

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.

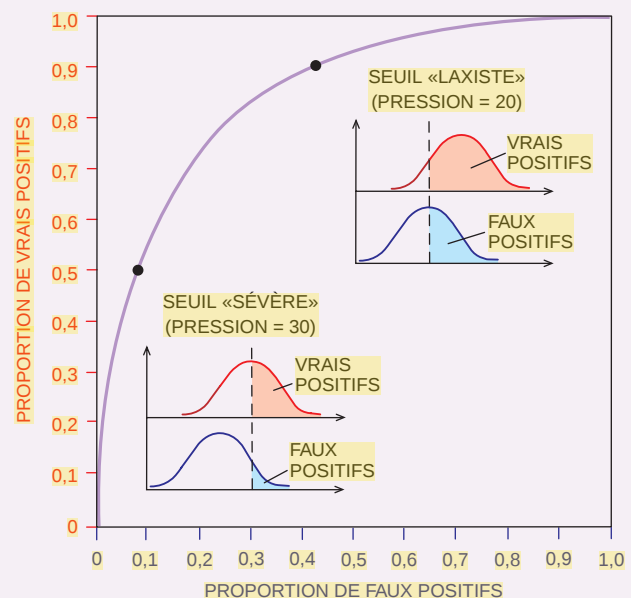
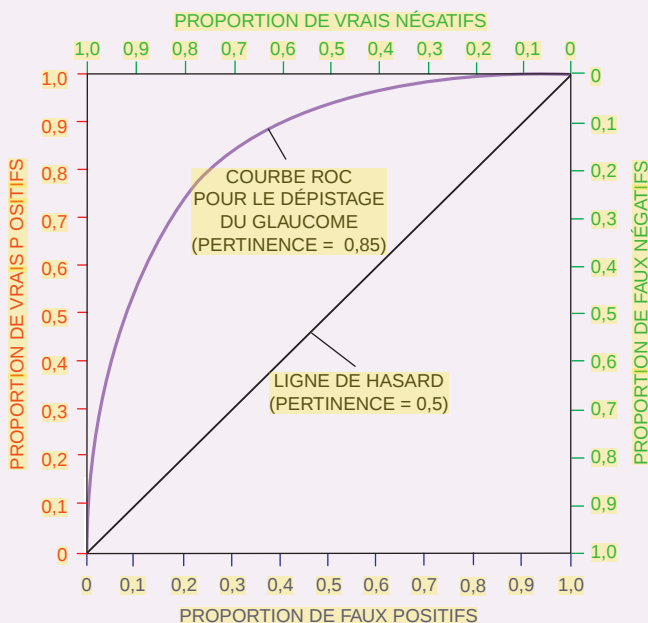
Le dépistage du cancer

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 diagnostique 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.



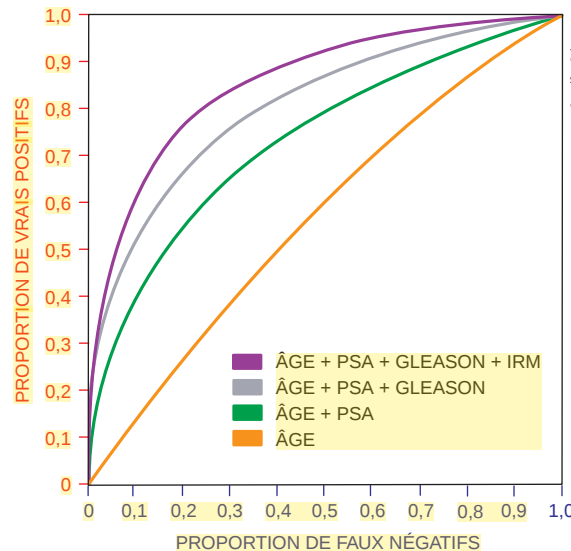
méthode ont une interprétation nettement plus pertinente, comparable en précision à celle des spécialistes.

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