



Peter Menzel

6. PENDANT UN PONTAGE CORONARIEN avec un robot, le chirurgien (à gauche) est assis à quelque distance du malade. Les autres membres de l'équipe chirurgicale observent le déroulement de l'opération sur un écran (en haut, à gauche).

de la cage thoracique : cette manœuvre déforme l'organe et diminue la quantité de sang qu'il pompe.

Pourtant, ces dernières années, des chirurgiens ont découvert des mesures simples qui évitent ces accidents. Mon collègue P. Gründeman a montré qu'une inclinaison de 15 à 20 degrés vers le bas de la table d'opération empêche une chute excessive de la pression sanguine. À l'hôpital de Recife, Ricardo Lima utilise le péricarde (la membrane qui entoure le cœur) pour soulever partiellement l'organe hors de la cage thoracique et accéder à l'envers du cœur sans compromettre la pression sanguine. Il a été suivi par de nombreux confrères.

D'autres solutions sont étudiées. Ainsi, en chirurgie abdominale, des opérations, telle l'ablation de la vésicule biliaire, s'effectuent par des incisions de la taille d'un trou de serrure, grâce à la chirurgie endoscopique : les chirurgiens insèrent un tube rigide relié à une caméra vidéo miniature (l'endoscope) par une incision et les instruments chirurgicaux par deux autres incisions : l'image de l'endoscope guide les mouvements du chirurgien. Pourquoi n'opérerait-on pas le cœur ainsi ?

À Stanford, des médecins ont mis au point des procédés de chirurgie cardiaque par endoscopie dans un thorax fermé et sur cœur arrêté, tandis que le malade est relié à une machine cœur-poumon.

La connexion à une machine cœur-poumon, ainsi que l'arrêt du cœur sans ouverture de la cage thoracique requièrent divers tubes et des cathéters manipulés à partir de l'aîne. Cependant, l'opération n'est pas toujours bien supportée. De plus, la suture d'un pontage est très exigeante, mais la panoplie des instruments actuels de la chirurgie endoscopique est restreinte, et la marge de manœuvre dans une cage thoracique fermée est étroite : les essais d'opération cardiaque par endoscopie ont été abandonnés après trois tentatives. Apparemment, on doit effectuer des ouvertures plus grandes (de six à neuf centimètres) pour obtenir des sutures fiables des greffons aux artères coronaires : au milieu de l'année 2000, plus de 6 000 malades avaient subi une telle opération des coronaires.

Un robot chirurgien

Les chirurgiens opéreront-ils un jour un cœur battant dans la cage thoracique fermée ? Les insuffisances de l'endoscopie classique seront peut-être compensées par des robots qui manipuleraient à distance des instruments chirurgicaux. Avec un tel équipement, les médecins visualisent en trois dimensions l'intérieur de la cavité thoracique, et les actions de l'opérateur sur une console seraient transmises aux instruments chirurgicaux dans le

thorax, sans tremblement et avec plus de précision.

Alain Carpentier et Didier Loulmet, de l'hôpital Broussais, à Paris, puis Friedrich Mohr, Volkmar Falk et Anno Diegeler, de l'Université de Leipzig, ont été les pionniers d'une telle chirurgie coronarienne à thorax fermé, avec cependant l'assistance d'une machine cœur-poumon. En 1998, ils ont combiné le procédé mis au point à l'Université de Stanford et le système de chirurgie endoscopique robotisé, nommé da Vinci.

En septembre 1999, à l'Université de l'Ontario, Douglas Boyd a effectué la première opération assistée par ordinateur, sur cœur battant et dans une cage thoracique close. Toutefois, l'opération a duré presque une journée. Au milieu de l'année 2000, des chirurgiens de Munich, de Leipzig, de Dresde et de Londres ont réduit le temps d'opération à moins de cinq heures pour des pontages coronariens simples, sur cœur battant et cage thoracique fermée ; les opérations ont été satisfaisantes.

Les techniques robotiques s'imposent progressivement. Bientôt les chirurgiens devront s'entraîner à la pratique des opérations coronariennes par endoscopie, de la même manière que les pilotes s'exercent dans des simulateurs de vol. Ensuite, d'autres innovations faciliteront sans doute la chirurgie des maladies cardiaques coronariennes : par exemple, grâce à un système à pression en cours de mise au point, les chirurgiens fixeront rapidement un pontage sans point de suture.

Les pontages coronariens seront peut-être un jour « dépassés » ; en attendant, la prévention des maladies cardiaques coronariennes et le perfectionnement constant des techniques restent des priorités.

Cornelius BORST est professeur de cardiologie expérimentale au centre médical de l'Université d'Utrecht, aux Pays-Bas.

Cornelius BORST et Paul F. GRÜNDEMANN, *Minimally Invasive Coronary Artery Bypass Grafting : An Experimental Perspective*, in *Circulation*, vol. 99, n° 11, pp. 1400-1403, 23 mars 1999.

Minimally Invasive Cardiac Surgery, sous la direction de Robbin G. Cohen et al., Quality Medical Publishing, 1999.

Minimal Access Cardiothoracic Surgery, sous la direction d'Anthony P.C. Yim et al., W.B. Saunders Company, 2000.

Comment rationaliser les décisions

JOHN SWETS • ROBYN DAWES • JOHN MONAHAN

Des outils mathématiques d'aide à la prise de décisions améliorent les diagnostics en médecine ou dans l'industrie.

Une opacité sur une mammographie est parfois le signe d'une tumeur. Doit-on l'opérer? Un comité qui étudie les dossiers de mise en liberté conditionnelle est alerté par le comportement violent d'un détenu : doit-on, malgré tout, le libérer? Dans un aéroport, un technicien observe une image anormale lors d'une inspection aux ultrasons d'une structure d'avion. Est-elle le signe d'une fissure inquiétante?

Souvent, les décisions sont difficiles à prendre, car les données sont incomplètes ou ambiguës. En médecine, en sécurité publique, dans l'industrie, dans les administrations, dans l'armée, au gouvernement, les enjeux sont parfois considérables, car des vies humaines sont en jeu.

Toutefois, les méthodes d'aide à la décision sont presque les mêmes, quels que soient les domaines. Des améliorations méthodologiques dans un domaine profitent donc généralement aux autres domaines. Ces méthodes, dont deux au moins sont validées, sont malheureusement méconnues ou inutilisées. L'une d'entre elles améliore la «pertinence» d'une décision : elle augmente la probabilité que la décision prise soit la bonne. L'autre améliore l'«utilité» d'une prise de décision ; elle évite les alertes inutiles. Ces méthodes sont statistiques et concernent donc un grand nombre d'individus ; au niveau individuel, elles influencent le choix du traitement thérapeutique, mais n'établissent pas une prédiction de l'évolution de la maladie d'une personne.

La décision n'impose pas toujours la mathématisation des problèmes. Dans certains domaines, telles la médecine

ou la météorologie, où de nombreux facteurs interviennent, les outils statistiques ne sont qu'une aide à la décision, des examens complémentaires ou des données nouvelles facilitant les décisions. En revanche, dans d'autres domaines, les analyses statistiques se sont souvent révélées plus pertinentes que les jugements subjectifs, même formulés par des professionnels expérimentés.

Nous examinerons ici les décisions limitées à deux éventualités. Une tumeur est-elle présente, oui ou non? Le délinquant peut-il être libéré sans risque de récidive, oui ou non? L'aile de l'avion est-elle défectueuse, oui ou non? De nombreux problèmes font apparaître des choix entre plus de deux possibilités, mais les décisions oui/non prédominent. Toutefois, on est généralement confronté à quatre solutions (et non deux) : les «vrais positifs» (une opacité se révèle être une tumeur) et les «vrais négatifs» (l'absence d'opacité correspond à une absence de tumeur), où la décision prise est sans ambiguïté ; les «faux positifs» (il n'y a pas de tumeur, malgré une opacité) et les «faux négatifs» (il y a une tumeur sans opacité), pour lesquelles la décision prise risque de ne pas être la bonne.

Les outils du métier

Si les tests diagnostiques donnaient les réponses attendues, les statistiques d'aide à la décision seraient évidemment inutiles. Toutefois, les résultats bruts des tests doivent généralement être interprétés. Par exemple, les ophtalmologistes mesurent la pression intraoculaire pour dépister les glau-

comes, des compressions du nerf optique et un accroissement de la dureté du globe, qui conduisent parfois à la cécité. Une pression très faible indique un œil sain, tandis qu'une pression élevée signale un glaucome... mais les valeurs intermédiaires?

Les informations statistiques éclairent partiellement la situation. Pour simplifier, admettons que la pression soit la seule mesure diagnostique disponible pour le dépistage du glaucome. Supposons qu'une pression inférieure à 10 (avec des unités arbitraires) corresponde toujours à un œil sain, que les pressions supérieures à 40 témoignent toujours d'un glaucome ; les mesures comprises entre 10 et 40 s'observent aussi bien sur des yeux sains que sur des yeux malades.

Pour élaborer des outils diagnostiques fiables, on doit d'abord identifier une grande population d'individus dont on mesure la pression intraoculaire, puis déterminer quels individus ont un glaucome au cours d'une période donnée. Pour chaque valeur de la pression, on détermine alors la probabilité qu'une personne ait un glaucome. Ces probabilités, et d'autres considérations présentées plus loin, définissent un point de partage rationnel, ou seuil diagnostique : les scores supérieurs ou égaux à ce seuil conduisent à un diagnostic positif (le patient souffre d'un glaucome), les scores inférieurs à un diagnostic négatif (le patient ne présente pas de glaucome).

Évidemment, un seul test diagnostique procure moins d'informations qu'une combinaison de tests. Pour accroître la pertinence d'un diagnostic, le décideur a intérêt à effectuer plusieurs examens qui fournissent

chacun une information en donnant un poids plus important aux mesures les plus prédictives de la pathologie étudiée. Les algorithmes mathématiques qui définissent les meilleurs tests et qui calculent le risque de glaucome d'après l'analyse des résultats combinés sont des méthodes statistiques prédictives.

Les données totalement objectives, telle la mesure de la pression, ne sont pas les seules utiles pour l'amélioration de la pertinence des méthodes statistiques prédictives ; des impressions subjectives peuvent également être quantifiées. On les objective en réalisant, par exemple, une liste de critères perceptibles (comme l'irrégularité d'un grain de beauté éventuellement malin), classés sur une échelle de 1 à 5.

On estime la pertinence globale des méthodes statistiques prédictives en étudiant les courbes ROC (de l'an-

glais *Receiver Operating Characteristic*, qui signifie «courbes des caractéristiques d'efficacité»). Ces courbes ont été utilisées pour la première fois au cours de la Seconde Guerre mondiale quand on a étudié la fiabilité des radars : on voulait quantifier leur capacité à distinguer les interférences aléatoires (le bruit de fond) du signal indiquant réellement la présence d'avions ennemis.

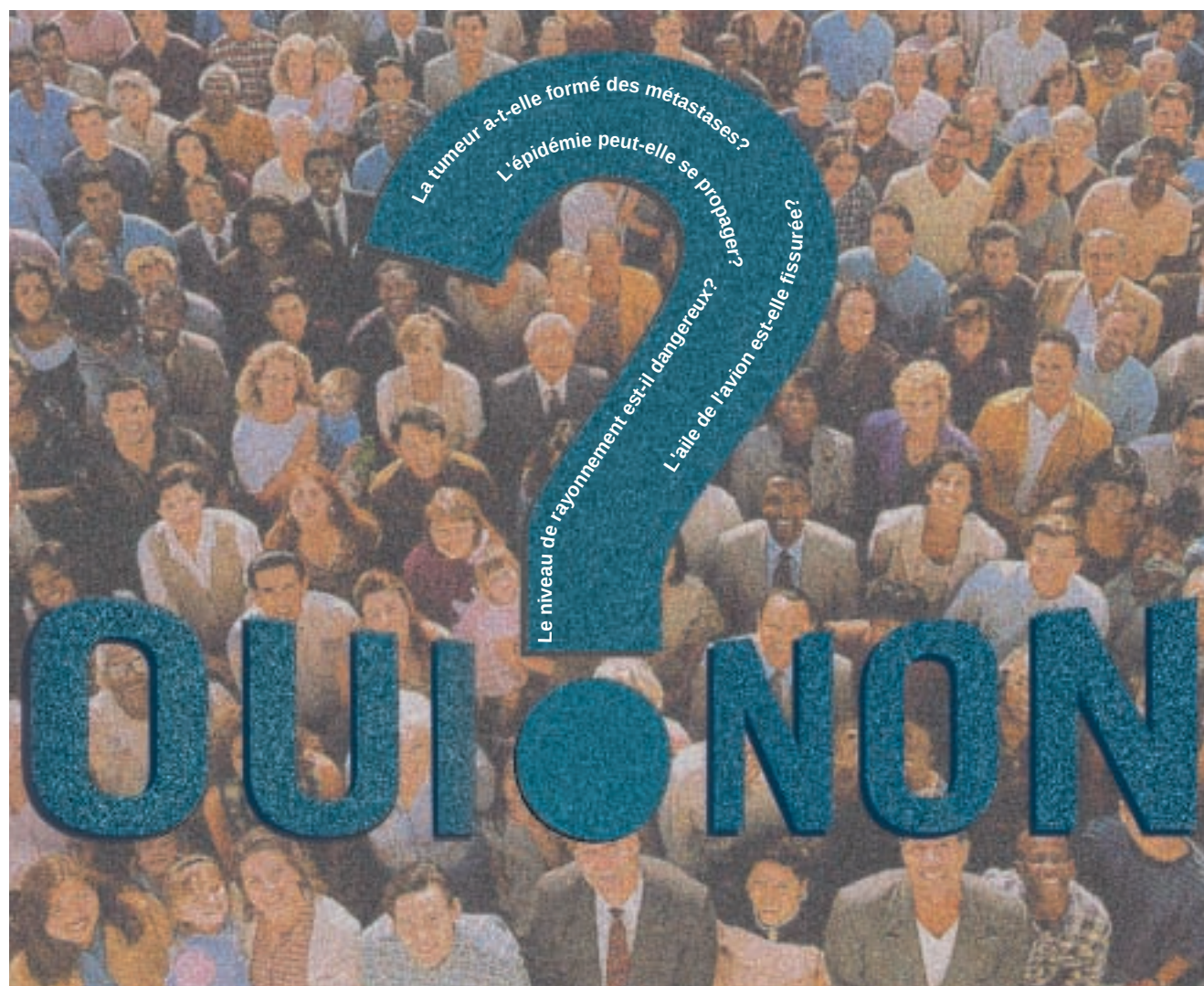
Les analyses des courbes ROC donnent une évaluation de la proportion d'individus atteints (et non atteints) en fonction d'une valeur seuil choisie : quelle proportion d'individus réellement atteints par la maladie étudiée seront dépistés (vrais positifs)? Et quel pourcentage d'individus sains seront par erreur considérés comme malades (faux positifs, ou «fausses alarmes»)?

Pour chaque seuil choisi, on détermine la proportion relative de vrais

positifs par rapport aux faux positifs. On obtient ainsi un graphe dont la courbure indique la pertinence : plus elle est forte, plus la prédiction est pertinente, car plus le rapport des vrais positifs sur les faux positifs est élevé.

Les vrais positifs et les faux positifs ne sont pas les seuls résultats possibles. Un diagnostic fondé sur des seuils donne aussi des vrais négatifs (des individus correctement estimés sains) et des faux négatifs (des individus malades non dépistés). Ces résultats complémentaires des précédents sont ignorés lors de la construction des courbes ROC. Par exemple, une proportion de vrais positifs égale à 80 pour cent correspond à une proportion de faux négatifs égale à 20 pour cent, ce qui représente une erreur considérable (un malade sur cinq n'est pas dépisté).

Comme les méthodes diagnostiques qui distinguent parfaitement les



1. LES QUESTIONS auxquelles un expert doit répondre par oui ou par non abondent, en médecine comme dans de nombreux autres domaines. Des méthodes statistiques confirmées, mais sous-utilisées, augmentent les chances de faire le bon choix.

individus malades des individus sains sont rares, chaque organisme doit décider de l'importance qu'il accorde à la détection de tous les vrais positifs ou de la majorité d'entre eux. Ce choix a des conséquences, car à un plus grand nombre de vrais positifs correspond également un plus grand nombre de faux positifs. Un seuil doit être adapté à chaque situation.

Revenons à l'exemple du glaucome. Les médecins n'utilisant que la pression intraoculaire trouveront la plupart des cas de glaucome s'ils choisissent un seuil diagnostique très «laxiste», égal à 10, par exemple (en supposant que l'échantillon de population du test ait montré que tous ceux qui étaient atteints d'un glaucome avaient une pression supérieure à ce seuil). Avec un tel choix, de nombreuses personnes saines seront déclarées malades ou faux positifs, et

inutilement inquiétées et traitées. Pour réduire ce type d'erreurs, les médecins peuvent choisir un seuil diagnostique assez «sévère», égal à 35 par exemple (très peu de personnes saines de l'échantillon avaient une pression aussi élevée), mais on dépisterait alors moins de la moitié de tous les individus malades ; les nombreux faux négatifs seraient alors privés de traitement.

On choisit le seuil en fonction des proportions de faux négatifs et de faux positifs qu'il engendre (ces proportions dépendent de la prévalence de la maladie dans la population étudiée). Les statisticiens ont identifié des principes et mis au point des outils mathématiques pour déterminer le seuil optimal. Par exemple, on préfère un seuil «laxiste» quand une maladie est fréquente dans une population ou quand il est très important de détec-

ter tous les cas avérés ; en revanche, le seuil doit être «sévère» quand la maladie est rare ou que le coût du traitement est élevé (en dehors de toute considération éthique).

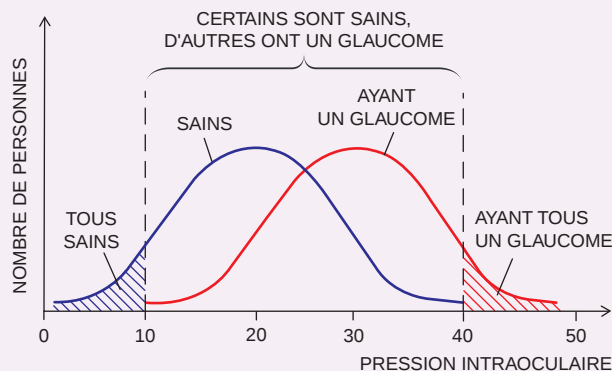
Bien que les méthodes statistiques prédictives et les courbes ROC soient peu utilisées par les médecins, les exemples de leur intérêt abondent. L'une des meilleures illustrations provient de la psychiatrie.

Régulièrement, les psychiatres et les psychologues doivent évaluer si des individus incarcérés ou perturbés risquent de devenir violents. Les personnes qui menacent la sécurité d'autrui doivent être identifiées et traitées, autant pour leur propre bien que pour celui de leur entourage. En revanche, le médecin n'a pas le droit d'intervenir dans la vie de personnes qui n'ont pas besoin de soins. Contre toute attente, en 1993, l'étude la plus

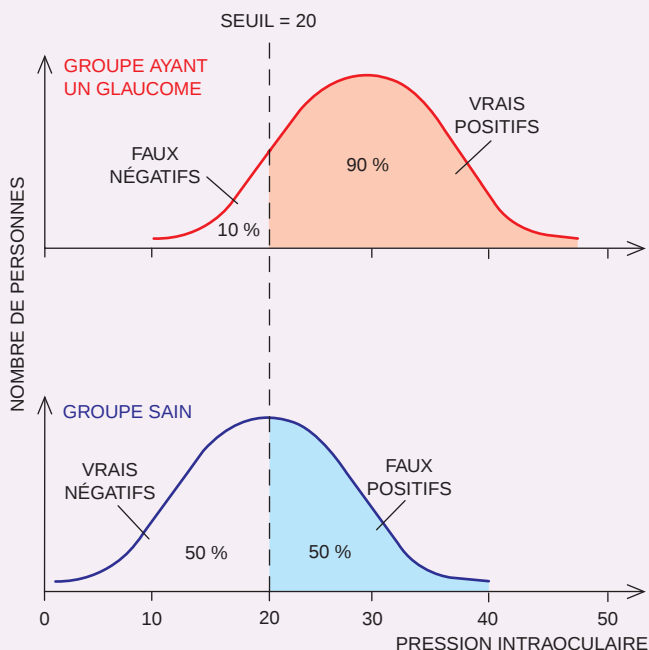
Améliorer la prise de décision en quatre étapes

Comment les médecins s'assurent-ils que le test diagnostique qu'ils utilisent est le plus pertinent, et qu'il distingue au mieux les individus malades de ceux qui sont sains ? La construction de courbes ROC est l'outil principal de mesure du degré de pertinence. L'étude d'un exemple permet de mieux comprendre son élaboration. Suivons les étapes d'une analyse qui évalue la pertinence du diagnostic du glaucome à partir d'une mesure de la pression intraoculaire des patients.

PREMIÈRE ÉTAPE : à partir d'un grand nombre de personnes dont on connaît la pression intraoculaire et la présence, ou non, d'un glaucome, on sépare les sujets sains et ceux qui sont atteints d'un glaucome. On marque sur un graphique le nombre d'individus pour chaque valeur de la pression intraoculaire. Le graphique de cette population montre que les pressions comprises entre 10 et 40 ne révèlent pas de façon fiable les personnes qui souffrent d'un glaucome.



DEUXIÈME ÉTAPE : pour chaque valeur de la pression, on détermine la probabilité qu'un diagnostic positif soit correct pour les patients. Cette probabilité est déterminée à partir de la fraction de l'échantillon pour laquelle la prédiction est exacte. L'aire totale sous la courbe représente 100 pour cent de chaque population. Si le seuil de pression intraoculaire choisi, pour la détermination des malades, est de 20, alors 90 pour cent des personnes ayant réellement un glaucome sont correctement diagnostiquées (vrais positifs) ; en revanche, 50 pour cent de personnes saines sont à tort déclarées atteintes (faux positifs).



complexe d'évaluation clinique par les méthodes statistiques a présenté un étonnant manque de pertinence : le diagnostic établi par des médecins qui examinaient des personnes amenées au service des urgences d'un grand hôpital psychiatrique était faux dans un cas sur deux, comme si on l'avait fait au hasard ; autrement dit, les psychiatres des urgences ne pouvaient absolument pas prévoir qui, parmi les malades, commettrait un acte violent au cours des six mois suivant le diagnostic.

Après cette étude, les biostatisticiens ont mis au point des méthodes statistiques prédictives pour évaluer le risque de violence. Le «guide d'évaluation du risque de violence», l'une des méthodes les plus étudiées, mesure douze variables, parmi lesquelles une évaluation du risque de développer une déficience mentale et l'inadapta-

tion à l'école primaire. Lors d'une vérification des prévisions établies par cette méthode, les délinquants sortis d'un hôpital psychiatrique étaient répartis en deux groupes : ceux pour lesquels le risque de récurrence violente était «élevé», ceux pour lesquels le risque était «faible». Cinquante-cinq pour cent des personnes du groupe à risque élevé ont récidivé, et seulement 19 pour cent l'ont fait dans le groupe à risque faible. La pertinence du guide était ainsi bien supérieure à celle d'un tirage au sort. Une méthode statistique prédictive plus récente s'est révélée encore meilleure pour la même prédiction. Toutefois, les méthodes statistiques restent controversées : les médecins doivent-ils les utiliser comme avis ou bien doivent-ils se conformer strictement aux indications des méthodes statistiques prédictives?

Les méthodes statistiques prédictives se révèlent également très utiles pour le dépistage du cancer du sein. Lors des examens classiques, les radiologues interprètent les mammographies de manière subjective.

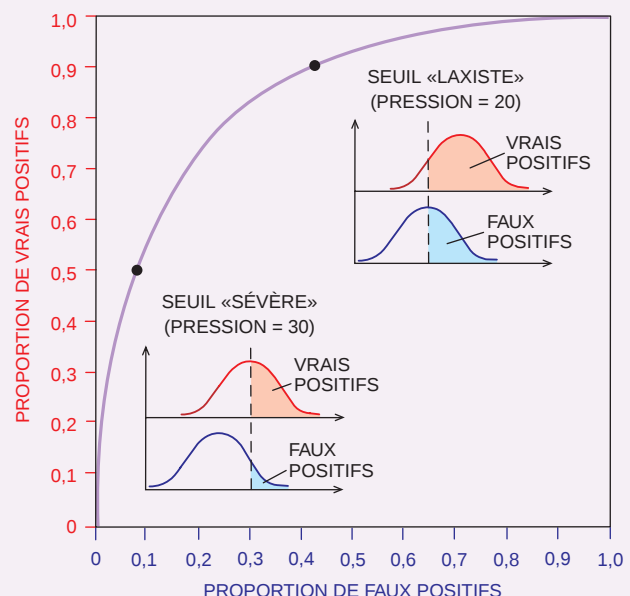
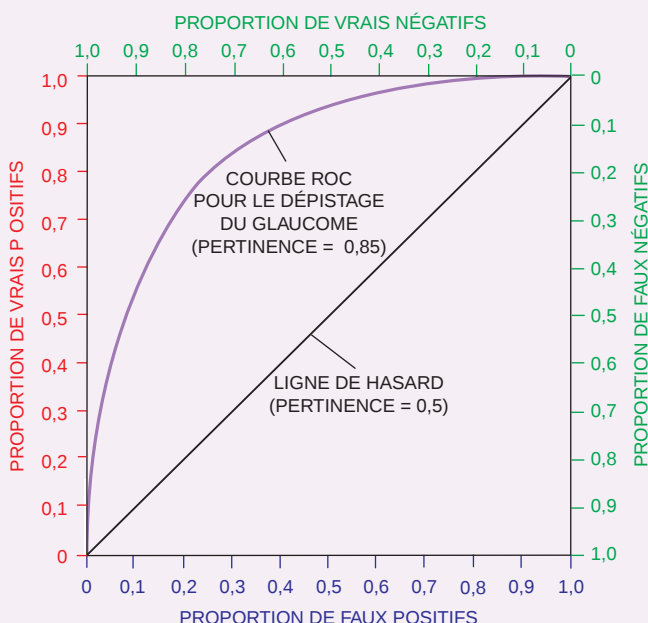
Le dépistage du cancer

Un protocole d'examen établi par des radiologues spécialisés, qui associe des méthodes statistiques prédictives, conduit à une interprétation plus pertinente : le protocole invite d'abord le radiologue à examiner des critères qualitatifs (tels que la régularité du contour de la masse), identifiés par des spécialistes du sein ; puis on calcule un risque de cancer du sein, dont on tient compte pour poser le diagnostic. Les informations apportées par l'analyse statistique sont une aide réelle : les radiologues non spécialisés ayant utilisé cette

TROISIÈME ÉTAPE : on construit la courbe ROC en reportant, pour chaque valeur seuil, la proportion de vrais positifs en fonction de celle des faux positifs. Si l'on obtient une droite, on doit conclure que le test diagnostic a 50 pour cent de chances de conduire au bon diagnostic : un tirage à pile ou face serait aussi efficace. Plus la courbe s'incurve, plus le test est pertinent (le rapport des vrais positifs sur les faux positifs augmente). La pertinence est l'aire de la surface sous la courbe ; elle augmente avec la courbure. Ici, le protocole de dépistage du glaucome est peu prédictif.

QUATRIÈME ÉTAPE : lorsque la pertinence atteinte est acceptable, on doit choisir le seuil pour un diagnostic oui/non. Le choix du seuil doit fournir une proportion de vrais positifs élevée sans entraîner une pro-

portion inacceptable de faux positifs. Chaque point de la courbe représente un seuil particulier, allant du plus «sévère», en bas à gauche, au plus «laxiste», en haut à droite. Les seuils «sévéres» (*cartouche inférieure*) limitent le nombre de faux positifs au prix de nombreuses personnes malades non diagnostiquées (forte proportion de faux négatifs, c'est-à-dire faible proportion de vrais positifs). Les seuils «laxistes» (*cartouche supérieure*) augmentent le nombre de personnes malades dépistées au prix de nombreux faux positifs (personnes saines diagnostiquées comme étant malades). Le seuil optimal pour une population donnée dépend de facteurs tels que la gravité de la maladie diagnostiquée, la prévalence de la maladie dans la population, la disponibilité de traitements pour ceux qui sont diagnostiqués, ainsi que les coûts économiques, sociaux ou émotionnels pour les faux positifs.



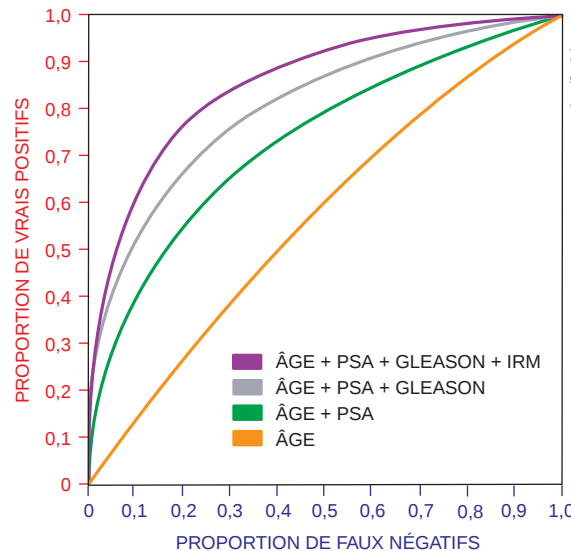
méthode ont une interprétation nettement plus pertinente, comparable en précision à celle des spécialistes.

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 fausement 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.

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.



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

Avec l'aimable autorisation de Tai Architects, Yamagisawa, Tokyo

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.

