

## Synthèse de l'étude :

### Recensement des modèles de prédiction des risques.

**L'objectif de cette étude a été de réaliser l'inventaire détaillé des modèles de prédiction des risques de mortalité chez les personnes atteintes de maladies chroniques** sur une période d'au moins 5 années. Les pathologies concernées se sont limitées aux cancers (dont pédiatriques), aux maladies cardiovasculaires (infarctus du myocarde et accident vasculaire cérébral), et au VIH-SIDA. Cette étude s'est déroulée dans le cadre du marché lancé par la direction générale de la santé portant sur les « études permettant d'améliorer les pratiques assurantielles sur la base des évaluations scientifiques des risques chez les personnes atteintes de maladies chroniques ».

#### *Points méthodologiques*

La recherche documentaire a été menée à partir de l'interrogation d'un ensemble de sources d'information nationales et internationales spécialisées en économie de la santé, en médecine, en statistique et en épidémiologie. Les modèles de prédiction des risques de mortalité mentionnés dans les publications ont été analysés selon les principes de GRACE (*Good Research for Comparative Effectiveness*), ces derniers ayant été enrichis pour prendre en compte la disparité des études et la spécificité du sujet étudié. L'analyse des modèles a été menée pathologie par pathologie. Les cancers ont été étudiés en distinguant leurs différentes localisations/sites.

**Ainsi, 97 modèles de prédiction des risques de mortalité ont été identifiés et distribués par pathologie:**

- Infarctus du myocarde : 14
- Accident vasculaire cérébral : 10
- VIH-SIDA : 6
- Cancers : 67 dont
  - o 5 cancers multi sites,
  - o 11 cancers du sein et des organes génitaux féminins,
  - o 6 cancers des organes digestifs,
  - o 5 cancers des organes respiratoires,
  - o 7 cancers des tissus lymphoïdes, hématopoïétiques et apparentés,
  - o 19 cancers divers
  - o 14 cancers pédiatriques

Que ce soit pour les maladies cardiovasculaires, le VIH-Sida ou les cancers, une grande majorité de ces 97 modèles de prédiction des risques de mortalité s'appuie sur la méthode de Kaplan-Meier, le plus souvent couplée avec des modèles des risques proportionnels de Cox et/ou le test du log-rank.

De plus, en fonction des pathologies, certains modèles s'appuient sur d'autres méthodes, comme par exemple :

- Pour les maladies cardiovasculaires : les modèles de risques proportionnels de Cox avec test des tendances linéaires des estimations de risque (pour l'infarctus du myocarde), le calcul des SMR et de leurs intervalles de confiance suivant la distribution de Poisson (pour l'accident vasculaire cérébral),
- Pour le VIH-Sida : l'utilisation de modèle exponentiel pour estimer les taux de décès par tranche d'âge,
- Pour les cancers :
  - o la survie nette estimée par la méthode de Pohar-Perme,
  - o la technique de Ederer II,
  - o l'analyse de régression « compétitive » de Fine et Gray,
  - o un modèle linéaire généralisé avec une structure d'erreur de Poisson,
  - o un modèle paramétrique flexible de Lambert et Royston,

- des modèles de survie paramétriques flexibles, avec splines cubiques restrictifs pour les effets de risque cumulés et indépendants du temps,
- des intervalles de confiance calculés selon la méthode de Greenland,
- l'estimation des taux de survie relative suivant les standards de comparaison de population (méthode de Hakulinen et analyses réalisées avec SEER),
- des modèles flexibles paramétriques de survie permettant de modéliser simultanément deux "end points" (critère) avec méthodes statistiques pour risques compétitifs,
- l'analyse de périodes, modèles linéaires généralisés avec structure d'erreur binomiale
- l'estimation des SMR (ratio de mortalité standardisé),
- l'analyse de régression "compétitive" de Fine et Gray pour prédire la mortalité spécifique au cancer de la prostate.

En ce qui concerne l'infarctus du myocarde et l'accident vasculaire cérébral, **la variable expliquée** des modèles est principalement le taux de mortalité global (notamment à 5 ans, 10 ans et 15 ans). Pour le VIH-Sida, les variables expliquées des modèles sont principalement la mortalité globale et la mortalité par Sida. Pour les cancers, la variable expliquée des modèles est très diverse pour chaque site de cancer étudié : mortalité globale, mortalité relative, mortalité spécifique au cancer, excès de mortalité, survie nette, survie brute, survie conditionnelle, survie sans maladie, survie globale, survie relative, survie spécifique à la maladie (DSS), survie spécifique au cancer (CSS), et pour des périodes allant jusqu'à 20 ans.

Les **variables explicatives** des modèles sont diverses :

Pour l'infarctus du myocarde, les modèles peuvent prendre en compte jusqu'à 9 variables explicatives, dont 3 en moyenne, la plus fréquente étant l'âge.

Pour l'accident vasculaire cérébral, les modèles prennent en compte jusqu'à 8 variables explicatives, dont 5 en moyenne, les plus fréquentes étant l'âge, le sexe et le type d'AVC.

Pour le VIH-Sida, les modèles prennent en compte jusqu'à 6 variables explicatives, dont 4 en moyenne, les plus fréquentes étant l'âge, le sexe et le mode de transmission.

En ce qui concerne les cancers, les modèles peuvent intégrer jusqu'à 26 variables. Plus précisément :

- pour les cancers des organes digestifs, les modèles peuvent intégrer jusqu'à 9 variables explicatives, dont 5 en moyenne, les plus fréquentes étant l'âge, le sexe, l'ethnie, le stade tumoral, et le site du cancer.

- pour les cancers des organes respiratoires, les modèles intègrent de 4 à 6 variables explicatives, les plus fréquentes étant l'âge, le sexe, le stade, le score de "Performance status", et le stade TNM.

- pour les cancers du sein et des organes génitaux de la femme, les modèles intègrent jusqu'à 6 variables explicatives dont 3 en moyenne, les plus fréquentes étant l'âge au diagnostic, l'ethnie, la taille de la tumeur, et l'histologie.

- pour les cancers des tissus lymphoïdes, hématopoiétiques et apparentés, les modèles intègrent jusqu'à 5 variables explicatives, dont 3 en moyenne, les plus fréquentes étant l'âge, le sexe, et l'année de diagnostic.

- en ce qui concerne les cancers divers, les modèles intègrent jusqu'à 26 variables, dont 5 en moyenne, les plus fréquentes étant l'âge, l'histologie, le grade/score de Gleason, le stade SEER, le sexe, l'épaisseur de la tumeur, et le site.

- concernant les cancers pédiatriques, les modèles intègrent jusqu'à 6 variables explicatives, dont 2 en moyenne, les plus fréquentes étant l'âge, la période de diagnostic, le sexe, la dose de radiothérapie et d'anthracyclines, le type de chimiothérapie, et la radiothérapie.

**Les échantillons** les plus importants sont le plus souvent issus de données de registres et de cohortes rétrospectives. La taille de l'échantillon des études est variable selon les pathologies :

- pour les modèles s'intéressant à l'infarctus du myocarde et à l'accident vasculaire cérébral, la taille des échantillons est relativement faible (souvent inférieure à 1000 sujets),
- pour les modèles s'intéressant au VIH-Sida, la taille des échantillons est relativement importante (50% des échantillons supérieurs à 10000 sujets),
- en ce qui concerne les cancers, les tailles des échantillons des études sont très variables :
  - o pour les cancers multiples, la taille des échantillons est très importante (la plupart des échantillons étant supérieurs à 100000 sujets),
  - o pour les cancers des organes digestifs, la taille des échantillons est importante (80% des échantillons avec plus de 20000 sujets),
  - o pour les cancers des organes respiratoires, la taille des échantillons est moyenne (80% des échantillons avec moins de 6000 sujets),
  - o pour les cancers du sein et des organes génitaux féminins, la taille des échantillons est moyenne (70% des échantillons avec moins de 6000 sujets),
  - o pour les cancers des tissus lymphoïdes, hématopoïétiques et apparentés, la taille des échantillons est moyenne (70% des échantillons avec moins de 9000 sujets),
  - o pour les cancers divers, la taille des échantillons est relativement importante (30% des échantillons avec plus de 15000 sujets),
  - o pour les cancers pédiatriques, la taille des échantillons est moyenne (75% des échantillons avec moins de 7000 sujets).

Par ailleurs, concernant la durée médiane de suivi des études, celle-ci est en moyenne de 7 ans (hors cancers pédiatriques). Plus précisément, en ce qui concerne l'infarctus du myocarde, la durée médiane de suivi des études est en moyenne de 10 ans (avec des écarts allant de 5 à 20 ans en fonction des cas, cette durée étant parfois mal documentée). Pour l'accident vasculaire cérébral, la durée médiane de suivi des études est en moyenne de 8 ans (avec des écarts allant de 1 jour à 30 ans en fonction des cas). Pour le VIH-Sida, la durée médiane de suivi des études est de 5 ans (la durée n'étant pas toujours documentée). Pour les cancers, la durée médiane de suivi des études est en moyenne de 7 ans (allant de 5 ans pour les cancers multiples jusqu'à 10 ans pour les cancers divers). Concernant les cancers pédiatriques, la durée médiane de suivi des études est de 28 ans en moyenne.

Enfin, lorsque l'on s'intéresse à la **performance prédictive** des modèles appliqués, force est de constater que les informations disponibles sont très disparates et souvent peu documentées:

- Les seuils de signification et les intervalles de confiance sont fournis le plus souvent (sauf pour les modèles concernant l'infarctus du myocarde où les seuils de signification statistique des modèles ne sont pas toujours fournis ou de manière partielle) ;
- Le traitement des valeurs manquantes est très peu documenté ;
- Le test d'adéquation du modèle est en général peu documenté pour les modèles relatifs à l'infarctus du myocarde, à l'accident vasculaire cérébral et au VIH-Sida. Il l'est plus au niveau des modèles de prédiction des risques de mortalité liés aux cancers. Quelles que soient les maladies chroniques concernées, le lissage des données et l'analyse de sensibilité ne sont pas ou très peu réalisés dans les modèles.

En conclusion, concernant l'objectif de la présente étude, il faut noter que si les études de mortalité montrent toutes un effet de l'âge, peu fournissent des facteurs prédictifs pour des gens en âge de souscrire des assurances, ou des estimations de taux de survie par sexe et par tranche d'âge. Il est donc difficile à partir des articles identifiés et analysés de fournir directement des valeurs de survie par tranche d'âge et par sexe.

L'analyse des publications montre une très grande disparité dans la présentation des résultats. Le manque de standardisation des résumés rend difficile l'analyse approfondie. Afin de tirer tout le bénéfice des publications, il serait utile que les revues comme la présentation des auteurs puissent convenir d'un

corpus d'éléments devant figurer dans les articles. Ainsi les durées médianes de suivi et de survie, avec les minima et maxima de suivi, devraient être obligatoirement présentes dans la section des résultats. Les méthodes statistiques devraient être plus clairement et plus complètement exposées, afin de prouver la qualité, la puissance et la robustesse des résultats.

Par ailleurs, la création de registres pour les maladies cardiovasculaires pourrait permettre de mieux analyser les tendances et les effets des traitements. Pour les cancers, les effets des traitements sont insuffisamment pris en compte dans les analyses. La personnalisation des traitements (e.g pharmacogénomique) et leur évolution rapide, qu'il s'agisse des médicaments ou des dispositifs médicaux (e.g stents), risquent de rendre difficile, voire impossible, d'évaluer l'impact à moyen et long terme des traitements sur les taux de mortalité sur le long terme.

Enfin, à l'exception du cancer, les comorbidités sont très rarement prises en compte dans les analyses. Or, bien souvent la mortalité par la maladie concernée peut rester très inférieure aux autres causes de décès.

**Au total**, cette étude apporte des informations importantes pour la réflexion, mais insuffisamment précises pour répondre correctement à la problématique assurantielle. En particulier, au vu du corpus de données existant, il apparaît difficile de répondre à des questions précises comme par exemple « Quelle est la durée de survie moyenne d'une femme de 40 ans traitée par chirurgie et chimiothérapie adjuvante pour cancer du sein ? ». Globalement, dans les études relevées, il apparaît un manque d'homogénéité dans la restitution des résultats publiés par les auteurs. Par contre les résultats collectés par la présente recherche pourraient faire l'objet **de modélisation des risques** utiles pour les personnes atteintes d'une maladie cardiovasculaire ou d'un cancer.