

**hospital
innovation
economics**

CHAIRE
ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS
 ECOLE D'ÉCONOMIE DE PARIS
PARIS SCHOOL OF ECONOMICS

Compte rendu du 3ème Atelier technique Hospinnomics 28.05.2020 :

Ajustement au risque dans les modèles de financement et les mesures de performance

Cette synthèse reprend les présentations et les discussions de la troisième séance de l'atelier technique. Servant de support au document final, les remarques et suggestions de tous participants à cette séance viendront enrichir la qualité de cette première version.

Introduction

Grégoire Mercier, CHU Montpellier

Cet atelier est le troisième organisé par Hospinnomics en collaboration avec l'EHESP (Nicolas Sirven) et le CHU de Montpellier, il porte sur l'ajustement au risque et sa place dans le système de santé français.

L'ajustement au risque est un sujet certes d'une très grande technicité, mais qui pose des questions au cœur de nos systèmes de santé, notamment l'arbitrage entre efficacité et équité. Si la crise mondiale liée au SARS-Cov2 domine les réflexions actuelles sur les évolutions à court terme de nos systèmes de santé, les enjeux de long terme, dont les évolutions de nos modèles de financement, bénéficient toujours de la poursuite des expérimentations nationales.

Les réflexions sur les modèles de financement et en particulier sur l'ajustement au risque restent donc au cœur des préoccupations sur les transformations du système de santé et seront renforcées à l'avenir.

L'ajustement au risque en santé, quels usages potentiels en France ?

David Bernstein¹, Direction de la Sécurité Sociale

1. Ajustement au risque : définition et exemples internationaux

L'ajustement au risque représente l'utilisation des ressources de santé au regard des caractéristiques des individus. Il consiste donc en des techniques d'actuariat, se distinguant ainsi de la médecine prédictive.

Les modèles d'ajustement utilisent les attributs d'un individu ou d'une population pour calculer une dépense de santé ou la probabilité d'occurrence d'un événement de santé. Les événements probabilisés peuvent être divers comme une hospitalisation non programmée, une perte d'autonomie et ses conséquences comme l'installation en EHPAD ou le besoin d'aide à domicile. Les dépenses constituent un indicateur synthétique de l'utilisation des ressources.

La prédiction a plusieurs objectifs : elle peut être comparée à un observable si elle porte sur le temps présent et ainsi servir de repère, ou être utilisée pour la prédiction de l'utilisation des ressources futures. Cette prédiction relève de l'utilisation des services de santé et non de la médecine à proprement parler.

La majorité des modèles prédictifs développés aujourd'hui concernent les dépenses, les recours aux urgences non programmés et la perte d'autonomie. Les applications à l'international sont multiples :

- **Allocation des ressources financières** : définition des budgets alloués à des assureurs dans des systèmes de compétition régulée (Allemagne, Pays-Bas, Suisse, Afrique du Sud). Dans ces modèles il existe divers opérateurs d'assurance, publics ou parapublics, qui participent au service public de santé en étant subventionné par une entité tierce, comme une caisse de péréquation des risques. Au Pays-Bas, la caisse nationale d'assurance maladie reçoit les cotisations sociales et distribue ces sommes à 4 assureurs en concurrence parmi lesquels les assurés ont le choix. Pour répartir ce budget, les Néerlandais ont, depuis 15 ans une expertise sur l'utilisation des modèles d'ajustement au risque ;
- **Rémunération des prestataires de soins** : gestion locale de la santé publique (Royaume-Uni, Etats-Unis, Espagne), au Royaume-Uni, l'unité locale médicale, dirigée par des ordres de généralistes, reçoivent leur budget du NHS selon une formule définie par un modèle d'ajustement au risque développé en Angleterre ;
- **Amélioration des parcours de soins** : segmentation de la patientèle en vue de définir des parcours et services adaptés (Royaume-Unis, Etats-Unis, Espagne, Suède, Canada, Australie) ;
- **Etudes et évaluations** : pour assurer une comparaison juste entre les établissements notamment dans la mesure de la qualité des soins (Royaume-Unis, Etats-Unis, Allemagne).

2. Allouer des ressources financières

L'ajustement au risque permet d'allouer des budgets à des régions ou à des localités, d'allouer des ressources à des assureurs en compétition régulée ou encore à des prestataires de soins.

¹David Bernstein et Emin Agamaliyev « L'essor des modèles prédictifs dans les systèmes de santé internationaux » *Etudes & Résultats* (Drees) 2017

Depuis plusieurs années, émergent des tiers coordinateurs comme les *Accountable Care Organizations* (ACO) aux Etats-Unis, les *Clinical Commissioning Groups* en Angleterre et les Communautés Professionnelles Territoriales de santé (CPTS) en France. Ces institutions d'un niveau loco-régional - intermédiaire entre le local et le régional - ont pour fonction la gestion et l'adaptation fine de l'organisation des soins. La construction de parcours de soins ou de regroupement de professionnels de santé relève de leurs responsabilités et induit la nécessité d'un paiement forfaitaire groupé, ajusté au risque, par nécessité d'équité entre les fournisseurs de soins.

En France, les réformes du financement des soins conduisent à se doter de tels outils, en particulier dans le cadre de l'article 51² au travers des financements au parcours de soins dans les expérimentations nationales : paiement à l'épisodes de soins (EDS), paiement en équipe de professionnels de santé (PEPS) et incitation à une prise en charge partagée (IPEP). Par ailleurs, le rapport Aubert³ préconise une forfaitisation du financement de la dialyse et un nouveau modèle de financement pour la psychiatrie basée sur une dotation populationnelle.

Dans toutes ces évolutions l'ajustement au risque est nécessaire : le passage du paiement à l'acte au paiement forfaitaire individuel à l'épisode de soins ou à la capitation conduit à un transfert du risque du payeur vers le fournisseur de soin, d'où l'importance que ce transfert soit juste.

3. Expériences étrangères

a. Prédiction des budgets

Les modèles utilisés pour le financement de groupements de professionnels affichent de bonnes performances. À Venise, la capacité d'explication du modèle pour le financement des *Aera Locale Di Sanita*⁴ est importante lorsqu'en plus de l'âge et du sexe, est introduit la classification de John Hopkins (Baltimore)⁵ qui permet d'atteindre un R² (part de la variance des dépenses de santé expliquée) de 48%.

En Suède, dans le cadre d'une réforme de financement pour les soins primaires, l'introduction de cette même classification permet une meilleure explication des dépenses de santé pour atteindre plus de 60% d'explication de la variance⁶.

Au Massachusetts, une réforme de santé a été engagée visant à payer de manière globale les maisons de santé. L'université de Boston y a été mandatée pour construire un modèle⁷ pour le financement de ces maisons avec un résultat important : la variance des dépenses de santé expliquée par le modèle (R²) atteint 0,952.

² Article 51 de la Loi de Financement de la Sécurité Sociale de 2018

³ Jean-Marc Aubert, directeur de la Drees, « Stratégie de transformation du système de santé, rapport final : Modes de financement et de régulation », https://solidarites-sante.gouv.fr/IMG/pdf/rapport_financement_et_regulation.pdf

⁴ Niveau administrative infrarégional.

⁵ Maria Chiara Corti, Francesco Avossa, Elena Schievano, Pietro Gallina, Eliana Ferroni, Natalia Alba, Matilde Dotto, Cristina Basso, Silvia Tiozzo Netti, Ugo Fedeli, Domenico Mantoan, « A case-mix classification system for explaining healthcare costs using administrative data in Italy », *European Journal of Internal Medicine*, Volume 54, 2018, Pages 13-16, ISSN 0953-6205.

⁶ Zielinski, Andrzej, Maria Kronogard, Hakan Lenhoff, et Anders Halling. « Validation of ACG Case-Mix for Equitable Resource Allocation in Swedish Primary Health Care ». *BMC Public Health* 9 (2009). <https://doi.org/10.1186/1471-2458-9-347>.

⁷ Randall Ellis, Arlene S Ash, Juan Gabriel Fernandez, Verisk Health, "Good-Enough" Risk Adjustment Models for Physician Payment and Performance Assessment; University of Massachusetts Medical School 4 June 20, 2015

Les budgets alloués sont-ils basés sur un budget global ou sur des budgets nécessaires selon le risque observé ?

Les pays utilisant ces modes de financement recherchent plutôt l'équité territoriale que l'égalité de soins sur le territoire. Souvent, il existe une enveloppe au niveau macroéconomique, qui est votée. Les outils d'ajustement au risque servent à répartir équitablement les ressources définies par cette enveloppe.

b. Prédiction d'événements de santé

Les modèles d'ajustement au risque sont aussi utilisés pour stratifier le risque des patients par la prédiction d'événements futurs comme les réadmissions non programmées à l'hôpital⁸.

Dans une étude de 2011, dans le contexte d'une mise à l'agenda des problèmes de réadmissions aux Etats-Unis, divers professionnels d'un service hospitalier - internes, médecins de divers niveaux d'expérience et infirmières - sont interrogés afin d'identifier les patients les plus à risque de connaître une réadmission non programmée dans les 30 jours⁹. Cette étude met en évidence l'impossibilité pour les professionnels de santé de prédire le risque individuel d'événement particulier parmi les patients et consacre ainsi l'intérêt de l'ajustement au risque, en particulier dans une démarche de stratification prédictive.

En Angleterre, les capacités de prédiction des modèles sont utilisées pour le développement d'une application¹⁰ permettant l'identification des patients à hauts risques de réadmission non programmée. A partir des caractéristiques de l'établissement et du patient, elle définit une probabilité de réadmission permettant de cibler certains patients lorsque cela est nécessaire pour l'attribution d'une technologie rare, non disponible pour l'ensemble des patients. Les caractéristiques prises en compte sont celles de l'établissement et du patient. La probabilité de réadmission dans les 30 jours est ensuite affichée, ce qui permet de hiérarchiser le risque individuel et ainsi d'allouer certaine technologie comme de la surveillance postopératoire à domicile. Cela peut permettre aussi de cibler les visites à domicile. De la même manière, au Pays de Galle, un service auprès des médecins généralistes permet de recevoir une liste hiérarchisée de patients, classé par risque d'hospitalisation non programmée.

La qualité de ces modèles pour la prédiction de la perte d'autonomie est mise en évidence par une étude¹¹ de 2011 au Royaume-Uni qui identifie les individus qui risquent de perdre leur autonomie et de rentrer dans un EphaD ou d'être consommateur d'aides à domicile. L'étude montre que la population identifiée par le modèle comme ayant de plus hauts risques a dix-

⁸ Notons avec intérêt le développement en 2016 à l'AP-HP d'un modèle prédictif des admissions non programmées à des fins de gestion des ressources humaines (gardes et astreintes) pour l'hôpital. Cela consiste en un modèle d'identification individuelle des hauts risques dans une population/patiente. <https://www.lebigdata.fr/hopitaux-de-paris-big-data-1312>

⁹Allaudeen N, Schnipper JL, Orav EJ, Wachter RM, Vidyarthi AR. Inability of providers to predict unplanned readmissions. *J Gen Intern Med.* 2011 Jul;26(7):771-6. doi: 10.1007/s11606-011-1663-3. Epub 2011 Mar 12. PMID: 21399994; PMCID: PMC3138589.

¹⁰ Billings J, Blunt I, Steventon A et al. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open* 2012;2:pii: e001667 10.1136/bmjopen-2012-001667

¹¹Martin Bardsley, John Billings, Jennifer Dixon, Theo Georgiou, Geraint Hywel Lewis, Adam Steventon, Predicting who will use intensive social care: case finding tools based on linked health and social care data, *Age and Ageing*, Volume 40, Issue 2, March 2011, Pages 265–270, <https://doi.org/10.1093/ageing/afq181>

sept fois plus de chances de consommer des aides médico-sociale que la personne moyenne de 75 ans.

En Suède, dans les grands centres de santé municipaux, les infirmières qui font des visites à domicile utilisent un algorithme afin de classer les visites par ordre d'importance¹². En Australie, dans le cadre du *Health Care Home Program*, les individus éligibles au programme sont identifiés via un outil analysant les dossiers patients chez les praticiens : les médecins des centres de santé sont invités par *Medicare* à télécharger un programme dans le système de gestion des dossiers médicaux. Le programme lit les dossiers et sort la liste des patients éligibles à ce programme¹³.

Les modèles peuvent aussi être utilisés pour cartographier la population d'une région et ainsi favoriser les politiques ciblées. En Catalogne, les modèles de stratification populationnelle en cours de développement utilisent notamment une classification synthétique des co-morbidités « groupe de morbidité ajustée »¹⁴. Elle pourrait permettre de prioriser les actions et l'élaboration de stratégies personnalisées de confinement/déconfinement dans le cadre de l'épidémie de COVID 19. Plus largement, les régions de Venise et de la Catalogne ont catégorisé et fait le portrait des différents types de patients selon les symptômes, la complexité, les multimorbidités complexes, permettant d'obtenir un portrait de la population afin de prioriser les actions de santé.

Ces modèles permettent-ils d'améliorer la qualité de vie des patients ou bien de réduire les dépenses ?

Ces modèles permettent d'affiner les actions et donc d'améliorer la qualité de vie lorsque celle-ci est la finalité de l'action publique. La littérature traite notamment de l'identification des hauts risques. L'identification et l'enrôlement des patients les plus fragiles dans un dispositif d'hospitalisation à domicile préventive (« virtual wards ») s'appuient en Angleterre sur un modèle prédictif qui anticipe l'hospitalisation des plus fragiles¹⁵. Cependant, notons que l'algorithme n'est qu'un outil d'aide à la décision, cette dernière ainsi que la responsabilité afférente demeurant au sein des acteurs du système de santé. Les évaluations de ce dispositif sont positives. Il reste que très souvent c'est la qualité statistique du modèle qui est évaluée, mais rarement son impact en vie réelle. Les modèles en eux-mêmes n'ont pas de véritable impact en termes de qualité des soins ou de dépenses. Ils sont des outils qui peuvent donner lieu à un vaste panel d'action, pertinentes ou non. Dans certaines localités d'Angleterre, un algorithme a été utilisé pour identifier les personnes atteintes de la Covid-19 se présentant aux urgences, mais pouvant bénéficier d'une prise en

¹² Carlsson, L., U. Börjesson, et L. Edgren. « Patient Based "burden-of-Illness" in Swedish Primary Health Care. Applying the Johns Hopkins ACG Case-Mix System in a Retrospective Study of Electronic Patient Records ». *The International Journal of Health Planning and Management* 17, n° 3 (septembre 2002): 269-82. <https://doi.org/10.1002/hpm.674>.

¹³ <https://www1.health.gov.au/internet/main/publishing.nsf/Content/predictive-risk-model-algorithm>

¹⁴ Villar-Garcia, Judit, Rosa Maria Vivanco-Hidalgo, Montserrat Cleries, Elisenda Martinez, David Monterde, Pol Perez-Sust, Luis Garcia-Eroles, Carol Sais, Montserrat Moharra, et Emili Vela. « Risk Factors for SARS-CoV-2 Infection, Hospitalisation, and Death in Catalonia, Spain: A Population-Based Cross-Sectional Study ». *MedRxiv*, 1 septembre 2020, 2020.08.26.20182303. <https://doi.org/10.1101/2020.08.26.20182303>.

¹⁵ « Predictive Modeling in Action: How "Virtual Wards" Help High-Risk Patients Receive Hospital Care at Home | Commonwealth Fund ». <https://www.commonwealthfund.org/publications/issue-briefs/2010/aug/predictive-modeling-action-how-virtual-wards-help-high-risk>.

charge à domicile, au lieu d'une hospitalisation, permettant ainsi de consacrer les lits d'hospitalisation aux cas les plus risqués¹⁶.

4. Conclusion

L'ajustement au risque est la recherche d'un indicateur synthétique décrivant les besoins de services de santé d'une population ou d'un individu, pour l'allocation équitable des ressources. Cette synthèse peut constituer un précieux outil d'aide à la décision pour les décideurs, confrontés quotidiennement à des dilemmes d'équité.

La France est moins avancée dans ce domaine, du fait des caractéristiques de son système de santé qui ne rend pas nécessaire cet outil utile en raison du paiement à l'acte et au système de Sécurité Sociale monopolistique.

Cependant, les réformes récentes ont fait sentir le besoin de les développer davantage pour des applications multiples et à destination de nombreux usagers potentiels : administration et régulateurs, opérateurs de soins, communautés professionnelles territoriales de santé (CPTS).

¹⁶ Thornton, Jacqui. « The "Virtual Wards" Supporting Patients with Covid-19 in the Community ». *BMJ* 369 (5 juin 2020). <https://doi.org/10.1136/bmj.m2119>.

Evolution des indicateurs qualité dans les modèles de financement, place de l'ajustement au risque

Marc-Antoine Sanchez, ERUDITE, Hospinnomics, SSA

L'Agence Technique de l'Information sur l'Hospitalisation (ATIH) a été chargée de développer les modèles de financement des expérimentations de l'article 51 de la loi de financement de la sécurité sociale (LFSS) 2018. Une réflexion y a été menée pour comprendre l'évolution des indicateurs et de leurs rôles dans les modèles de financement eux-mêmes.

1. Evolution des indicateurs qualité

La mesure de la qualité des soins est une préoccupation majeure des politiques de santé. Elle est aujourd'hui prise en compte dans le financement du système de santé. De nombreux indicateurs se développent ainsi pour améliorer la mesure de la qualité des parcours. Les expérimentations sur ce sujet étant nombreuses à l'étranger et différents rapports d'experts proposent des adaptations possibles au cas français. C'est le cas du rapport d'Olivier Véran concernant l'évolution des modes de financement des établissements de santé, publié en 2016 qui proposait de s'inspirer d'expériences étrangères. Cela a d'ailleurs été fait pour le projet d'Incitation Financière à l'Amélioration de la Qualité (IFAQ), fortement emprunté au modèle Value Based Payment (VBP). L'idée principale est bien de décloisonner les modèles pour renforcer la place du financement de la qualité des soins. Des modèles mixtes se mettent ainsi en place, pour coupler le financement à l'activité et à la qualité.

Ces dernières années ont vu émerger la volonté d'installer le patient au cœur du système de soins. Les indicateurs ont récemment évolué pour prendre en compte les critères de jugement des patients, mesurée au travers des patient-reported outcome measures (PROMs) et des patient-reported experience measures (PREMs). L'international consortium for health outcomes measures (ICHOM) vise d'ailleurs à promouvoir la place des indicateurs de qualité de soins centrés sur le patient comme devant être au cœur des dispositifs de financement. Les études s'intéressant à la qualité ont montré par exemple que des expériences patient positives permettaient une diminution de la consommation des soins primaires et secondaires, une meilleure adhésion aux programmes de prévention et une meilleure compliance aux traitements^{17,18}. L'une des innovations des expérimentations consiste à considérer la qualité comme une part intégrante du financement des soins.

2. La co-construction des indicateurs dans le financement

La co-construction est née d'une volonté politique forte d'impliquer les professionnels de santé dans le processus de définition et de mise en œuvre des expérimentations article 51. Compte tenu des contraintes de temps pour l'élaboration des indicateurs qualité, la démarche dite de *bottom-up* a constitué un process innovant et efficace. Les indicateurs ont donc été définis à partir des propositions des cliniciens eux-mêmes. La co-construction des indicateurs avec les professionnels est une obligation à laquelle il faut aujourd'hui s'astreindre au titre de contraintes politiques¹⁹ ; elle permet notamment l'acceptation progressive de la mise en place des expérimentations par les professionnels, qui est un objectif important des réformes.

¹⁷ Doyle, Cathal, Laura Lennox, et Derek Bell. 2013. « A Systematic Review of Evidence on the Links between Patient Experience and Clinical Safety and Effectiveness ». *BMJ Open* 3 (1): e001570. <https://doi.org/10.1136/bmjopen-2012-001570>.

¹⁸ Anhang Price, Rebecca, Marc N. Elliott, Alan M. Zaslavsky, Ron D. Hays, William G. Lehrman, Lise Rybowski, Susan Edgman-Levitan, et Paul D. Cleary. 2014. « Examining the Role of Patient Experience Surveys in Measuring Health Care Quality ». *Medical Care Research and Review: MCRR* 71 (5): 522-54. <https://doi.org/10.1177/1077558714541480>.

¹⁹ Bras, Pierre-Louis. 2017. « Sortir de la T2A par le haut : la mesure de la qualité des soins ». *Journal de gestion et d'économie médicales* Vol. 35 (6): 245-63.

Les attentes sont grandes car ces modèles mettent explicitement la qualité au centre de la réflexion sur le système de soins. C'est une dimension importante pour les professionnels de la santé qui ont des motivations intrinsèques à améliorer la qualité des soins. La manière de mesurer la qualité des soins dans ces expérimentations repose essentiellement sur des indicateurs de résultats. Cela s'inscrit dans une tendance plus générale au passage d'indicateurs dédiés aux processus à des indicateurs de résultats et s'accompagne de la mise à dispositions des moyens nécessaires à la production de soins de qualité ; c'est un changement de paradigme important qui peut notamment conduire à valoriser les actes de prévention et ainsi trancher avec les modalités incitatives du paiement à l'acte.

3. Les incitations financières pour améliorer la qualité

Les incitations doivent être perçues comme suffisamment intéressantes par les établissements, pour les encourager à participer et à changer leurs pratiques. L'impact de l'introduction d'indicateurs qualité dans les modèles de financement en termes d'amélioration des pratiques est encore à prouver, les résultats d'IFAQ étant pour le moment mitigés²⁰. Plusieurs possibilités permettent d'introduire une part de financement à la qualité : un système de bonus/malus, et la définition d'un compartiment alloué à la qualité au sein même du modèle. Ces questions ont été essentielles dans la construction des modèles de financement à l'ATIH. Les expériences étrangères ont permis de se faire une idée des méthodes fonctionnelles.

Pour l'expérimentation Episode de Soins (EDS) de l'Article 51, le bonus est défini par la comparaison de deux mesures à deux cibles. La première mesure est celle de la dépense observée, mesurée à un "forfait cible", la seconde est celle du score qualité. Quatre cas de figure sont alors définis : 1) Dépense observée inférieure au forfait cible et score de qualité supérieur ou égal au seuil, cela entraîne l'obtention d'un bonus, 2) dépense observée inférieure au forfait cible mais score de qualité inférieur au seuil, cela n'entraîne pas de paiement pour la qualité, 3) dépenses supérieures au forfait et qualité supérieure ou égale au seuil, cela n'entraîne pas de malus ni de bonus, 4) dépense supérieure au forfait cible et score de qualité en dessous du seuil défini, cela entraîne un malus. Les mesures de la qualité et du forfait sont ajustées aux caractéristiques des patients.

L'ajustement au risque est nécessaire dans le calcul de ces indicateurs pour ne pas défavoriser les hôpitaux qui auraient des patientèles plus à risque et éviter la sélection des patients. L'ajustement prend en compte différentes caractéristiques patients, environnementales, ou même territoriales. Il doit se faire à l'entrée du parcours de soin surtout dans la mise en œuvre d'un paiement prospectif. Il constitue une des innovations majeures de l'article 51. Cet ajustement doit garantir l'adaptabilité des modèles, évoluant en même temps que les expérimentations. Ainsi les indicateurs utilisés dans la construction des modèles doivent aussi servir à l'évaluation.

4. Les incertitudes des expérimentations

Des incertitudes persistent concernant quatre points essentiels dans le déploiement de ce type de modèles :

- **Le recueil des données.** Le recueil des données reste une source d'incertitude forte. Un mode de financement fondé sur des données de qualité des soins doit s'assurer de la qualité de ces données mais aussi de l'uniformité du codage parmi les professionnels de santé ;

²⁰ Lalloué, Benoît, Shu Jiang, Anne Girault, Marie Ferrua, Philippe Loirat, et Etienne Minvielle. 2017. « Evaluation of the Effects of the French Pay-for-Performance Program-IFAQ Pilot Study ». *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care* 29 (6): 833-37. <https://doi.org/10.1093/intqhc/mzx111>.

- **L'impact sur la qualité des soins.** Les incitations financières sont nécessaires, mais elles ne sont pas suffisantes pour une amélioration de la qualité. L'accompagnement des acteurs est également indispensable ;
- Il faut aussi répondre aux **attentes des patients** : la transparence dans la communication des résultats permet d'assurer une visibilité nécessaire à la comparaison entre établissements. Ceci favorise la concurrence à la qualité entre établissements et le pilotage ;
- Plus largement il y a un **enjeu de diffusion** et de partage de l'information entre les professionnels de ville et d'hôpital. La transparence des indicateurs permet de mieux collaborer contribuant ainsi à améliorer les pratiques ;
- Enfin, la **lisibilité des modèles** est essentielle pour favoriser la diffusion et la compréhension par les professionnels.

5. Conclusion

En conclusion, la définition de ces indicateurs est passée d'une logique d'expert à une co-construction avec les professionnels en prenant en compte les contraintes économiques et pratiques. La méthode de co-construction semble particulièrement appropriée tant les enjeux sont multidimensionnels. Cette démarche s'inscrit dans l'application des recommandations faites il y a quelques années concernant la manière de construire les indicateurs qualité. Cependant, le coût à payer pour cette méthode est l'absence de validation par les instances compétentes, au risque de créer des indicateurs moins fiables ou moins performants, alors que les résultats des expérimentations françaises et étrangères sur le sujet s'accordent à dire qu'il s'agit là d'un écueil majeur des indicateurs actuels. La dernière étape est toutefois envisagée dans l'évolution de la construction de ces indicateurs afin de donner plus de place à la recherche.

Modèles prédictifs, comparaison avec des modèles traditionnels et différents outils de *machine learning*

Alexandre Vimont, AP-HP, URC ECO

1. Introduction

Le *machine learning* (ML) est un champ de l'intelligence artificielle. La méthode consiste en un processus d'entraînement du modèle permettant d'améliorer ses performances en exploitant les données disponibles de façon itérative. Le raisonnement est inductif : de l'observation à la généralisation. Le processus se divise en plusieurs étapes :

- i) Le choix d'une méthode d'apprentissage ;
- ii) Le choix des variables à utiliser ;
- iii) La paramétrisation du modèle et apprentissage ;
- iv) La validation du modèle.

Le modèle prédictif se différencie du cadre classique qui cherche à améliorer la compréhension des données. Dans les modèles de ML, l'apprentissage du modèle lui donne son caractère prédictif. Le modèle n'est donc pas nécessairement le plus juste, il est celui qui permet le mieux de prédire les nouvelles observations.

Dans le cadre des travaux sur l'ajustement au risque, l'apprentissage supervisé a été utilisé pour prédire les dépenses de santé, c'est-à-dire que les variables sont connues en amont. Dans le cadre du ML non supervisé, les variables ne sont pas connues.

2. Minimiser les erreurs

Le modèle d'apprentissage cherche à minimiser deux sources d'erreur : l'erreur sur les données d'apprentissage et l'erreur sur la validation des données. Plus la complexité du modèle est importante plus les erreurs d'entraînement diminuent. Toutefois les erreurs de validation forment un U avec le niveau de complexité : le surapprentissage contribue à faire augmenter les erreurs de validation. L'objectif est de trouver un compromis entre ces deux types d'erreurs.

3. Comparaison des modèles de ML et modèles classiques

Les modèles utilisés pour la prédiction des dépenses de santé sont un réseau de neurones et un modèle de forêt aléatoire. Dans le réseau de neurones, les variables explicatives sont entrées dans le modèle et des couches (neurones) sont introduites entre ces données et les données de sorties. Enfin, des fonctions d'activation viennent moduler l'entrée des variables utilisées pour la prédiction. Ces fonctions sont régies par des poids qui modulent l'information. Ces réseaux sont appréciés car ils permettent d'appréhender des relations non linéaires. Ils permettent également une convergence rapide des paramètres.

Dans le cas des forêts aléatoires, l'algorithme effectue un apprentissage sur de nombreux arbres de décision entraînés sur des sous-ensembles de données légèrement différents (rééchantillonnage). La structure hiérarchique des nœuds permet d'identifier et de sélectionner rapidement des interactions non linéaires entre variables.

Les modèles comparés dans le travail présenté ici sont un Modèle Linéaire Généralisé (MLG), une forêt aléatoire avec un maximum de 50 arbres et 50 nœuds et un réseau de neurones à deux couches cachées entraînés sur l'Echantillon Généraliste des Bénéficiaires (EGB) qui représente 1/97ème du SNDS. Les dépenses de santé sont définies par la part remboursé par l'Assurance Maladie, la part remboursée par les mutuelles, ainsi que le reste à charge des

patients. La prédiction des dépenses est faite en population générale à l'exception des individus avec troubles mentaux (séjours psychiatriques non inclus dans l'EGB). La variable à prédire est la dépense individuelle de santé, ville et hôpital confondus. Les variables utilisées pour la prédiction sont :

- Démographiques : sexe, âge, CMU-C, indice de défavorisation ;
- Cartographie des pathologies, indice de Charlson, ALD (présence, durée) ;
- Dépenses annuelles de l'année précédente ;
- Montant des postes de dépense de l'année précédente ;
- Nombre de consommation de soins de l'année précédente ;
- Nombre de délivrance par groupe ATC de l'année précédente ;

La performance des modèles est mesurée par le R², Mean Absolut Error (MAR), Hit Ratio (HiR), Matrice de confusion sur un échantillon de validation.

Transparence et explicabilité des résultats

Les modèles de ML sont souvent qualifiés de « boîtes noires » car il est compliqué d'expliquer les fondements sur lesquels se base la prédiction. Même si dans le cadre des forêts aléatoires le poids de chaque variable explicative est mesurable, l'interprétation reste toutefois limitée. La recherche actuelle en algorithmique, réalisée principalement en langage R ou python, via le portail d'accès à l'EGB a été très limitée jusqu'à aujourd'hui. Dans le cadre précédent du logiciel unique il a été fastidieux de proposer des modèles de prédiction innovants, mais la récente disponibilité du logiciel R et la multiplication des possibilités d'accès aux données pourrait la favoriser.

4. Statistiques descriptives

Les coûts sont essentiellement portés par les hospitalisations, plus de 50% des individus consomment moins de 1500 euros annuellement, 60% des individus consomment relativement le même montant d'une année à l'autre. Les individus associés à des coûts élevés (≥ 15.000€) qui représentent 2% de l'échantillon accumulaient 45% des coûts globaux.

5. Performance des modèles

Comme indiqué précédemment, les modèles de ML nécessitent des données riches et volumineuses pour être plus performants que les modèles linéaires classiques. Ainsi, sur les données démographiques tous les modèles se valent.

Variables	Indicateur de performance du modèle	GLM	Réseau de neurones	Forêt aléatoire
Démographiques	R ² ajusté	0,018	0,017	0,023
	MAE	2252	2294	2219
	HiR	38,94	32,70	44,75
Démographiques + Pathologies (CNAM)	R ² ajusté	0,137	0,120	0,258
	MAE	2073	2005	1864
	HiR	45,22	46,29	58,66
Démographiques + Pathologies (CNAM) + coûts totaux de 2015	R ² ajusté	0,350	0,320	0,346
	MAE	1787	1722	1788
	HiR	48,66	48,89	57,11
Démographiques + Pathologies (CNAM) + coûts totaux et coûts par type de service	R ² ajusté	0,341	0,206	0,376
	MAE	1661	1898	1601
	HiR	49,80	40,90	57,51

Démographiques + Pathologies (CNAM) + coûts totaux et consommation de ressources (quantité)	R ² ajusté	0,342	0,311	0,515
	MAE	1675	1670	1375
	HiR	47,60	47,73	62,04
Démographiques + Pathologies (CNAM) + coûts totaux et consommation de ressources (quantité) + nombre de délivrance (ATC)	R ² ajusté	0,345	0,175	0,488
	MAE	1681	1989	1478
	HiR	47,01	34,70	59,61

En introduisant la cartographie des pathologies de la CNAM, la forêt aléatoire devient plus performante. L'introduction des coûts de l'année N-1 (2015) introduit une ré-égalisation des modèles. Les modèles linéaires sont performants sur les données de type dépenses de santé ce qui explique la ré-égalisation de la performance des modèles qui fait suite à l'introduction des données des dépenses de santé de l'année précédente. L'introduction des postes de coûts ne change pas véritablement la performance des différents modèles. Les forêts aléatoires deviennent réellement plus performantes lorsque sont ajouté les consommations de ressources non pas en montant mais en quantité. A ce stade, les forêts aléatoires prédisent plus de la moitié de la variance alors que les autres modèles ont un R² inférieur à 0.35. Enfin, l'introduction du nombre de délivrance par la classification anatomique thérapeutique et chimique (ATC) n'apporte pas beaucoup plus de performance.

L'observation du croisement des classes prédites et des classes observées montre que le modèle GLM sur-prédit le montant des années précédentes. Le modèle GLM perd en performance en particulier dans les valeurs extrêmes : les dépenses supérieures à 5000 euros par ans. Cela est dû au fait que l'importance relative des différentes variables reste le même et que l'effet de ces variables est multiplicatif entre elles, alors qu'il est pondéré dans les modèles de ML.

Quels logiciels utiliser en priorité pour l'exploitation de tels modèles ?

Ce travail a été fait sur SAS avec des procédures déjà programmées. Il s'est avéré au cours des travaux de recherche que l'architecture des procédures préexistantes dans SAS est trop simple et ne permet pas complexifier le modèle au point de pouvoir prendre en compte les spécificités des dépenses de santé. De futurs travaux auront pour objectif d'utiliser les nouveaux logiciels mis en place sur la plateforme d'exploitation de l'EGB, notamment le langage R.

6. Conclusion

Le type de modèle à utiliser dépend de l'objectif qui est suivi. Si la recherche de la performance est première alors les forêts aléatoires sont plus attrayantes, car elles permettent de mieux identifier les individus atypiques. La performance d'un réseau de neurones, souvent conditionnée à son architecture, était trop limitée par son architecture. Des études ont montré des résultats prometteurs sur le SNDS avec des architectures plus complexes²¹. La performance des modèles GLM n'augmente pas avec la granularité et la quantité d'information contrairement aux forêts aléatoires.

Ce travail a également permis de faire émerger un axe potentiel de recherche dans la prévision des dépenses. Puisqu'une grande partie des coûts sont portés par une minorité d'individus et que ces coûts sont eux même portés par les hospitalisations, identifier les individus atypiques à travers un surrisque d'hospitalisation permettrait de mieux anticiper les dépenses.

²¹ Kabeshova A, Yu Y, Lukacs B, Bacry E, Gaïffas S. ZiMM: A deep learning model for long term and blurry relapses with non-clinical claims data. J Biomed Inform. 2020 Oct;110:103531.

