remplacent. Les médecins également veulent garder la maîtrise de leurs diagnostics et de leurs recommandations, et donner un compte rendu argumenté de leur décision. Les résultats d'une méthode statistique prédictive sont parfois difficiles à prendre en compte, surtout quand la logique qui sous-tend l'analyse n'est pas évidente.

Nous comprenons ces préoccupations, mais les bénéfices apportés par les outils statistiques justifient certainement que les personnes amenées à prendre des décisions, qui ont la vie d'autrui entre leurs mains, y réfléchissent.

John SWETS est spécialiste de la «politique» de santé américaine et radiologue à Boston. Robyn DAWES est professeur de sciences sociales à l'Université de Carnegie Mellon. John MONAHAN dirige l'association de recherche Loi-santé mentale de la Fondation MacArthur.

A.W. PARTIN et al., *Combination of Prostate-Specific Antigen, Clinical Stage and Gleason Score to Predict Pathological Stage of Localized Prostate Cancer*, in *Journal of the American Medical Association*, vol. 277, n° 18, pp. 1445-1451, 14 mai 1997.

Kenneth A. FREEDBERG et Jeffrey H. SAMET, *Think HIV : Why Physicians Should Lower their Threshold for HIV Testing*, in *Archives of Internal Medicine*, vol. 159, n° 17, pp. 1994-2003, 27 septembre 1999.

John A. SWETS, Robyn M. DAWES et John MONAHAN, *Psychological Science Can Improve Diagnostic Decisions*, in *Psychological Science in the Public Interest* (supplément à *Psychological Science*), vol. 1, n° 1, pp. 1-26, mai 2000.