Dépistage systématique ?

On étudie la mise en place d'un test diagnostique qui doit permettre de dépister une maladie m dont la fréquence (ou *prévalence*) dans la population est notée p (avec $0 < p < 1$).

On prélève au hasard une personne ayant été soumise au test.

On définit les événements suivants :
– T : « le test est positif » ;
– M : « la maladie est présente ».

Pour tout test, le fabricant indique :
– la probabilité $P_M(T)$ qu'un individu malade ait un test positif, appelé *sensibilité* du test, notée S_e ;
– la probabilité $P_{\overline{M}}(\overline{T})$ qu'un individu sain ait un test négatif, appelée *spécificité* du test, notée S_p.

Le test est idéal quand $S_e = S_p = 1$.

1. Illustrer par un arbre pondéré ; y faire figurer les probabilités p, S_e et S_p et le compléter.

2. **a.** Exprimer $P(M \cap T)$, $P(M \cap \overline{T})$, $P(\overline{M} \cap T)$ et $p(\overline{M} \cap \overline{T})$ en fonction de p, S_e et S_p.

b. Montrer que la probabilité que le test délivre une juste conclusion est égale à : $p(S_e - S_p) + S_p$.

3. On appelle :
– **valeur prédictive positive** du test (VPP), la probabilité $P_T(M)$ d'être malade, sachant que le test est positif ;
– **valeur prédictive négative** du test (VPT), la probabilité $P_{\overline{T}}(\overline{M})$ d'être non malade, sachant que le test est négatif.

a. Calculer $P(T)$.

b. Exprimer les deux valeurs prédictives VPP et VPN en fonction de p, S_e et S_p.

c. Le test apporte une information intéressante si VPP>p. Montrer que dans ce cas : $S_e + S_p > 1$.

4. Application au diagnostic du paludisme

La prévalence du paludisme est de 90 % en Afrique et de $\frac{1}{1\,000}$ en France.
Le test biologique utilisé a pour sensibilité $S_e = 0,95$ et pour spécificité $S_p = 0,85$.

a. Calculer VPP et VPN pour l'Afrique et pour la France.

b. En déduire ce que l'on peut dire en terme de probabilités à un patient africain et à un patient français selon que son test est positif ou négatif.

5. Influence de la prévalence sur VPP et VPN

En conservant les caractéristiques du test précédent ($S_e = 0,95$ et $S_p = 0,85$), on considère les fonctions v et w définie sur $]0\,;1[$ par : $v(p) = P_T(M)$ et $w(p) = P_{\overline{T}}(\overline{M})$.

a. Donner les expressions de $v(p)$ et de $w(p)$ en fonction de p.

b. Sur une calculatrice ou un tableur, tabuler les valeurs de $v(p)$ et de $w(p)$ pour p variant de 0,01 à 0,99, avec un pas de $\frac{1}{100}$ et obtenir les deux courbes.

6. Répondre aux questions suivantes, en argumentant :

a. Que peut-on dire de l'influence de la prévalence p sur les paramètres VPP et VPN ?

b. Lorsque p n'est pas trop faible, en quoi la positivité du test est-elle un élément important du diagnostic ?

c. Lorsque p est faible, $v(p)$ l'est aussi ; pourquoi le Rapport de Vraisemblance : RV= $\dfrac{P_M(T)}{P_{\overline{M}}(T)}$ est-il à prendre en compte dans le processus de diagnostique ?

d. Pour une maladie rare, quels inconvénients présente un test de dépistage systématique de toute une population ?

Éléments de correction

1.

2. **a.** $P(M \cap T) = pS_e \quad P(M \cap \overline{T}) = p(1 - S_e) \quad P(\overline{M} \cap T) = (1 - p)(1 - S_p) \quad P(\overline{M} \cap \overline{T}) = (1 - p)S_p$

b. $P(\text{« le test est correct »}) = P(M \cap T) + P(\overline{M} \cap \overline{T}) = pS_e + (1 - p)S_p = p(S_e - S_p) + S_p$

3. **a.** $P(T) = pS_e + (1 - p)(1 - S_p)$

b. $VPP = P_T(M) = \dfrac{P(M \cap T)}{P(T)} = \dfrac{pS_e}{pS_e + (1 - p)(1 - S_p)}$

$VPN = P_{\overline{T}}\overline{M} = \dfrac{P(\overline{M} \cap \overline{T}}{P(\overline{T}} = \dfrac{(1 - p)S_p}{1 - pS_e - (1 - p)(1 - S_p)}$

c. $VPP > p \iff S_e > pS_e + pS_p - p + 1 - S_p \iff S_e(1 - p) + S_p(1 - p) > 1 - p \iff S_e + S_p > 1$ (étant donné que $1 - p > 0$)

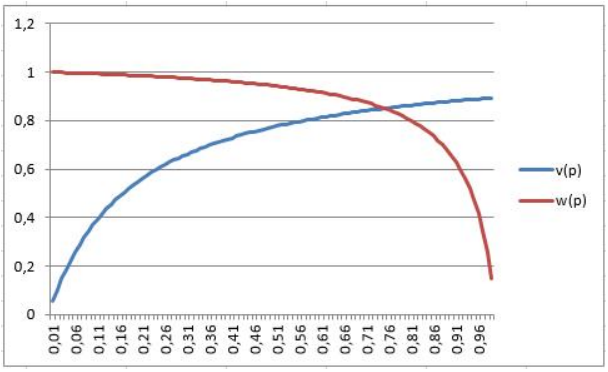
a. Pour l'Afrique : $VPP \approx 0,982$ et $VPN \approx 0,654$
Pour la France : $VPP \approx 0,006$ et $VPN \approx 0,999$

b. Un patient en Afrique avec un test positif est presque sûr d'être atteint ; un patient français avec un test négatif est presque sûr d'être sain.

4. **a.** Pour l'Afrique : $v(p) = \dfrac{0,855}{0,855 + (1 - p)0,15}$ et $w(p) = \dfrac{0,85(1 - p)}{0,145 - (1 - p)0,15}$

Pour la France : $v(p) = \dfrac{0,00095}{0,00095 + (1 - p)0,15}$ et $w(p) = \dfrac{0,85(1 - p)}{0,99905 - (1 - p)0,15}$

b.



5. **a.** Avec p faible : VPN est très bonne mais VPP est mauvaise.
Avec p grand : VPN est mauvaise mais VPP est très bonne.

b. si p est grand (proche de 1) : avec un test positif, on est presque sûr que la personne est malade.

c. si p est faible (proche de 0) : VPP est un mauvais indicateur.

$RV = \dfrac{P_M(T)}{P_{\overline{M}}(T)} = \dfrac{S_e}{1 - S_p} = \dfrac{0,95}{0,15} = 6,3$: dans l'idéal, RV est « infini » ; s'il est grand, c'est que le test a du sens. S'il est trop petit, on ne peut pas apporter de conclusion fiable à la suite du test.

d. Dans ce cas, on aura de nombreux « faux positifs » ce qui peut entraîner des problèmes :
- psychologiques, au moment où on annonce à quelqu'un qu'il est sûrement atteint par une maladie ;
- de suivi, dans la mesure où il faudrait dans l'idéal effectuer d'autres tests sur le patient pour éviter un traitement inutile (voir à effets secondaires pénibles) ;
- autres ...