

# **Performance des lectures radiologiques dans le programme valaisan de dépistage du cancer du sein**

Lausanne, février 2006

Dr PD Jean-Luc Bulliard

Prof. Fabio Levi

Unité d'épidémiologie du cancer

Institut universitaire de médecine sociale et préventive, Lausanne

# 1. Introduction

## 1.1. Préambule

Le critère ultime de l'efficacité d'un Programme de dépistage du cancer du sein par mammographie est la baisse de la mortalité de ce cancer attribuable au programme dans la population concernée. Comme cet impact n'est observable qu'à long terme, une évaluation périodique des performances intermédiaires d'un programme, selon des normes codifiées<sup>1</sup> visant à garantir la qualité et l'efficacité du dépistage par mammographie tout en minimisant ses effets adverses, est indispensable. En Suisse, une telle évaluation des programmes de dépistage est obligatoire<sup>2</sup>. L'expérience pilote, conduite au milieu des années 90 dans 3 districts vaudois, a démontré que les normes de qualité et d'efficacité fixées au niveau européen étaient, en règle générale, atteignables dans le contexte suisse<sup>3</sup>.

L'évaluation du Programme valaisan de dépistage du cancer du sein est assurée depuis son démarrage en 1999 par l'Unité d'épidémiologie du cancer (UEC) de l'Institut universitaire de médecine sociale et préventive (IUMSP) de Lausanne. Cette unité est responsable du suivi épidémiologique de 3 programmes cantonaux de dépistage (Vaud, Valais et Jura) et participe activement aux activités du réseau international des programmes de dépistage du cancer du sein (International Breast cancer Screening Network). Le précédent rapport d'évaluation d'août 2004 (RE2004)<sup>4</sup> relevait notamment la participation régulière des femmes, l'excellente qualité et l'adéquation des investigations médicales complémentaires.

Bien qu'indiquant un niveau de performances globalement satisfaisant et proche des normes internationalement admises, le RE2004 soulignait la fréquence excessive de reconvoications des femmes lors de leur première participation, ce qui induit un nombre élevé de « faux positifs ». Parmi les pistes proposées pour améliorer la qualité des lectures, un examen de la stratégie en vigueur pour la lecture multiple était suggéré. La prise de conscience engendrée

---

<sup>1</sup> Europe Against Cancer. *European guidelines for quality assurance in mammography screening*. 3rd ed, Perry N, Broeders M, de Wolf C, et al. eds, 2001, Luxembourg: Office for Official Publications of the European Communities, 366 p.

<sup>2</sup> Ordonnance Fédérale du 23 juin 1999 sur la garantie de la qualité des programmes de dépistage du cancer du sein réalisés par mammographie

<sup>3</sup> Bulliard J-L, De Landtsheer J-P, Levi F. Results from the Swiss mammography screening pilot programme. *Eur J Cancer* 2003,**38** :1760-68.

<sup>4</sup> Bulliard J-L, Levi F. Programme valaisan de dépistage du cancer du sein par mammographie: Evaluation des deux premiers tours de dépistage (1999-2004). IUMSP, Lausanne, 2004. (téléchargeable depuis le site [www.iumsp.ch](http://www.iumsp.ch))

par le RE2004 au niveau des radiologues et de la direction du Programme valaisan est à l'origine de ce rapport thématique<sup>5</sup>.

## 1.2. Contenu du rapport

L'objectif principal de ce travail est de fournir à la direction du Programme valaisan les éléments nécessaires pour comprendre les déterminants de la performance des lectures radiologiques et estimer l'effet de toute mesure éventuelle d'amélioration. Pour cela, ce rapport vise à répondre à 4 questions. A savoir :

- 1) Quelle a été l'évolution récente des indicateurs d'efficacité du Programme ?
- 2) Pourquoi le taux de reconvoction est-il élevé en tour prévalent ?
- 3) Quels facteurs inhérents au Programme influencent les performances des lectures ?
- 4) Quelles mesures permettraient d'agir sur les performances du Programme ?

Ce document se compose de 4 chapitres. Après un rappel du contexte de l'évaluation du dépistage du cancer du sein et de l'origine de ce rapport thématique (chapitre 1), l'efficacité du programme valaisan de dépistage, après presque 6 ans d'activité, est discutée (chapitre 2). Ce deuxième chapitre décrit les performances du programme en s'intéressant surtout à l'évolution récente des quelques indicateurs d'efficacité pour lesquels des efforts ont été entrepris pour améliorer l'adéquation aux normes européennes (question 1 ci-dessus). Ce chapitre comprend l'évaluation des cancers diagnostiqués entre 2 invitations chez les participantes (cancers dits d'intervalle).

La qualité des lectures radiologiques est examinée de manière détaillée dans le chapitre 3, qui forme l'ossature du rapport (questions 2 à 4). Les résultats des lectures y sont analysés séparément selon qu'il s'agisse d'une première ou deuxième lecture, d'un tour prévalent ou incident, et du niveau d'expérience du radiologue (volume de lectures, 1<sup>e</sup> ou 2<sup>e</sup> lecteur). Les conclusions et recommandations sont présentées dans le chapitre 4. Un bref glossaire des termes les plus importants complète ce rapport.

---

<sup>5</sup> Avenant du 30.11.2005 au mandat d'évaluation entre l'IUMSP et la direction du Programme valaisan de dépistage du cancer du sein.

## 2. Surveillance de l'efficacité du dépistage

### 2.1. Considérations préliminaires

Les critères d'éligibilité des femmes pour l'évaluation sont identiques à ceux utilisés les années précédentes (voir RE2004). Les résultats présentés reflètent la situation du programme valaisan de dépistage au 21/06/2005 (date de l'extraction des données du Centre valaisan de dépistage (CVD)) et incluent les cancers d'intervalle parmi les participantes<sup>6</sup> des 3 premières années du programme (01/10/1999 au 30/09/2002), enregistrés par le Registre valaisan des Tumeurs (RVsT) au 22/06/2005 (date de l'appariement entre les fichiers du RVsT et du CVD). Les appariements ont été réalisés par le RVsT, avec un logiciel dont l'efficacité est avérée<sup>7</sup>. La procédure de croisement a fait l'objet de vérifications manuelles.

Les données considérées n'incluent pas l'ensemble des mammographies, puisque l'activité liée aux invitations survenues après le 31/03/2005 n'est pas prise en compte. Cette restriction est nécessaire pour s'assurer que le recul entre l'invitation au dépistage et la transmission des données pour l'évaluation permette la réalisation des mammographies et l'enregistrement des informations correspondantes. De même, l'identification et la documentation des cancers d'intervalle exigent un délai minimal de 2 ans après l'examen de dépistage.

Ce rapport comptabilise 45'089 mammographies effectuées auprès de 24'754 femmes, pour un total de 246 cancers dépistés chez 240 femmes (tableau 1). Ces mammographies se répartissent entre le premier (51%), le deuxième (38%) et le troisième tour (11%) d'invitations. La forte proportion de premières mammographies (tour prévalent) chez les femmes de 50 à 54 ans reflète la dynamique du programme qui intègre chaque année les Valaisannes atteignant 50 ans. En tour incident (participantes récurrentes), le plus jeune âge est 52 ans de sorte que le volume de mammographies réalisées et de cancers détectés dans la tranche d'âge 50-54 ans y sont plus modestes.

---

<sup>6</sup> Seules les participantes dûment autorisant la transmission de leurs données à des fins statistiques et d'amélioration de la qualité du programme de dépistage ont été considérées.

<sup>7</sup> Dal Maso L, Braga C, Franceschi S. Methodology used for "software for automated linkage in Italy" (SALI). *J Biomed Inform* 2001;**34**:387-95.

Tableau 1 : Répartition des mammographies effectuées\* et des cancers dépistés par classe d'âge et tour de dépistage

Classe d'âge	Nb. mammographies		Nb. cancers dépistés	
	prévalent	incident	prévalent	incident
50-54 ans	9'827	3'483	36	7
55-59 ans	5'985	5'196	38	16
60-64 ans	4'819	5'876	40	37
65- 70 ans	4'123	5'780	40	32
<b>Total</b>	<b>24'754</b>	<b>20'335</b>	<b>154</b>	<b>92</b>

\* Mammographies réalisées entre le 01/10/1999 et le 30/04/2005.

A noter que 76% des 10'855 mammographies réalisées au cours des 13 mois séparant cette analyse du RE2004 portaient sur des participantes récurrentes. A l'instar, le nombre de nouveaux cancers dépistés en l'espace de 13 mois s'est accru de 10% en tour prévalent (154 contre 140) alors qu'il a doublé en tour incident (92 contre 44). La croissance de la part de l'activité de dépistage dévolue aux mammographies incidentes devrait se poursuivre et il faut anticiper dès 2006 un plus grand nombre total d'examen en tour incident qu'en tour prévalent.

## 2.2. Indicateurs d'efficacité précoce du dépistage

Les normes d'efficacité du dépistage distinguent les tours prévalent et incident. Il est en effet admis que les femmes qui reviennent au dépistage « bénéficient » de l'effet de leur comportement en matière de prévention. Le risque de cancer du sein dans une population asymptomatique, régulièrement suivie, est plus bas que pour une population de premières participantes (tour prévalent) ; de plus, les cancers détectés en tour incident présentent généralement un pronostic plus favorable de par l'élimination des cancers « prévalents » (tableau 2).

Tableau 2: Evolution des indicateurs d'efficacité précoce du Programme valaisan de dépistage du cancer du sein et comparaisons avec les normes européennes\*

Indicateurs d'efficacité	Tour		Tendance récente**	Normes européennes	
	prévalent	incident		prévalent	incident
<b>Cancers dépistés</b>					
Taux de détection*** (‰)	6,2	4,5	Favorable	≥8,8	4,4
In situ (%)	13,6	15,2	Stable	10-20	10-20
Invasifs ≤1 cm (%)	25,6	21,4	Favorable	≥20	≥25
Stade ≥II (%)	33,6	36,8	Favorable	≤25	≤20
Sans atteinte gangl. (%)	75,2	76,9	Stable	≥70	≥75
<b>Cancers d'intervalle****</b>					
Première année (%)	30,8		Défavorable	≤30	
Deuxième année (%)	61,6		Stable	≤50	

\* Les normes européennes s'appliquent aux femmes de 50 à 64 ans, alors que les indicateurs pour le Programme valaisan couvrent la population féminine de 50 à 70 ans. Des différences peuvent en résulter dans la valeur de certains indicateurs.

\*\* Appréciation par les évaluateurs de la tendance de la dernière année par rapport au RE2004.

\*\*\* L'incidence du cancer du sein en Valais, pour la période quinquennale précédant le démarrage du programme (1994-98), a été considérée comme incidence en absence de dépistage.

\*\*\*\* Exprimé en % du taux d'incidence « en l'absence de dépistage » (voir point \*\*\*); le complément (à 100%) de l'incidence proportionnelle correspond à la sensibilité du programme. Les taux de cancers d'intervalle pour la première et deuxième année sont respectivement de 0,9‰ (16 cas) et 1,8‰ (32 cas).

En tour prévalent, 1 cancer est détecté pour 161 mammographies réalisées (1 pour 221 en tour incident). Toutefois, l'appréciation de l'adéquation du taux de détection requiert de la prudence puisque les normes européennes sont basées sur une incidence du cancer du sein en l'absence de dépistage, une situation donc purement théorique. Un dépistage individuel s'est progressivement développé en Valais, avant même le démarrage du programme (34% des Valaisannes de 50 à 69 ans déclaraient en 1997 avoir eu une mammographie au cours des 2 dernières années<sup>8</sup>). Le manque de données fiables sur l'ampleur de ce phénomène rend difficile l'estimation de l'incidence attendue entre 1999 et 2004 « en l'absence de dépistage ». A cet égard, la norme en tour incident reflète mieux le contexte valaisan du dépistage. *L'évolution de cet indicateur d'efficacité montre, pour la première fois dans*

<sup>8</sup> OFS, Enquête suisse sur la Santé, 1997.

*l'histoire du programme, que la norme minimale requise pour le taux de détection est atteinte (4,5‰ contre 4,4‰).*

Le degré d'extension des cancers détectés lors de la première participation demeure en adéquation avec les recommandations européennes, à l'exception de la proportion de tumeurs à un stade avancé (33,6% contre 25%). En tour incident, les proportions élevée de cancers à un stade avancé (36,8% contre 20%) et faible de cancers invasifs de petite taille (21,4% contre 25%) avaient déjà été relevées et commentées (RE2004). La tendance récente indique une amélioration du profil pronostique depuis l'année précédente, de sorte que le degré d'extension des tumeurs est à présent comparable entre les tours prévalent et incident (les différences ne sont pas statistiquement significatives). *Ces signes encourageants méritent cependant un effort et une surveillance continue afin de s'assurer, à moyen terme, que les indicateurs d'efficacité précoce pour les tumeurs « incidentes » atteignent les normes européennes.* Signalons qu'une amélioration du profil pronostique entre les tours prévalent et incident n'est pas systématiquement observée dans les principaux programmes de dépistage, sans que cela paraisse - de prime abord - préjudiciable à l'efficacité de ces programmes<sup>9</sup>.

Comme l'intervalle observé entre 2 mammographies est de l'ordre de 24 mois, seuls les cancers diagnostiqués dans les 2 ans suivant la mammographie de dépistage ont été comptabilisés comme cancers d'intervalle (48 cancers pour un taux global de cancers d'intervalle de 2,7‰). Ce nombre de cancers d'intervalle, couvrant les 3 années initiales du programme (01/10/1999 au 30/09/2002), est trop modeste pour tirer des conclusions quant à son adéquation aux normes européennes. A cet égard, les tendances récentes ne sont fournies qu'à titre d'indication générale. La fréquence des cancers d'intervalle dans les 12 mois qui suivent la mammographie représente environ 30% de l'incidence estimée en l'absence de dépistage alors qu'en deuxième année, elle représente les 60% de cette incidence (moyenne biennale : 46,2%). Enfin, 2 des 48 cancers d'intervalle concernent des femmes chez qui les investigations ont infirmé la suspicion de lésion indiquée par le dépistage (« faux négatifs » des investigations).

---

<sup>9</sup> Boer R, de Koning H, van Oortmarssen G, Warmerdam P, van der Maas P. Stage distribution at first and repeat examinations in breast cancer screening. *J Med Screen* 1999,**6**:132-8.  
Sarkeala T, Anttila A, Saarenmaa I, Hakama M. Validity of process indicators of screening for breast cancer to predict mortality reduction. *J Med Screen* 2005,**12**:33-7.

### 3. Indicateurs de qualité des lectures

#### 3.1. Qualité globale et évolution temporelle

Les normes de qualité radiologique du dépistage communément admises sont plus strictes en tour incident qu'en tour prévalent. En effet, les radiologues disposent en tour incident de l'iconographie de l'examen précédent. Les clichés comparatifs aident souvent à interpréter la mammographie et permettent ainsi de réduire substantiellement le taux de reconvoction. Le tableau 3 décrit les indicateurs de qualité des lectures du programme valaisan. La persistance d'un bon niveau de performance pour tous les indicateurs de qualité des investigations liées au dépistage (rendement biopsique, taux de cytoponction avec matériel suffisant, de biopsie bénigne, etc...) a été vérifiée ; ces indicateurs ne sont pas présentés.

Tableau 3: Indicateurs de qualité des lectures du Programme valaisan de dépistage du cancer du sein et comparaisons avec les normes recommandées

Indicateurs de qualité des lectures (%)	Tour		Valeurs recommandées *	
	prévalent	incident	prévalent	incident
Taux de 3 <sup>e</sup> lecture**	14,9	7,2	≤10	≤ 7
Taux de reconvoction	7,0	3,3	<7	<5
VPP du test de dépistage	8,7	14,0	≥12,5	≥12,5

\* Les normes européennes sont indiquées pour le taux de reconvoction ; en l'absence de normes internationales pour les autres indicateurs, les recommandations sont basées sur l'expérience des évaluateurs et des programmes cantonaux<sup>10</sup>.

\*\* Cet indicateur peut varier selon les critères utilisés dans la grille de lecture radiologique ; la valeur proposée est avant tout indicative.

En première participation, le taux de reconvoction, tous âges confondus, se maintient à la limite supérieure des recommandations européennes (5 à 7%). Il est trop élevé chez les 50-54 ans (8,4%), et d'environ 6% dans les autres catégories d'âge (données non présentées). En tour incident, ce taux diminue remarquablement et satisfait l'objectif fixé (3 à 5%).

---

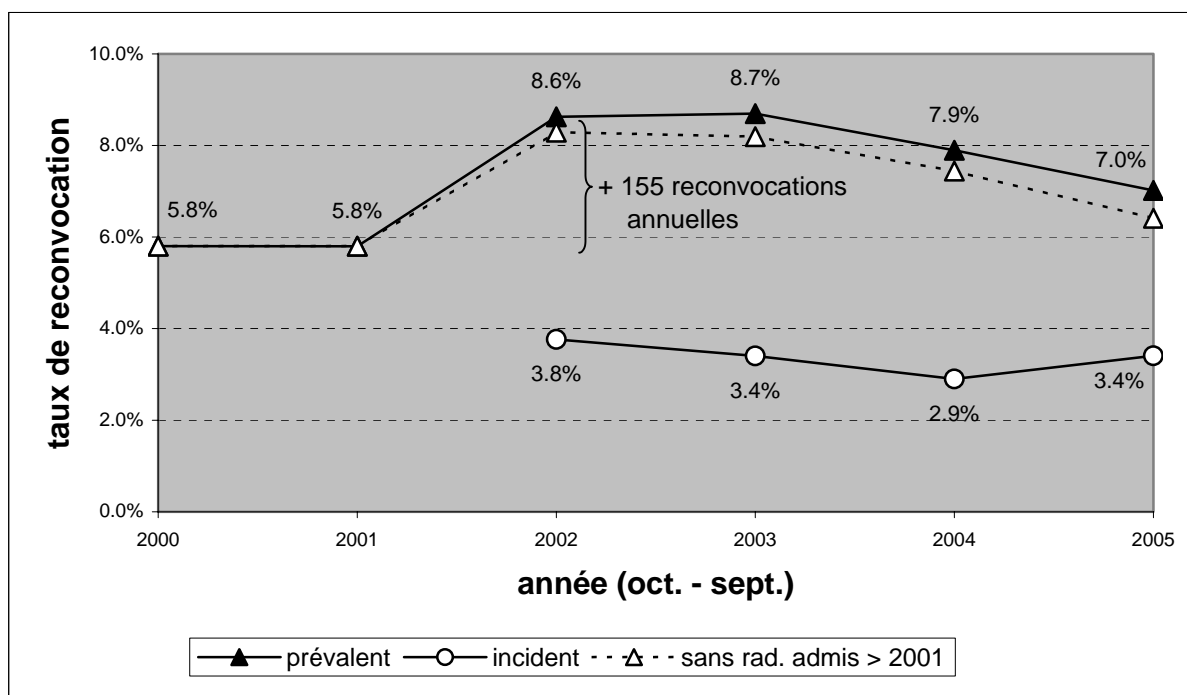
<sup>10</sup> De Landtsheer J-P, Delanoy Ortega B, Jemelin C. Dépistage du cancer du sein: analyse comparée de trois programmes suisses. *Méd Hyg* 2000;**58**:1407-10.  
Programme vaudois de dépistage du cancer du sein : Rapport d'activité 2004.  
Programme genevois de dépistage du cancer du sein : Rapport d'activité 2004.



La proportion de 3<sup>e</sup> lectures arbitrales, effectuées lors de conclusions radiologiques discordantes, influence grandement le taux de reconvoication. De manière générale, la 3<sup>e</sup> lecture permet, en Valais, de réduire de moitié environ les convocations pour des examens complémentaires. Ainsi, un taux (trop) élevé de lectures discordantes favorise un haut taux de reconvoication.

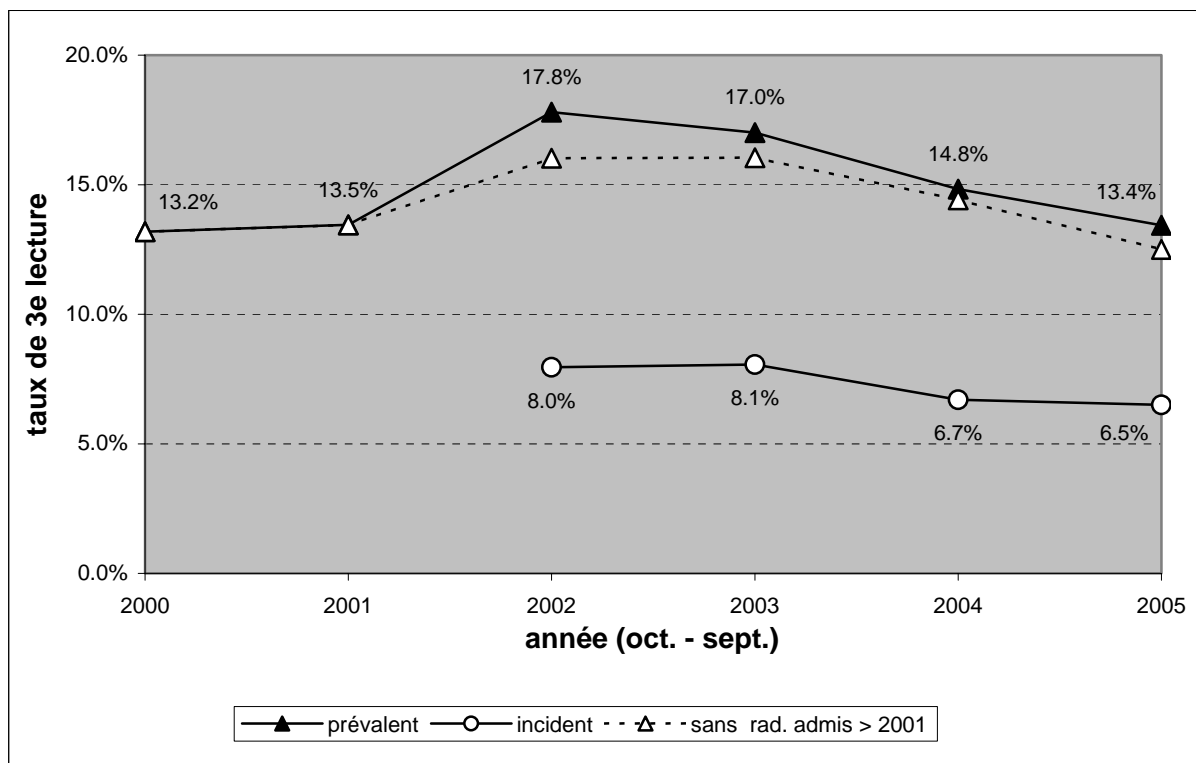
L'évolution chronologique du taux de reconvoication montre une forte augmentation, en tour prévalent, après les 2 premières années de dépistage (10/1999-09/2001) (Figure 1). Cet accroissement, de près de 50% au cours des 3<sup>e</sup> et 4<sup>e</sup> année d'activité du programme (2002-3), s'est traduit par un supplément annuel de 155 femmes rappelées pour des investigations complémentaires. Il reflète largement l'augmentation concomitante du taux de 3<sup>e</sup> lecture en tour prévalent (Figure 2).

Figure 1 : Tendence chronologique du taux de reconvoication en tour prévalent et incident



Ces évolutions temporelles sont observées dans chaque tranche d'âge quinquennale (données non présentées). Ainsi, le fait que les femmes de 50 à 54 ans représentent une part croissante des examens prévalents dès 2002 n'explique pas les accroissements des taux de reconvoication et de 3<sup>e</sup> lecture après les 2 premières années de dépistage.

Figure 2 : Tendence chronologique du taux de 3<sup>e</sup> lecture en tour prévalent et incident



Plusieurs causes semblent expliquer ce phénomène. Les principales sont :

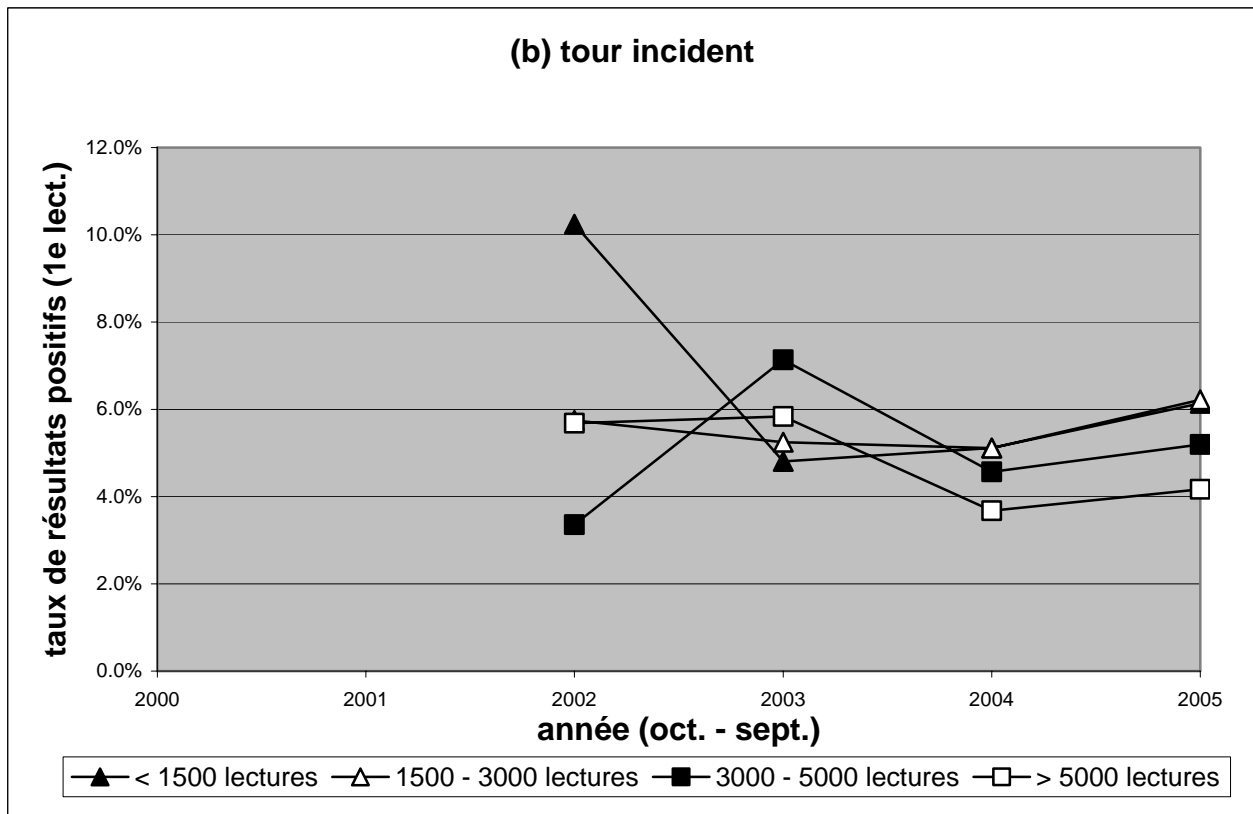
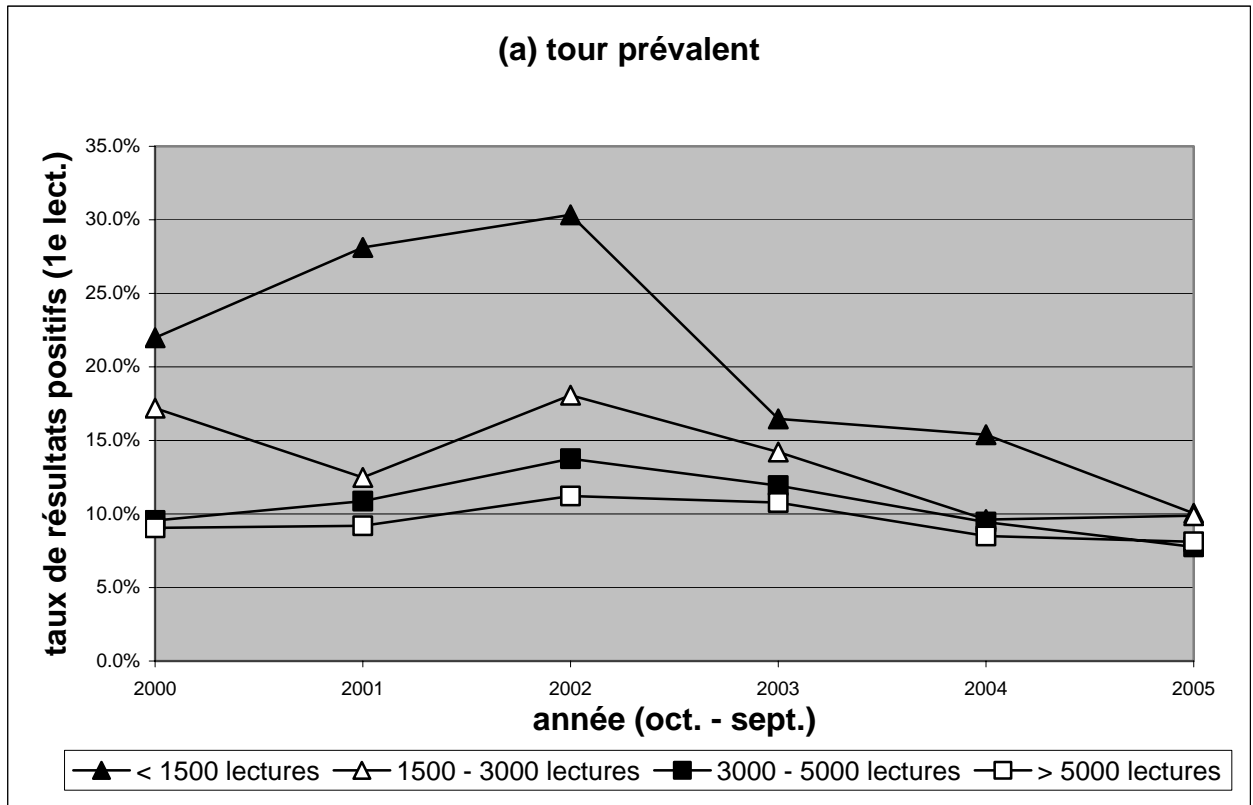
- 8 nouveaux radiologues ont été accrédités entre octobre 2001 et septembre 2003 ; cet apport de 50% de lecteurs supplémentaires explique, à lui seul, un tiers de l'augmentation du taux de 3<sup>e</sup> lecture (Figure 2, trait pointillé). En revanche, il influence modérément le taux de reconvoication (Figure 1, trait pointillé) qui serait, en tour prévalent et sans ces nouveaux lecteurs, de 6,7% (au lieu de 7,0%), soit dans la fourchette des recommandations européennes (5 à 7%).
- Le premier retour d'information annuel reçu par les radiologues sur leurs performances individuelles remonte à octobre 2001. Toute « réaction » afin de réduire la probabilité de manquer une tumeur se traduit par une augmentation à court terme des mammographies positivées. Ce changement de comportement a été observé dans les programmes vaudois et valaisan, indépendamment du contenu des recommandations individuelles accompagnant ce retour d'information (un faux négatif « pèse » plus qu'un faux positif pour un clinicien). Ce facteur subjectif n'est pas directement quantifiable.
- L'avènement dès octobre 2001 des lectures incidentes a pu rendre soudainement plus difficile une interprétation sans comparatifs iconographiques, en particulier pour les lecteurs moins expérimentés. Cette hypothèse est corroborée par l'évolution chronologique du taux de mammographies positivées selon le volume total de lectures effectuées (voir Figure 3).

La diminution des taux de reconvoication et de 3<sup>e</sup> lecture en tour prévalent, amorcée en 2004, est réjouissante. *Cette tendance devra toutefois se poursuivre avant qu'un impact sur le taux global de reconvoication devienne perceptible.* Cette amélioration est essentiellement due à une baisse appréciable du taux de mammographies positivées en 1<sup>ère</sup> lecture par les radiologues les moins expérimentés (Figure 3). En tour incident, le taux de reconvoication est relativement stable tant sur le plan de la variation chronologique (Figure 1) que sur celui de l'âge (entre 3,0 et 3,6% selon la tranche d'âge ; données non présentées). Si elle se confirme, la tendance actuelle va également vers une homogénéisation des taux de résultats positifs en 1<sup>ère</sup> lecture, indépendamment du niveau d'expérience du radiologue (Figure 3).

Relevons ici l'importance de la prise en compte du type de tour de dépistage dans les indicateurs d'activité du programme. Une statistique générale, tous tours confondus, indiquerait par exemple des taux annuels de reconvoication de 7,1% (2002), 5,1% (2003), 4,2% (2004) et 4,3% (2005), occultant les tendances observables sur la Figure 1. Cette apparente contradiction s'explique par la contribution grandissante dans ces calculs des mammographies incidentes, dont les résultats sont plus souvent négatifs que pour les mammographies prévalentes.

La valeur prédictive positive (VPP) de la mammographie est un indicateur d'une grande importance pratique. En moyenne, sur 10 valaisannes dont la mammographie a été positivée, une est diagnostiquée avec un cancer. La VPP augmente en tour incident et, surtout, avec l'âge (tour prévalent : 6,1% et 14,4% chez les 50-59 et 60-69 ans ; tour incident : 8,2% et 18,4%, respectivement). Comparativement à d'autres programmes de dépistage, une VPP de l'ordre de 10% paraît plutôt faible.

Figure 3 : Evolution du taux de mammographies positivées en 1<sup>ère</sup> lecture selon le volume de lectures et le tour de dépistage



### 3.2. Déterminants de la qualité des lectures

La qualité des lectures d'un radiologue dépend d'une part des caractéristiques des patientes (niveau individuel de risque, statut ménopausique, densité du tissu mammaire, ...) et d'autre part de ses performances intrinsèques. La distribution de l'âge et de la densité mammaire ne révèle pas de gros déséquilibres entre les femmes vues par chaque radiologue. Il a ainsi été admis que, globalement, les mammographies interprétées par chaque radiologue présentaient le même degré de difficulté de lecture et étaient de qualité technique comparable. S'affranchir de cette hypothèse nécessiterait l'interprétation des mêmes clichés par tous, dans les mêmes conditions de travail, ou le recours à un modèle mathématique capturant tous ces facteurs et leurs interactions. La première solution requiert la mise sur pied d'une étude *ad hoc*, éloignée de la réalité du dépistage, avec relecture d'un nombre limité de mammographies. La deuxième implique un degré de complexité dépassant le cadre de ce rapport et une traduction opérationnelle des résultats statistiques pouvant s'avérer difficile sur le plan pratique.

Quatre facteurs influant sur la qualité de l'interprétation des mammographies ont été identifiés. Il s'agit du :

- 1) Type de tour (prévalent ou incident)
- 2) Contexte de lecture (première ou deuxième)
- 3) Type de lecteur (premier ou second)
- 4) Volume de lectures effectuées (expérience du radiologue)

Le type de tour de dépistage est prépondérant. L'apport de clichés comparatifs en tour incident augmente substantiellement les performances radiologiques et réduit les différences de performance, ceci indépendamment des 3 autres facteurs. Ce constat, le plus grand potentiel d'amélioration de la qualité des lectures en tour prévalent, et le fait que la direction du programme ne peut influencer sur le type de mammographies à interpréter (prévalente ou incidente) ont conduit à se concentrer sur une comparaison des résultats des lectures en tour prévalent (tableau 4). *En se focalisant sur le tour prévalent, l'analyse des facteurs de qualité met en exergue certains points moins favorables du programme.* Les enseignements tirés de cette analyse, ainsi que les mesures éventuelles visant à améliorer la qualité des lectures prévalentes bénéficieront aussi aux lectures en tour incident.

Tableau 4: Taux de faux positifs, de détection, de cancers non vus, et valeur prédictive positive par contexte de lecture, type de lecteur et nombre de mammographies lues (tour prévalent, chiffres en %)

Facteurs	Faux positifs		Détection		Ca. non vus		VPP	
	L1	L2	L1	L2	L1	L2	L1	L2
<b>Type de lecteur</b>								
1 <sup>er</sup> lecteur	12,4	-	0,58	-	0,07	-	4,5	-
2 <sup>e</sup> lecteur	10,3	6,8	0,50	0,53	0,06	0,08	4,7	7,2
<b>Nb. de lectures*</b>								
< 1500	18,1	-	0,56	-	0,08	-	3,0	-
1500 – 3000	13,5	11,6	0,63	0,21	0,08	0,21	4,4	1,8
3000 – 5000	10,5	8,8	0,53	0,49	0,08	0,06	4,8	5,3
> 5000	9,2	6,0	0,47	0,55	0,04	0,08	4,9	8,4
<b>Total</b>	<b>11,4</b>	<b>6,8</b>	<b>0,54</b>	<b>0,53</b>	<b>0,06</b>	<b>0,08</b>	<b>4,5</b>	<b>7,2</b>

L1 : 1<sup>e</sup> lecture, L2 : 2<sup>e</sup> lecture, VPP : valeur prédictive positive

\* Seules les 1<sup>e</sup> et 2<sup>e</sup> lectures effectuées dans le cadre du programme sont comptabilisées.

La ventilation et la standardisation des échelles de mesure (en %) dans le tableau 4 permettent de comparer et d'apprécier séparément l'effet sur les performances du contexte de lecture, du type de lecteur et de l'expérience du radiologue. Notons tout d'abord que la fréquence des cancers non vus est statistiquement trop faible pour contribuer à expliquer des différences absolues dans la qualité des lectures.

En se concentrant sur les 2<sup>e</sup> lecteurs – les seuls à effectuer des 1<sup>e</sup> et des 2<sup>e</sup> lectures, on remarque que *le contexte de lecture influence grandement la performance*. Les 2<sup>e</sup> lecteurs obtiennent de meilleures performances en 2<sup>e</sup> plutôt qu'en 1<sup>e</sup> lecture : leur VPP est plus élevée (4,7% contre 7,2%) ce qui s'explique par un taux de faux positifs plus bas (10,3% contre 6,8%) pour des taux de détection sensiblement égaux (0,50% contre 0,53%).

Afin de prendre en compte l'influence du contexte de lecture sur les performances, la comparaison entre radiologues porte sur les résultats en 1<sup>e</sup> lecture uniquement. Les résultats suggèrent alors un niveau de performances proche entre 1<sup>e</sup> et 2<sup>e</sup> lecteurs (VPP de 4,5% et 4,7%). Les 2<sup>e</sup> lecteurs ont un taux de faux positifs légèrement plus bas (10,3% contre

12,4%), mais détectent environ 15% de cancers de moins que les 1<sup>e</sup> lecteurs (0,50% contre 0,58%).

Plus un radiologue est expérimenté, plus son niveau de performances s'élève (tableau 4). Ainsi, le taux de faux positifs décroît alors que le taux de détection et la VPP augmentent avec le nombre de lectures (hormis le taux de détection en 1<sup>e</sup> lecture, basé sur des petits nombres puisque les radiologues les plus expérimentés sont surtout sollicités pour interpréter les 2<sup>e</sup> et 3<sup>e</sup> lectures). En outre, l'effet du volume de lectures sur les performances est plus marqué en 2<sup>e</sup> lecture. En particulier, *on observe un niveau de performances amoindri en dessous de 3000 lectures*. Des analyses plus poussées confirment une rapide amélioration des principaux indicateurs de performance pour des volumes totaux de lectures inférieurs à 3'000 mammographies (résultats non présentés) ; au-delà, le gain résiduel de performance en fonction du nombre de lectures effectuées se réduit. L'effet de l'expérience du radiologue sur la qualité des lectures se retrouve, quoique de manière moins flagrante, si l'on considère le volume annuel moyen - plutôt que le volume total - des lectures (résultats non présentés).

### 3.3. Interprétation de l'analyse de la qualité des lectures

L'analyse de la qualité des lectures montre que le taux de faux positifs capture la majeure partie des différences globales de performance, notamment de VPP. Cet indicateur se révèle donc particulièrement utile pour évaluer la qualité des lectures radiologiques et cibler des interventions éventuelles. Outre le type de tour de dépistage (prévalent ou incident), qui influe considérablement sur les performances des radiologues mais sur lequel il est difficile d'agir, 2 facteurs s'avèrent déterminants dans la qualité de l'interprétation des mammographies : le contexte de lecture et le volume de lectures.

Les radiologues valaisans atteignent globalement de meilleures performances en 2<sup>e</sup> qu'en 1<sup>e</sup> lecture (section 3.2, tableau 4). Deux différences fondamentales caractérisent ces 2 contextes de lecture et peuvent expliquer ces performances différenciées :

- La 1<sup>e</sup> lecture s'effectue au cabinet tandis que la seconde est réalisée au CVD où les radiologues peuvent entièrement se concentrer sur l'interprétation des clichés avec un moindre risque d'interruption par d'autres sollicitations professionnelles.
- En 1<sup>e</sup> lecture, le radiologue tend à se placer dans une approche individuelle « médecin - patient » plutôt que dans une démarche de type purement populationnelle, qui caractérise

le contexte de la 2<sup>e</sup> lecture. Ce phénomène agirait surtout sur le taux de faux positifs des 2<sup>e</sup> lecteurs, ce que corroborent les résultats (10,3% contre 6,8%).

Il se peut de surcroît que la conséquence de l'interprétation radiologique soit perçue différemment selon le type de lecture réalisée (sentiment de « filet de sécurité » en 1<sup>e</sup> lecture). Par contre, il est peu probable que des différences d'équipements (*ie.* mammographe et négatoscope) entre le cabinet et le CVD contribuent à la différence de performances mesurées entre ces 2 contextes de lecture.

Les résultats attestent que la qualité des lectures augmente avec l'expérience du radiologue (le nombre cumulé de mammographies lues semble importer plus que le nombre annuel moyen de lectures). Le niveau de performances progresse rapidement jusqu'à 3'000 - voire 5'000 - lectures, seuil à partir duquel l'expérience acquise confère plus de constance dans la qualité des interprétations iconographiques. Cependant, le nombre moyen de lectures par radiologue reste comparativement faible en Valais<sup>11</sup>. La taille du bassin de population concerné et le nombre élevé de radiologues impliqués – signe par ailleurs de la bonne acceptabilité du programme – freinent la courbe d'apprentissage des radiologues.

L'analyse de la qualité des 1<sup>e</sup> lectures indique, en revanche, une faible différence de performances entre 1<sup>e</sup> et 2<sup>e</sup> lecteurs. Ceci reflète la stratégie du programme valaisan de ne pas opérer de distinction entre les types de lecteurs et confirme, de manière générale, que les radiologues volontaires pour être 2<sup>e</sup> lecteurs présentaient les mêmes compétences initiales que les autres radiologues. *Partant d'un niveau égal de capacité, les 2<sup>e</sup> lecteurs acquièrent plus rapidement de l'expérience de par le volume supérieur de mammographies qu'ils interprètent.* Ainsi, les 7 radiologues qui ont interprété moins de 1500 mammographies n'ont fait que des premières lectures alors que les 12 deuxième lecteurs ont, en moyenne, lu environ 3'000 mammographies depuis le début du programme (dont 7 qui ont interprété plus de 5000 mammographies, y compris les 5 radiologues 3<sup>e</sup> lecteurs).

Dans le contexte suisse où le critère de qualité concernant le volume annuel de lectures est difficilement applicable<sup>11</sup> (moins de 18'000 mammographies sont réalisées annuellement dans le plus grand programme cantonal de dépistage<sup>12</sup>), la sélection des 2<sup>e</sup> lecteurs est une stratégie possible, appliquée avec succès par plusieurs programmes romands (Genève, Vaud, Jura). Le recul et les données actuels sont insuffisants pour anticiper l'évolution des

---

<sup>11</sup> Les recommandations européennes préconisent un volume annuel d'au moins 5'000 lectures pour les radiologues 2<sup>e</sup> lecteurs.

<sup>12</sup> Programme vaudois de dépistage du cancer du sein : Rapport d'activité 2004.



différences de performance entre les 1<sup>e</sup> et 2<sup>e</sup> lecteurs en Valais. Celle-ci dépendra largement de l'évolution des performances individuelles et des choix d'accréditation et de répartition des types de lecteurs.

Afin d'estimer l'effet qu'aurait une sélection des 2<sup>e</sup> lecteurs, les taux de 3<sup>e</sup> lecture et de reconvoation du programme ont été calculés pour différents volumes de mammographies interprétées par les 2<sup>e</sup> lecteurs. En tour prévalent et à niveau d'expérience comparable des 1<sup>e</sup> lecteurs, les taux de 3<sup>e</sup> lecture se montaient à 22,4%, 19,7% et 16,6% selon que le 2<sup>e</sup> lecteur avait effectué moins de 3000 lectures, entre 3'000 et 5'000 lectures, ou plus de 5'000 lectures. Un gradient de même direction mais de moindre amplitude a été observé en tour incident (8,5%, 7,9% et 7,5%, respectivement). *Cette stratégie sélective (mais fictive) indique que l'utilisation de la plus grande spécificité de 2<sup>e</sup> lecteurs expérimentés permettrait d'amenuiser le taux de discordance entre radiologues.* En réduisant le taux de reconvoation (tableau 3), la 3<sup>e</sup> lecture permet aussi d'atténuer l'effet d'une sélection des 2<sup>e</sup> lecteurs. Ainsi, même s'il n'est pas négligeable, l'effet de l'expérience du 2<sup>e</sup> lecteur sur le taux de reconvoation des femmes est moindre (8,6% contre 7,9% pour des nombres de lecture inférieurs et supérieurs à 3'000 unités) que sur le taux de 3<sup>e</sup> lecture, et se manifeste surtout au-delà d'une expérience correspondant à 5'000 lectures (7,6%).

Les divers éléments statistiques présentés dans ce chapitre concordent et suggèrent que:

- 1) le taux de reconvoation est positivement (et fortement) corrélé au taux de 3<sup>e</sup> lecture
- 2) plus le 2<sup>e</sup> lecteur est expérimenté, plus le taux de 3<sup>e</sup> lecture diminue.

Même si un lien direct entre ces 2 constatations n'est pas formellement démontré, il est fort probable qu'une équipe restreinte de 2<sup>e</sup> lecteurs expérimentés permette d'abaisser le taux de reconvoation, à l'instar des pratiques et expériences faites dans plusieurs programmes romands de dépistage.

## 4. Conclusions

### 4.1. Limitations

Cette étude appréhende et quantifie les facteurs associés à la performance des lectures radiologiques dans le cadre du Programme valaisan de dépistage du cancer du sein. Elle fournit également des informations nécessaires pour apprécier l'effet d'interventions éventuelles visant à améliorer les performances du programme. Quelques réserves méritent toutefois d'être mentionnées :

- Ce travail original demande une **validation externe**. Elle pourrait se faire à 3 niveaux :
  - 1) La participation des radiologues valaisans à un test de lecture, tel que celui proposé aux radiologues vaudois<sup>13</sup>, permettrait une double validation, externe et objective.
  - 2) Une reproduction de cette analyse dans un autre programme romand est recommandable, notamment pour confirmer la relation entre les performances radiologiques et le volume de lecture. Ce travail scientifique est prévu par les évaluateurs en 2006.
  - 3) Une comparaison inter-cantonale de la qualité des clichés iconographiques, selon une échelle de score commune et reconnue, serait souhaitable pour s'assurer de la comparabilité du matériel avec lequel les radiologues sont évalués.
- Le **recul** actuel est **insuffisant** pour considérer les conclusions de ce travail comme définitives. Les petits effectifs sur lequel il s'appuie incitent notamment à la prudence.
- **L'interdépendance des facteurs** influençant la qualité des lectures ne peut être prise en compte que par une analyse statistique complexe (modèle multivariable). Ce travail scientifique sera réalisé par les évaluateurs au-delà du mandat d'évaluation.

### 4.2. Synthèse et recommandations

Reprenant les questions abordées dans cette étude (section 1.2), les éléments de réponse suivants peuvent être apportés :

---

<sup>13</sup> Un set de 200 mammographies dont la catamnèse est connue a été développé par l'UEC. Validé, puis utilisé par des radiologues vaudois, ce test de lecture est en cours d'adaptation en vue notamment de le proposer à d'autres programmes de dépistage.

### *Quelle a été l'évolution récente des indicateurs d'efficacité du programme ?*

Les indicateurs d'efficacité du programme suggèrent une légère progression depuis 2004, surtout au niveau du profil pronostique des cancers détectés en tour incident. Il est cependant probable que cette tendance favorable reflète, en large partie, des fluctuations aléatoires puisque ces indicateurs restent basés sur un petit nombre de cancers. Le taux de détection (tour incident), les proportions de cancers *in situ*, de petites tailles (tour prévalent) et sans envahissement des ganglions satisfont aux normes édictées par l'Union européenne. Ce n'est pas le cas des cancers dépistés à un stade avancé qui restent, proportionnellement, trop fréquents.

Les éléments disponibles ne permettent pas encore de cerner pourquoi le profil pronostique n'est pas meilleur en tour incident qu'en tour prévalent ou les raisons qui pourraient expliquer la forte proportion de tumeurs détectées à un stade avancé dans cette population. Une comparaison des paramètres cliniques des cancers dépistés par le Programme avec ceux des cancers d'intervalle et des cancers diagnostiqués en dehors du programme fournirait des renseignements utiles à cet égard. Ce travail est envisagé à court terme.

### *Pourquoi le taux de reconvoction est-il élevé en tour prévalent ?*

Une analyse chronologique (figure 1) a montré que le taux de reconvoction en tour prévalent a subi une forte élévation, au-delà des normes requises, dès 2002. Les causes principales de ce phénomène résident vraisemblablement dans l'accréditation de nombreux nouveaux lecteurs sur une brève période et la « réaction » des radiologues suite à la communication de leurs statistiques de lecture par le CVD. Ce dernier facteur n'est pas directement quantifiable mais une augmentation des mammographies positivées afin de réduire le risque de manquer un cancer a été observée dans plusieurs programmes suite au retour d'information sur les performances individuelles.

La baisse récente des taux de reconvoction (et de faux positifs) en tour prévalent, ainsi que la diminution de l'écart des performances entre radiologues, soulignent une progression dans la qualité des lectures. Ces signes encourageants ne se concrétisent cependant pas encore par une amélioration globale (1999-2005) du taux de reconvoction, à cause de l'ampleur des améliorations et de la maigre contribution des mammographies réalisées en tour prévalent (50% de toutes les mammographies prévalentes ont été réalisées dans les 2 premières années du programme). Le taux de reconvoction de 7,0% en tour prévalent pour

la période la plus récente (octobre 2004 – mai 2005) reste, au demeurant, élevé puisqu'il coïncide avec la limite supérieure recommandée au niveau européen (5 à 7%).

Bien qu'en phase de résorption (figure 3), un taux élevé de reconvocation peut avoir des conséquences à terme. Dans des contextes sanitaires comparables au Valais, plusieurs programmes de dépistage ont documenté une baisse sensible du taux de participation suite à un résultat faussement positif<sup>14</sup>. A l'échelle valaisanne, l'augmentation subite des taux de discordance et de reconvocation en 2002 a généré annuellement quelque 240 3<sup>e</sup> lectures et 155 reconvocations supplémentaires, avec un coût et des effets secondaires non négligeables (p.ex. anxiété).

Au demeurant, les causes probables de cette augmentation seraient sujettes à intervention: le nombre de nouveaux radiologues accrédités par année et une sensibilisation accrue des radiologues quant à l'effet de leur réaction suite à la connaissance de leurs statistiques de lecture. Bien qu'un faux négatif (cancer non vu) « pèse » plus qu'un faux positif pour un clinicien, l'analyse des déterminants de la qualité des lectures a démontré le poids globalement négligeable des cancers non vus dans la différence de qualité des performances.

*Quels facteurs inhérents au Programme influencent les performances des lectures ?*

D'importantes différences de performance entre radiologues ont été mises en évidence (tableau 4). Celles-ci relèvent surtout des conditions de lecture (les radiologues sont plus performants en tour incident et en 2<sup>e</sup> lecture) et de l'expérience du radiologue (mesurée par le volume total de mammographies lues dans le programme). Par ailleurs, le taux de faux positifs est l'indicateur qui discrimine le plus clairement les performances entre radiologues.

L'analyse des facteurs de qualité fournit les enseignements suivants :

- i. Une comparaison « non biaisée » des performances entre radiologues doit se concentrer sur les 1<sup>e</sup> lectures du programme ; comme pour tout indicateur de performance, une séparation entre tours prévalent et incident est essentielle.
- ii. L'apport de clichés comparatifs (tour incident) se traduit par une amélioration des performances, plus marquée chez les radiologues qui ont un faible volume de lectures (surtout des 1<sup>e</sup> lecteurs).

---

<sup>14</sup> Bulliard J-L, De Landtsheer J-P, Levi F. Reattendance in the Swiss mammography screening pilot programme. *J Med Screen* 2004, **11**, 59-64.

- iii. Sur le plan de la lecture, le taux de faux positifs contribue pour près de 95% au taux de mammographies positivées; cet indicateur qui reflète la spécificité radiologique permet une excellente surveillance de la qualité des lectures au niveau individuel.

*Quelles mesures permettraient d'agir sur les performances du Programme ?*

Ce travail atteste que le recours à des radiologues expérimentés améliore sensiblement la qualité du dépistage. A conditions de lecture égales, les radiologues qui ont interprété moins de 3000 mammographies positivent 1,5 à 2 fois plus souvent des anomalies qui se sont avérées non cancéreuses que leurs collègues qui ont cumulé plus de 5'000 lectures depuis le début du programme, ceci sans péjoration des taux de détection et de cancers non vus (tableau 4). A cet égard, la 2<sup>e</sup> lecture fournit une opportunité d'acquérir plus rapidement de l'expérience : les 2<sup>e</sup> lecteurs interprètent annuellement déjà environ 3 fois plus de mammographies que les 1<sup>e</sup> lecteurs. Cependant, le nombre de radiologues accrédités dans la population valaisanne concernée ne permet pas d'offrir assez de lectures pour que la plupart des radiologues puisse rapidement accroître leur expérience.

Plusieurs mesures de conscientisation des radiologues et de renforcement de la formation continue ont déjà été appliquées, concrétisant les recommandations émises (RE2004). L'évolution favorable en 2005 des indicateurs de qualité et d'efficacité du Programme valaisan de dépistage pourraient être les premiers fruits d'une expérience grandissante des radiologues et des efforts accrus de sensibilisation menés par la direction du Programme. Outre les colloques périodiques de mammographies et le retour annuel d'information sur les performances individuelles, le Programme a notamment, dès novembre 2004, informé les radiologues des possibilités de stages pratiques et de cours de formation (avec participation financière du programme), a fourni à chaque Service ou Institut de radiologie un cédérom d'entraînement et d'autoévaluation, et administré une enquête par questionnaire portant sur divers aspects du dépistage et du programme valaisan.

Si les mesures actuelles s'avéraient insuffisantes ou effectives à (trop) long terme, d'autres mesures éventuelles, même provisoires, seraient envisageables. Ces mesures complémentaires peuvent être de 2 ordres : globales ou individuelles. Des mesures globales au niveau de l'organisation du Programme, telles l'établissement d'un nombre minimum de mammographies à lire (par année, ou pour être admis comme 1<sup>e</sup> ou 2<sup>e</sup> lecteurs) ou une centralisation dans certains cas des 1<sup>e</sup> lectures au CVD (notamment lors d'une première participation) devraient impérativement être discutées et acceptées par l'ensemble des radiologues concernés. Une désolidarisation des radiologues pourrait se répercuter sur

d'autres plans qui constituent la force du Programme (excellente participation et acceptabilité du programme).

Les statistiques agrégées occultent parfois des variations importantes entre individus d'un même groupe. Si les indicateurs de performances suggèrent une tendance vers une plus grande homogénéité entre radiologues, quelques différences notables subsistent parmi les 1<sup>er</sup> ou les 2<sup>es</sup> lecteurs, ou pour des radiologues avec des volumes de lectures comparables. Une sensibilisation et une responsabilisation individuelles des radiologues se différenciant trop des performances moyennes des autres lecteurs serait susceptible d'homogénéiser la qualité des performances radiologiques. Cette approche devrait être privilégiée car elle implique et informe utilement les radiologues sans (grande) répercussion sur l'organisation des lectures.

# Glossaire

## **Tour de dépistage**

Période qui inclut l'ensemble des événements survenant lors d'une séquence d'invitation. D'une durée théorique de 2 ans, un tour débute avec l'invitation de la femme (ou tout autre contact éventuel) et se termine, au plus tard, avec sa prochaine invitation (ou tout autre événement qui justifie de ne plus l'inviter, tel qu'un cancer détecté ou une inéligibilité). Les dates de début et de fin de tour sont propres à chaque femme.

## **Tour prévalent/incident**

Un tour de dépistage est dit prévalent (ou initial) lorsque la femme participe pour la première fois au programme lors du tour en question. Si elle a déjà participé lors d'un tour précédent, le tour est dit incident (ou subséquent).

## **Taux de 3<sup>e</sup> lecture (ou taux de discordance)**

Proportion des mammographies pour lesquelles la conclusion des 1<sup>e</sup> et 2<sup>e</sup> lectures indépendantes diverge et nécessite une 3<sup>e</sup> lecture. Cet indicateur est sensible aux critères utilisés dans la grille de lecture radiologique et au traitement des divergences mineures.

## **Taux de reconvoction**

Proportion des femmes dépistées qui sont rappelées pour des investigations complémentaires à cause d'une anomalie suspecte (les clichés qui présentent un défaut technique sont généralement refaits sur place et ne sont pas comptés dans cette statistique).

## **Taux de résultats positifs (de la lecture)**

Proportion des mammographies jugées positives par un radiologue et nécessitant des examens complémentaires.

## **Taux de détection**

Proportion des femmes dépistées chez qui les investigations complémentaires ont confirmé le diagnostic de cancer (comprend les cancers détectés lors d'un examen de contrôle). Cet indicateur est un estimateur de la sensibilité du dépistage.

**Valeur prédictive positive (VPP) du test de dépistage**

Proportion des femmes rappelées chez qui les investigations complémentaires ont confirmé le diagnostic de cancer (comprend les cancers détectés lors d'un examen de contrôle).

**Taux de faux positifs**

Proportion des femmes rappelées chez qui les investigations complémentaires ont exclu un diagnostic de cancer. Cet indicateur est un estimateur de la spécificité du dépistage.

**Cancer d'intervalle**

Cancer du sein non détecté par le programme de dépistage chez une participante. Un cancer détecté lors d'un examen de contrôle (généralement à 12 mois) ou ne correspondant pas à la lésion radiologique signalée (par exemple, dans l'autre sein) n'est pas considéré comme un cancer d'intervalle du Programme. En pratique, seuls les cancers diagnostiqués dans un laps de temps correspondant à l'intervalle observé entre 2 mammographies sont comptabilisés, soit 2 ans à partir de la dernière mammographie effectuée dans le cadre du programme valaisan. La cause de non détection du cancer (erreur technique, cancer occulte à la mammographie, examens complémentaires inadéquats, etc...) n'est pas prise en compte : elle fait partie de la classification et non de la définition du cancer d'intervalle.

**Taux d'incidence en absence de dépistage**

Taux estimé de l'incidence du cancer du sein dans une population dépistée si aucun dépistage n'avait eu lieu. Cette valeur théorique permet d'apprécier les taux de détection et de cancer d'intervalle et de les comparer à des normes internationales. Ce taux est généralement estimé à partir du taux d'incidence pour une période précédant l'activité de dépistage (taux historique) ou sur la base d'une projection statistique (taux modélisé).