

Evaluation des lecteurs de mammographies de dépistage : nécessaire démarche qualité du programme en parallèle de la dématérialisation, indicateurs indispensables à l'évolution d'une 2^{ème} lecture non inférieure à l'IA.

N. Sellier, A. Koivoigui, T. Le Trong, A. Bernoux, G. Abihsera, H. Delattre, D. Lamarque

Contexte : Le modèle français de dépistage s'appuie sur une 2^{ème} lecture (L2) non indépendante qui est pourtant une recommandation européenne. L'intégration de l'IA dans le parcours du dépistage pose la même problématique.

L'évaluation des 2^{èmes} lecteurs radiologues (L2R) constitue un double enjeu médical et économique alors que le bénéfice de la L2 diminue relativement à l'amélioration continue des techniques et des radiologues.

La L2 devant être non inférieure à l'IA, il faut disposer d'indicateurs pour sélectionner des futurs 2^{èmes} lecteurs (L2R+) plus ou aussi pertinents que l'IA.

Matériel : 766 607 mammographies de dépistage (MG) de 2 départements d'Ile de France, 93 (439 770 MG retenues, 12 ans, 18 L2R) et 95 (326 837 MG retenues, 9 ans, 17 L2R) ayant bénéficié de la même 3^{ème} lecture d'expertise (L3E) permettant de réduire la positivité des L2R. Profil des lecteurs de dépistage : moyenne d'activité des L1 de 400 MG/an ; expérience de 5 000 à 50 000 MG cumulées pour les 35 L2R.

Résultats en référence du Gold Standard d'une 3^{ème} lecture non aveugle sans cancer d'intervalle (CI) connu pour les dossiers vus par le L3E :

- Nombre total de cancers détectés: 5829 dont 4480 en L1, 375 en L2 (6,43 %) et 974 cancers d'intervalle (16,7 % avec petite disparité départementale de 15 % et 19 %).
- Positivité des L2 à 2 % et positivité retenue après L3E de 1,16 %.
- Importantes variations du taux des positifs de L2 confirmés par la L3E : de 40 % à 78,1% dans le 93, de 30,9 % à 76,8% dans le 95 mais des moyennes identiques de 56,8 et 56,6 % et avec dans chacun des départements 4 lecteurs nettement au-dessus de la moyenne.
- Ces disparités de positivité des 2^{èmes} lecteurs (L2R) ne dépendent pas du nombre de lectures effectuées.
- Taux de rappel évité aux femmes de 42,5 % avec importantes variations selon les LR2 (de 20,7 à 59,9 %).
- Différence des taux de détection de cancers (TDC) et de cancers d'intervalle (TCI) des LR2 et des L1.
- Stabilité dans le temps de la disponibilité de l'antériorité MG : entre 55 et 58 %,
- Gestion du risque du L3E qui est applicable au L2R+ : différence statistiquement significative du comportement du L3 selon la disponibilité de l'antériorité, avec plus grande positivité en l'absence de MG antérieure (75% versus 49 % avec antériorité).

Interprétation des résultats :

- La positivité adéquate des L2R n'est pas dépendante de l'expérience quantitative.
- La lecture d'expertise (L3E ou L2R+) a un impact fort sur le taux de rappel.
- L'existence d'antériorité influe de façon significative et efficace sur la décision de positivité et sur le taux de cancer d'intervalle qui était dans notre étude 2,6 fois supérieur au taux de détection de la L2.
- L'absence d'antériorité est une défaillance du processus de dépistage qui met le L2R en situation d'infériorité.

Discussion : La positivité de la L2 se décline en deux catégories :

- la plus importante, celle des ACR 0 garant de la vigilance et de la qualité du dépistage au travers des compléments d'examen qui sont demandés aux 1^{ers} lecteurs. L'IA y ajoute maintenant un rôle d'alerte.
- celle réduite des ACR 4 ou 5, visant à diminuer le taux de cancer d'intervalle. L'IA y joue sa part importante en les réduisant de près de moitié.

La haute spécificité du L2R+ est applicable à l'IA après scoring des MG par IA de triage.

Méthodes d'évaluation et évaluation souhaitable :

Sans dématérialisation et en l'absence de registre, l'évaluation individuelle actuelle basée sur les positifs et le Bi-Rads (notamment ACR 3 des L1 et ACR 0 des L2R), sur les taux de rappel et leur évolution dans le temps, est insuffisante.

La dématérialisation et l'environnement numérique de santé doivent permettre de corriger ces manquements :

- disponibilité de l'antériorité pour ne pas mettre le L2R en situation d'infériorité, d'autant plus que l'IA intègre maintenant la comparaison des scores des MG antérieures.
- récupération des data ce que va permettre le DMP et le projet DRIM-box national : VPP de biopsie et VPP de rappel pour une taille inférieure à 10 mm (cancers subtils).
- TCI par L2 et analyse rétrospective experte des MG avec la réserve que les vrais CI sont plus agressifs que les cancers détectés du dépistage.
- analyse des épisodes de sensibilité en incidence et par classes d'âge.
- index composite histologique (grade de CCIS, grade tumoral SBR, taille < 10 mm, lobulaire infiltrant).
- évaluation de l'impact global de l'IA sur le comportement des lecteurs et l'arbitrage des cas discordants.

Pour la fiabilité des indicateurs, l'ouverture des résultats du L1 en fin de séance de L2 classique ou après IA de triage serait souhaitable.

Conclusion : L'évaluation et la dématérialisation constituent la voie d'amélioration du dépistage. Le profil des L2R+ avec la diffusion de l'IA sera déterminant (forme d'accréditation et de label qualité) sachant qu'en L1, l'accès de l'IA à l'antériorité jouera un rôle d'alerte important. L'effort de nos pratiques pour améliorer la sensibilité et contrôler le taux de rappel, doit porter sur l'observance du dépistage, la disponibilité pour le L2R de l'antériorité MG et la communication entre lecteurs des discordances.