



Aging at work, multidomain determinants and the risk of dementia: supervised/unsupervised machine learning analyses in the Survey of Health Ageing and Retirement in Europe (SHARE)

AWDem, année 2018

Dr Laurent CLERET DE LANGAVANT (Laboratoire NPI, INSERM U955
Equipe 1, Ecole Normale Supérieure, UPEC)

8/12/2023

Soutenu par : IReSP & INSERM

Résumé

Vieillir au travail, déterminants multiples et risque de démence : apprentissage machine supervisé et non supervisé dans l'étude SHARE (Survey of Health Ageing and Retirement in Europe)

Dr Laurent CLERET DE LANGAVANT

CONTEXTE

La démence, ou trouble neurocognitif majeur, est caractérisée par un déclin cognitif progressif et une perte d'autonomie. Elle est due à la maladie d'Alzheimer et aux maladies apparentées, et reste largement sous-diagnostiquée (50% dans les pays développés). Puisqu'aucun traitement curatif des démences n'est disponible, la prévention revêt une importance cruciale. On estime que 40 % des démences pourraient être évitées grâce au contrôle des facteurs de risque modifiables. La prévention doit être proposée alors que l'individu est encore en activité professionnelle. Afin d'apporter des éléments scientifiques en faveur de la prévention, la cohorte SHARE paraît une source essentielle : c'est une étude Européenne dans laquelle des milliers d'individus de plus de 50 ans et tirés au sort sont évalués tous les deux ans depuis plus de 15 ans. L'étude SHARE permet d'évaluer la transition entre la fin de l'activité professionnelle et le passage à la retraite. Cependant, dans cette étude, le diagnostic de démence n'est pas établi.

OBJECTIFS

Nous voulons identifier les personnes atteintes de démence dans la cohorte SHARE en utilisant l'apprentissage machine non supervisé. Avec la suite de cette identification de la démence probable, nous avons décidé de créer un algorithme supervisé qui prédit le risque de démence probable à 10 ans en fonction des facteurs de risque modifiables.

MÉTHODOLOGIE

Nous avons développé un algorithme non-supervisé d'identification des cas de démence probable au sein de l'étude SHARE de façon longitudinale sur un échantillon de plus de 15000 personnes. Celui-ci utilisait seulement des variables de cognition (mémoire, calcul, langage) et d'autonomie. Dans cet algorithme non supervisé, les personnes de SHARE étaient automatiquement regroupées si elles avaient des caractéristiques similaires. Nous avons utilisé la cohorte English Longitudinal Study of Ageing (ELSA) pour la réplification. Nous avons ensuite créé un algorithme supervisé de prédiction de la démence basé sur l'identification des démences probables dans SHARE et en utilisant les facteurs de risque modifiables. Enfin, nous l'avons testé dans la cohorte ELSA.

PRINCIPAUX RÉSULTATS

L'algorithme non supervisé identifie les cas de démences probables à toutes les vagues de SHARE et de ELSA. Les personnes ayant une démence probable sont souvent âgées, de sexe féminin et ont souvent des facteurs de risque : faible niveau d'éducation, surdité, hypertension artérielle, consommation excessive d'alcool, tabagisme, dépression, isolement social, inactivité physique, diabète et obésité. Grâce à l'algorithme de prédiction, il est

possible d'estimer le risque d'avoir une démence probable à 10 ans en fonction des facteurs de risque déclaré par une personne.

APPORTS OU IMPACTS POTENTIELS

Nous avons développé un algorithme de prédiction des démences à 10 ans qui identifie les personnes à risque de plus de 50 ans alors qu'elles sont encore en activité professionnelle. Des stratégies de prévention multi-domaines des démences pourraient leur être proposées.

Abstract

Aging at work, multidomain determinants and the risk of dementia: supervised/unsupervised machine learning analyses in the (SHARE)

Dr Laurent CLERET DE LANGAVANT

CONTEXT

Dementia, or major neurocognitive disorder, is characterized by progressive cognitive decline and loss of independence. It is mainly due to Alzheimer's disease and related diseases, and remains largely underdiagnosed (50% in developed countries). Since there is currently no cure for dementia, prevention is of crucial importance. It is estimated that 40% of dementia cases could be prevented by controlling modifiable risk factors for dementia. Prevention must be offered while the individual is still working. In order to provide scientific evidence in favor of prevention, the SHARE cohort appears to be an essential source: it is a European study in which thousands of randomly selected individuals over 50 years of age have been evaluated every two years for more than 15 years. The SHARE study assesses the transition from the end of employment to retirement. However, in this study, the diagnosis of dementia is not established.

OBJECTIVES

We wanted to identify people with dementia in a large cohort in the European General Population (SHARE) using an original method, unsupervised machine learning. Once people with probable dementia were identified in the SHARE cohort, we decided to create a supervised algorithm that predicts the risk of probable dementia at 10 years based on dementia risk factors.

METHODS

We developed an unsupervised algorithm to identify cases of probable dementia in the SHARE study longitudinally on a sample of more than 15000 people. It used only cognition variables (memory, numeracy, language) and autonomy. In this unsupervised algorithm, people in SHARE were automatically grouped together if they had similar characteristics. We used the English Longitudinal Study of Ageing (ELSA) cohort for replication. We then created a supervised dementia prediction algorithm based on the identification of probable dementias in SHARE and using modifiable risk factors. Finally, we tested it in the ELSA cohort.

MAIN RESULTS

The unsupervised algorithm identifies probable dementia cases at all waves of SHARE and ELSA. People with probable dementia are often elderly, female, and often have risk factors: low education, deafness, high blood pressure, excessive alcohol consumption, smoking, depression, social isolation, physical inactivity, diabetes, and obesity. Thanks to the prediction algorithm, it is possible to estimate the risk of having probable dementia at 10 years based on the risk factors reported by a person.

OUTCOMES OR POTENTIAL IMPACTS

We have developed an algorithm for predicting dementia at 10 years of age that identifies at-risk people over the age of 50 while they are still working. Multi-domain dementia prevention strategies could be proposed.

Synthèse longue

Messages clés du projet

- L'apprentissage machine non supervisé appliqué aux données longitudinales de la cohorte en population générale SHARE permet d'identifier les participants ayant une démence probable.
- L'apprentissage machine supervisé permet de prédire le risque futur de démence probable en utilisant des données élémentaires comme les facteurs de risque de démence.
- A l'issue du projet, un prototype d'algorithme qui prédit le risque de démence futur permet d'identifier les personnes qui pourraient le plus bénéficier d'un programme de prévention intensif et multi-domaines des démences.

Contexte

Généralités sur les démences

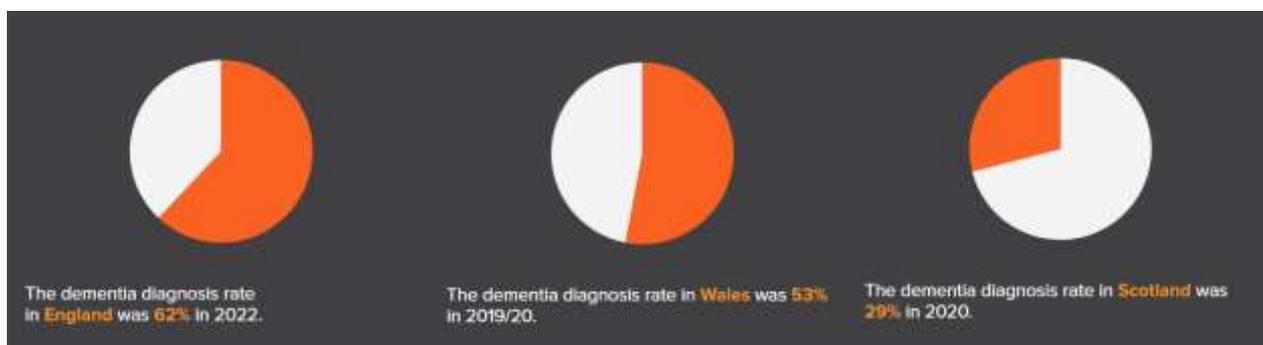
Depuis 2013, le terme « démence » est renommé « trouble neurocognitif majeur » dans la 5ème édition du DSM (DSM-V) avec de nouveaux critères diagnostiques : il s'agit d'une condition responsable d'un déclin cognitif significatif par rapport à un niveau antérieur et d'une perte d'autonomie dans les actes du quotidien. Cependant, le terme employé le plus fréquemment au sein de la communauté médicale et scientifique reste « démence ».

La majorité des cas de démence est due à un ensemble de pathologies nommé « maladie d'Alzheimer et maladies apparentées » (MAMA). La maladie d'Alzheimer représente à elle seule 60 à 70% des cas de démences. Les MAMA touchent particulièrement les personnes âgées.

Au niveau mondial, selon les dernières estimations de l'OMS de 2019, la démence toucherait près de 55 millions d'individus. Selon les prévisions, la prévalence risque de presque tripler d'ici 2050 pour atteindre 152 millions d'individus. La démence est la première cause de perte d'autonomie chez la personne âgée et la 7ème cause de mortalité tous âges confondus. Selon les chiffres de 2016 de Santé Publique France, près de 1,2 million de personnes sont touchées par les MAMA en France. En 2050, on estime que 2,2 millions de personnes seront atteintes par les MAMA en France.

Au niveau mondial, la démence est largement sous-diagnostiquée. Une méta-analyse par Lang *et al.* de 2017 incluant 23 études internationales montrée que le pourcentage de démence non détectée était de 61,7% [55% – 68%]. Les réticences personnelles et familiales, la prise en charge défaillante par le système de soins ou l'accès aux structures adaptées avec du personnel spécialisé sont autant de raisons qui sont à l'origine de ces chiffres.

Figure 1 : D'après Alzheimer's research UK, la démence est largement sous-diagnostiquée en Angleterre (62% de taux de diagnostic), au Pays de Galles (53%) et en Ecosse (29%).



Il n'existe pas à ce jour de traitement curatif des démences. Les traitements disponibles pour la maladie d'Alzheimer comme les anticholinestérasiques n'ont qu'un effet symptomatique. De plus, les nouveaux traitements à base d'anticorps monoclonaux qui ciblent les plaques amyloïdes de la maladie d'Alzheimer ne sont disponibles que dans un nombre limité de pays et ne sont pas considérés comme curatifs de cette maladie. En France, la prise en charge des patients déments repose sur un accompagnement non-médicamenteux et des mesures médico-sociales pour soulager le patient et les aidants.

En l'absence de traitement curatif, la prévention est la stratégie de choix afin de réduire le nombre de personnes atteintes de démence ou, a minima, de retarder la survenue du déclin cognitif. Les démences sont favorisées par des facteurs non modifiables comme le grand âge, le sexe féminin et certains déterminants génétiques comme l'allèle 4 de l'ApoE. Cependant, au moins 12 facteurs de risque modifiables de démence ont été identifiés par la communauté scientifique : le faible niveau d'éducation, la perte auditive, le traumatisme crânien, l'hypertension artérielle, la consommation excessive d'alcool, l'obésité, le tabagisme, la dépression, l'isolement social, l'inactivité physique, la pollution de l'air et le diabète. On estime que près de 40% des cas de démence seraient évitables grâce au contrôle de ces facteurs de risque modifiables de démence. Cette estimation est confortée par les résultats de l'étude FINGER qui a montré que le contrôle de plusieurs facteurs de risque modifiables en même temps chez les sujets à fort risque de démence était associé à une baisse du risque de déclin cognitif.

Cohortes de vieillissement en population générale

1. Survey of Health, Ageing and Retirement in Europe (SHARE)

Lancée en 2004, l'étude SHARE étudie le vieillissement et ses conséquences chez les personnes âgées de 50 ans et plus. L'étude SHARE a d'abord été mise en place dans 12 pays (Allemagne, Autriche, Belgique, Danemark, Espagne, France, Grèce, Israël, Italie, Pays-Bas, Suède et Suisse). Puis avec le temps, elle a été déclinée dans plus de 25 pays Européens.

Figure 2 : date des différentes vagues de l'étude SHARE selon g2aging.org



Dans chaque pays participant, un échantillon représentatif de la population générale de plusieurs milliers de personnes est sélectionné. Ces personnes répondent à des questions et

tests lors d'interviews tous les deux ans environ. Ainsi, dans SHARE sont récoltées des informations socio-démographiques (âge, sexe, profession, études), des données cognitives (performances à des tests), des données sur leur autonomie (capacité de marcher, faire les courses, gérer son budget), des données sur les pathologies et les comorbidités (diabète, hypertension, dépression), sur le réseau social (nombre d'interactions sociales), ainsi que sur les revenus et finances. Quand la personne sélectionnée ne parvient pas à répondre à une question, un membre proche du foyer est autorisé à répondre à sa place.

L'étude SHARE constitue une opportunité exceptionnelle pour mieux comprendre les déterminants du vieillissement pathologique et notamment des démences. Cependant, le diagnostic de démence ou de maladie d'Alzheimer n'y est pas collecté de façon très fiable. En effet, la maladie d'Alzheimer ou la démence est un diagnostic auto-rapporté par les personnes interviewées (participant ou proxy) grâce à une seule question depuis la deuxième vague : « Est-ce qu'un médecin vous a déjà dit que vous aviez une maladie d'Alzheimer, une démence, une sénilité ? ». La réponse à cette question apporte une information sur une éventuelle MAMA, mais cela dépend de la volonté du participant de répondre à cette question délicate et surtout reste soumis au problème de sous-diagnostic des démences (62% en moyen comme décrit au-dessus). Ainsi, le nombre de cas de MAMA rapporté dans SHARE est probablement très inférieur à la réalité. C'est une limite majeure pour étudier la démence dans cette cohorte.

2. English Longitudinal Study of Ageing (ELSA)

L'étude anglaise ELSA s'intéresse au vieillissement et ses conséquences chez les personnes de 50 ans et plus en Angleterre. L'étude ELSA suit un protocole similaire à celui de SHARE.

Figure 3 : Dates des différentes vagues de la cohorte ELSA selon g2aging.org



Les participants d'ELSA ont été interviewés par des enquêteurs en face à face. Il est à noter que le recueil des informations lors du décès d'un participant entre deux phases n'était pas aussi exhaustif que dans SHARE.

Contrairement à SHARE, la maladie d'Alzheimer et la démence font l'objet de deux questions séparées auprès du participant ou de son proxy : « Votre médecin vous a-t-il déjà dit que vous aviez une maladie d'Alzheimer ? » « Votre médecin vous a-t-il déjà dit que vous aviez une démence ? » Cependant, encore une fois, le sous-diagnostic des MAMA a pour conséquence une sous-estimation des cas réels de MAMA dans l'étude ELSA.

Objectifs

Le premier objectif de ce projet était d'identifier parmi les participants inclus dans les cohortes longitudinales SHARE et ELSA ceux qui étaient probablement atteints de démence et qui étaient sans doute beaucoup plus nombreux que les personnes ayant déclarés avoir reçu ce diagnostic par un médecin.

Le second objectif était la création d'un algorithme de prédiction de la démence probable identifiée grâce au premier objectif en utilisant les facteurs de risque modifiables de démence. Un tel algorithme permettrait d'identifier les personnes les plus à risque et de leur proposer des stratégies de prévention multi-domaines à l'instar de ce qui a été testé comme efficace dans l'étude FINGER.

Identification de la démence

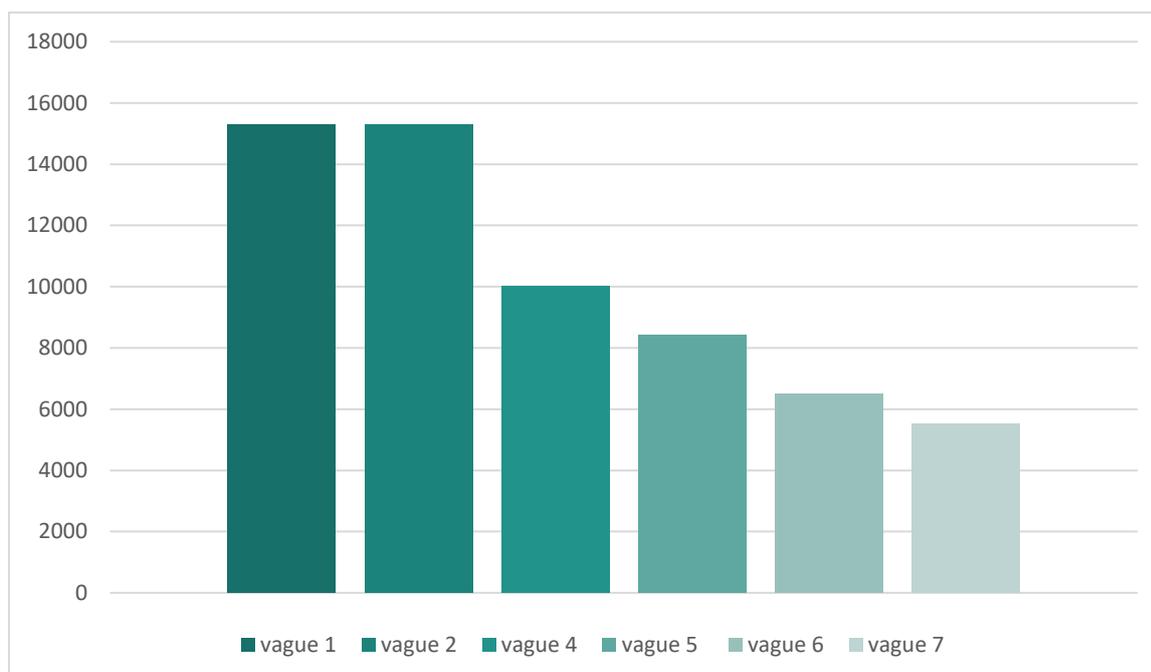
Méthodologie

Cleret de Langavant et al. (2018) ont montré l'intérêt de l'utilisation d'un algorithme d'apprentissage statistique non-supervisé au sein de l'étude SHARE pour identifier un sous-groupe de participants avec une forte probabilité de démence. Cette étude utilisait des données transversales, avec de très nombreuses variables incluant les facteurs de risque de démences. L'algorithme combinait une Analyse en Composantes Principales (ACP) et une Classification Hiérarchique sur Composantes Principales (CHCP) pour identifier 3 clusters, dont l'un correspondait à un statut de démence probable. Même si cette approche était novatrice, elle avait également des limites car elle empêchait la construction d'algorithmes de prédiction des démences fondé sur l'existence de facteurs de risque. Il était donc nécessaire d'ajuster cette méthode d'identification dans l'objectif de créer un algorithme de prédiction.

1. Population d'étude

Nous avons utilisé les données des vagues 1 à 7 de SHARE allant de 2004 à 2017. Nous avons sélectionné les participants âgés de 50 ans présents à la première vague avec un suivi consécutif. Les participants venaient de 12 pays : Allemagne, Autriche, Belgique, Danemark, Espagne, France, Grèce, Israël, Italie, Pays-Bas, Suède et Suisse. La population d'étude comprenait 15278 individus ayant au moins participé aux deux premières vagues. Du fait du phénomène d'attrition, très important dans SHARE, et des décès l'échantillon a diminué au fil du temps ($N_{\text{SHARE_W4}} = 10008$; $N_{\text{SHARE_W5}} = 8418$; $N_{\text{SHARE_W6}} = 6485$; $N_{\text{SHARE_W7}} = 5533$).

Figure 4 : Attrition générale dans la cohorte SHARE



2. Clustering

Afin d'être le plus proche de la définition du DSM-5 de la démence (trouble cognitif et perte d'autonomie), nous avons sélectionné les variables renseignant sur les performances cognitives et sur le niveau d'autonomie. Les variables ayant plus de 30% de données manquantes ont été exclues. Celles restantes ont été imputées, si besoin, en utilisant la commande `imputeMFA` du package `missMDA`.

Nous avons utilisé une Analyse Factorielle Multiple (AFM) via le package `FactoMineR`. L'AFM a été réalisée avec les données de l'ensemble des vagues. L'AFM a été suivie d'une classification hiérarchique sur composantes principales CHCP en 3 clusters utilisant les 5 premières composantes principales créées. Chaque participant, à chaque vague, a été assigné à un cluster (1, 2 ou 3) par l'algorithme. Quand un participant était classé dans le cluster de démence probable, il était attribué à celui-ci de manière permanente.

3. Validation de l'identification de la démence probable

Afin de valider le cluster regroupant les participants probablement atteints de démence, nous avons :

- I. Comparé les résultats obtenus avec la variable de démence auto-rapportée ;
- II. Étudié la prévalence de l'appartenance à ce cluster par sexe et classe d'âge ;
- III. Vérifié l'association de l'appartenance à ce cluster avec des facteurs de risque modifiables de démence grâce à des modèles multi-états.

4. Prise en compte de l'attrition

Le phénomène d'attrition est très important au sein de SHARE. Nous avons utilisé la méthode de pondération de probabilité inverse à l'aide du package `ipw`. Pour chaque vague, une régression logistique a été calculée en utilisant les caractéristiques des participants (âge, sexe et pays d'origine) à la vague précédente. Les poids statistiques ont été obtenus en

inversant la probabilité de participation calculée par le modèle. Ces poids ont été introduits au moment de l'imputation et du clustering.

5. Réplication

L'ensemble des analyses précédentes a été répliqué avec les données de la cohorte ELSA de la vague 1 à la vague 9 avec deux exceptions. L'information sur le lieu de résidence (urbain ou rural) des participants n'était pas disponible, donc il n'a pas été possible d'étudier l'effet de la pollution comme facteur de risque. ELSA n'ayant de recueil exhaustif des décès, il était impossible de les prendre en compte dans les modèles multi-états.

Résultats

1. Résultats dans l'étude SHARE

a. Identification du cluster de démence probable

A l'inclusion, la répartition des participants entre les 3 clusters est inégale (Tableau 1). Le Cluster 1 (N = 11369) et le Cluster 2 (N = 3374) regroupent la majorité des participants de SHARE tandis que le Cluster 3 (N = 535) est le plus petit. Ce dernier concentre des individus plus âgés (âge moyen 76,5), le plus souvent des femmes (N = 368 [68,8%]) et moins éduqués (N = 436 [80,1%] avec un niveau d'éducation en deçà du secondaire). Par rapport aux personnes des Clusters 1 et 2, ils ont une perte d'autonomie importante (score ADL moyen de 3,1 [1,7] et score IADL moyen de 4,2 [1,8]) et des performances cognitives moindres (nombre de mots en rappel immédiat de 2,6 [1,9] et fluence verbale de 10,4 [6]). Le Cluster 3 correspond donc au cluster de « démence probable ».

Tableau 1 : Caractéristiques des participants de l'étude SHARE suivant les résultats de la répartition du clustering à la vague 1.

	Cluster 1 (n=11369)	Cluster 2 (n=3374)	Cluster 3 (n=535)
Age (années)	64,6 (9,6)	65,2 (9,7)	76,5 (11)
Sexe			
Femmes	6,251 (55%)	1704 (50,5%)	368 (68,8%)
Hommes	5,118 (45%)	1670 (49,5%)	167 (31,2%)
Education			
En deçà du secondaire	6863 (51%)	593 (45,8%)	436 (80,1%)
Education secondaire	3973 (29,6%)	402 (31,1%)	68 (12,5%)
Education supérieure	2604 (19,4%)	299 (23,1%)	40 (7,4%)
Score mobilité [0–7]	1 (1,4)	1 (1,4)	4,9 (1,5)
Autonomie			
Score ADL [0–6]	0,1 (0,3)	0,1 (0,4)	3,1 (1,7)
Score IADL [0–7]	0,2 (0,5)	0,2 (0,6)	4,2 (1,8)
Cognition			
Rappel immédiat mots [0–10]	5 (1,7)	4,7 (1,9)	2,6 (1,9)
Fluence verbale [0–67]	19,5 (7,1)	18,7 (7,3)	10,4 (6)

b. Comparaison à la démence auto-rapportée

A chaque vague, le nombre de personnes identifiées comme probablement démentes après classification hiérarchique (N = 1013 [6.6%] à la vague 2) est supérieur au nombre de personnes identifiées par la variable de démence auto-rapportées (N = 278 [1,8%] à la vague 2).

Au fil des vagues, l'AUC varie entre 0,745 (0,712 – 0,777) et 0,830 (0,800 – 0,861), montrant de bonnes capacités de discrimination de l'identification de la démence probable par notre méthode comparée à la démence auto-rapportée. La sensibilité, quant à elle, varie de 0,579 (0,520 – 0,637) à 0,702 (0,646 – 0,759), alors que la spécificité est toujours au-dessus de 90%.

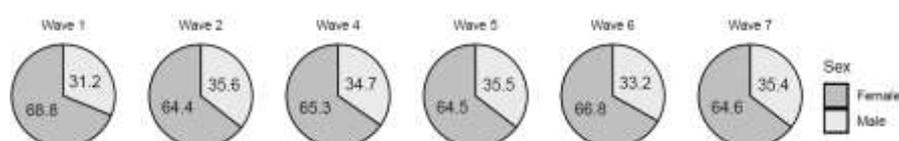
Tableau 2 : Comparaison de l'identification de la démence probable et de la démence auto-rapportée.

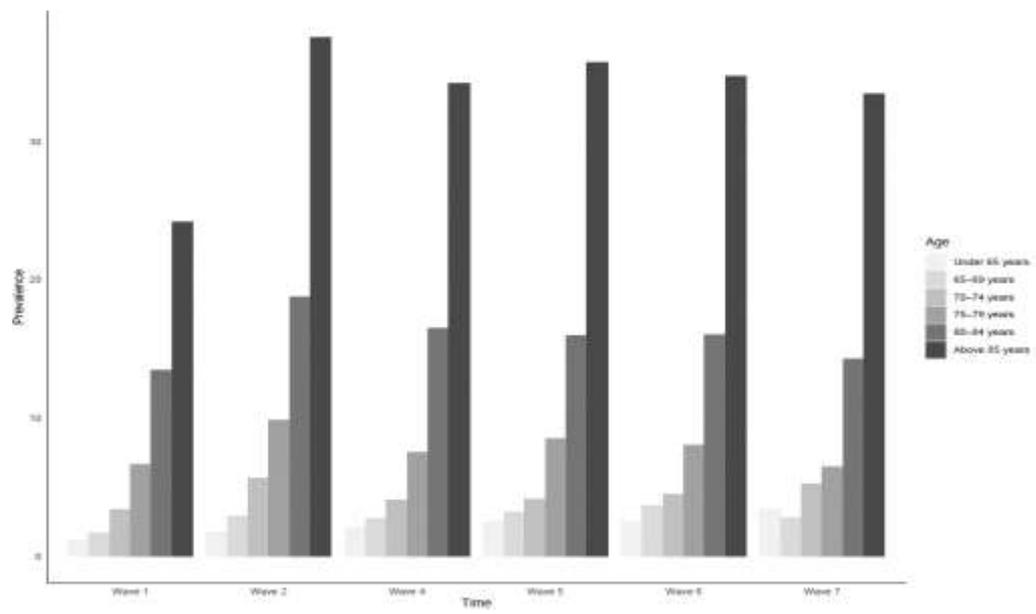
Vagues	Démence auto-rapportée	Démence probable	Sensibilité	Spécificité	Aire sous la courbe (AUC)
Vague 1 (2004)	NA	535 (3,5%)	NA	NA	NA
Vague 2 (2006)	278 (1,8%)	1013 (6,6%)	66% (61 – 72)	94% (94 – 95)	80% (78 – 83)
Vague 4 (2010)	252 (2,5%)	691 (7%)	70% (65 – 76)	95% (94 – 95)	82% (79 – 85)
Vague 5 (2012)	279 (3,3%)	677 (8%)	65% (59 – 70)	94% (93 – 94)	79% (76 – 82)
Vague 6 (2014)	273 (4,2%)	585 (9%)	58% (52 – 64)	93% (92 – 94)	75% (72 – 79)
Vague 7 (2016)	276 (5%)	514 (9,3%)	56% (50 – 61)	93% (92 – 94)	74% (71 – 78)

c. Prévalence par sexe et par âge

A chaque vague, les participants probablement déments selon notre méthode sont environ pour deux tiers des femmes et pour un tiers des hommes, comme il est classiquement rapporté pour la démence selon la littérature. Le nombre de cas de participants probablement déments suit un quasi-doublement par tranche d'âge de 5 ans pour chaque vague (Figure 1).

Figure 5 : Répartition des participants identifiés comme probablement déments dans SHARE. Représentation par sexe et par tranche d'âge de 5 ans de la vague 1 à la vague 7. Les nombres rapportés sont des pourcentages.

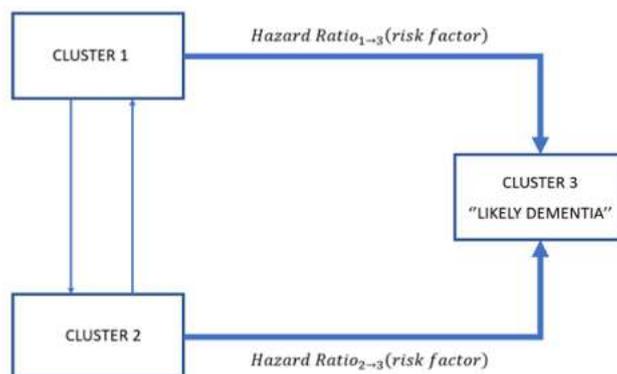




d. Facteurs de risque modifiables de démence

Nous avons testé les facteurs de risque associés à la transition depuis un cluster regroupant des individus sans démence (Clusters 1 et 2) vers un cluster regroupant des personnes avec une démence probable (Cluster 3) grâce à un modèle multi-états.

Figure 6 : Transitions d'intérêt dans le modèle multi-états



Neuf des onze facteurs de risque de démence modifiables étudiés sont associés à un risque accru de transition depuis le Cluster 1 vers l'état de démence probable :

- Faible niveau d'éducation (HR = 1,92 [1,58 – 2,33]),
- Perte auditive (1,74 [1,45 – 2,09]),
- Hypertension (1,35 [1,14 – 1,16]),
- Tabagisme (1,45 [1,13 – 1,87]),
- Dépression (2,51 [1,06 – 3,07]),
- Isolement social (1,66 [1,39 – 1,98]),
- Inactivité physique (3,66 [2,97 – 4,51]),
- Diabète (2,4 [1,94 – 2,96]),
- Obésité (1,7 [1,39 – 2,07]).

Certaines de ces associations étaient également significatives pour le passage du Cluster 2 au cluster de « démence probable » :

- Dépression (2,39 [1,62 – 3,53]),
- Isolement social (2,31 [1,51 – 3,53]),
- Inactivité physique (3,21 [2,12 – 4,87])
- Obésité (1,58 [1,08 – 2,32]).

Ces résultats ont été reproduits dans plusieurs analyses de sensibilité prenant en compte le décès et omettant les cas de démence probables identifiés aux deux premières vagues.

2. Réplication dans ELSA

La grande majorité des résultats obtenus dans SHARE a été retrouvée avec les données de la cohorte ELSA (N = 7840).

Prédiction de la démence

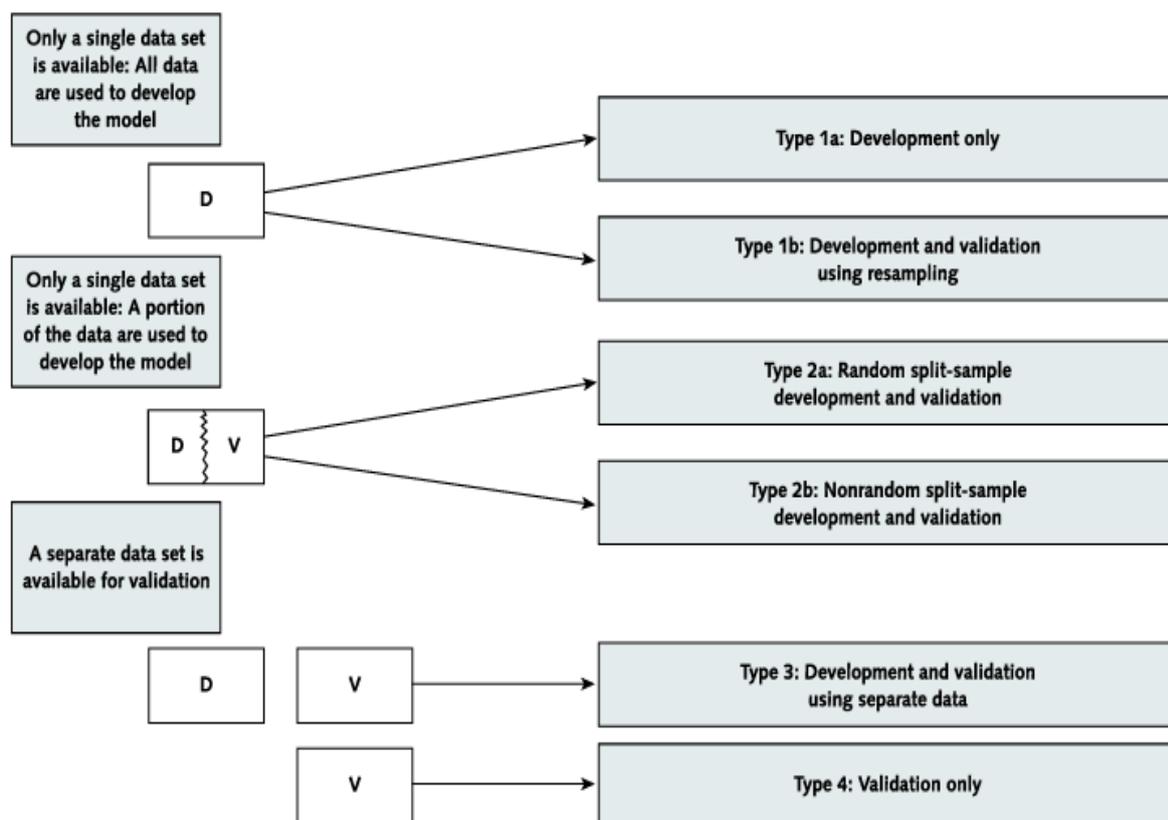
Notre but est de développer un algorithme de prédiction à 10 ans de la démence en utilisant les données des cohortes SHARE et ELSA, ainsi que l'identification des cas de démence probable précédemment effectuée.

Méthodologie

1. Populations d'étude

Selon les recommandations de Kattan et Gerds, il est préférable de tenir compte des risques compétitifs dans le développement des algorithmes de prédiction. Selon les recommandations du groupe MELODEM (MEthods in LONGitudinal DEMentia research), dans le cas de la démence, il est important de tenir compte du risque compétitif de décès. Nous avons également suivi les recommandations de type 3 selon TRIPOD pour développer et valider un algorithme prédictif.

Figure 7 : Recommandations TRIPOD pour le développement et la validation d'un algorithme de prédiction. Nous avons choisi le type 3.



Par conséquent, nous avons choisi d'utiliser les données de SHARE pour le développement de notre algorithme car le statut vital des participants y rapporté. Les données de l'étude ELSA ont été utilisées pour la validation de celui-ci.

Parmi les participants de SHARE âgés de 50 ans ou plus avec un suivi longitudinal et des visites consécutives (N = 15278), nous en avons exclu 105 en raison d'incohérences dans les dates d'interview et de décès (N = 15173) et 514 car ils étaient identifiés comme atteints de démence probable à l'inclusion (N = 14659).

Parmi les participants d'ELSA âgés de 50 ans ou plus avec un suivi longitudinal et des visites consécutives (N = 7840), nous avons soit (1) exclu 27 participants avec un statut de démence auto-rapportée positif ou non-renseignée à l'inclusion (N = 7813) soit (2) exclu 659 participants identifiés comme atteints de démence probable à l'inclusion (N = 7181).

2. Modèle illness-death

Le modèle illness-death est un type de modèle de survie s'intéressant aux transitions entre états : sain, malade et décédé. La méthode choisie utilisait une distribution de type Weibull à risque proportionnels pour les transitions entre les états. Ce modèle permet de prédire le risque de démence à 10 ans.

Après le développement du modèle, nous avons cherché à valider celui-ci en utilisant différentes mesures : la calibration du modèle (c'est-à-dire la concordance entre la probabilité prédite et la probabilité observée), la discrimination à l'aide de l'AUC et l'erreur de prédiction du modèle à l'aide du score de Brier.

3. Covariables

D'après les résultats trouvés dans l'identification de la démence probable, nous avons retenus neuf facteurs de risque de démence modifiables pour ce modèle de prédiction : bas niveau d'éducation, perte auditive, hypertension, tabagisme, dépression, isolement social, inactivité physique, diabète et obésité. Tous ces facteurs étaient binaires et mesurés à l'inclusion.

Résultats

1. Caractéristiques des participants

Dans SHARE, près d'un individu sur dix a développé une démence probable (N = 1404 [9,6%]) au cours du suivi (médiane [écart interquartile], 8,6 ans [2,7 – 12,4]). Ceux identifiés comme probablement atteints de démence sont plus âgés que ceux indemnes à l'inclusion (73,6 ans [9,6] vs 63,7 ans [9,2]) et le plus souvent sont de sexe féminin (N = 825 [58,8%]). Hormis pour le tabagisme, les participants atteints de démence probable sont proportionnellement plus touchés par les facteurs de risque de démence modifiables mesurés à l'inclusion que ceux non-identifiés comme tels. S'agissant de la mortalité, 2511 (17,1%) participants sont décédés au cours du suivi dont 578 (41,2%) chez ceux avec une démence probable et 1933 (14,6%) chez ceux indemnes.

Environ un participant sur vingt d'ELSA a été déclaré comme étant atteint de démence (N = 368 [4,7%]) au cours du suivi (médiane [écart interquartile], 11,9 [4,3 – 16,2]). De la même

manière que dans la cohorte de développement, les personnes avec démence auto-rapportée sont plus âgées que les personnes indemnes (73,3 ans [8,5] vs 64,2 ans [9,8]), sont le plus souvent des femmes (N = 224 [60,9%]). Ils abordent le même profil que les participants de SHARE à l'inclusion s'agissant des facteurs de risque de démence modifiables en termes de distribution. S'agissant des résultats avec la démence probable dans l'étude ELSA, une grande proportion de participants était identifiée comme étant atteinte de démence probable (N = 1491 [20,8%]) au cours du suivi (11,7 [4,6 – 16,2]).

2. Développement du modèle

Tous les facteurs de risque de démence modifiables, pris simultanément dans le modèle illness-death, sont statistiquement liés à une augmentation du risque de développement d'une démence probable.

A partir du modèle développé qui est utilisable pour les individus de 50 ans et plus, il est possible de prédire à un certain horizon temporel (T) pour un nouvel individu plusieurs probabilités (dont la somme est égale à 1) :

- P_{00} : la probabilité d'être vivant et sain (sans démence) à T.
- P_{01} : la probabilité d'être vivant et atteint de démence probable à T.
- P_{02} : la probabilité d'être décédé à T.

Par exemple, pour un homme âgé de 61 ans qui serait fumeur, peu actif physiquement, avec un faible niveau d'éducation et une baisse de l'audition, ces probabilités à 10 ans sont :

- $P_{00} = 72,5\%$ [68,4% – 76,6%].
- $P_{01} = 5,1\%$ [4,2% – 6,2%].
- $P_{02} = 22,4\%$ [19,1% – 26%].

Du fait de l'utilisation d'un modèle paramétrique de type Weibull, il est possible d'établir des prédictions au-delà du temps d'observation maximal disponible. La démence découle en effet d'un processus physiopathologique long (entre 10 et 20 ans) et il est intéressant d'avoir des prédictions longtermes avant la date théorique de l'apparition de la démence. Dans l'étude SHARE, nous disposons d'un recueil sur 13,75 ans maximum. Nous avons donc exploré l'horizon de prédiction de 15 ans. Pour le même sujet que précédemment, l'algorithme fournit à 15 ans les prédictions suivantes :

- $P_{00} = 49,3\%$ [43,2% – 55%].
- $P_{01} = 2,8\%$ [2,4% – 3,3%].
- $P_{02} = 47,9\%$ [42% – 53,8%].

3. Validation du modèle

La validation du modèle illness-death est en cours. Elle nécessite l'application des prédictions à une cohorte longitudinale où le diagnostic de démence est cliniquement

démonstré. Nous avons demandé l'accès à de telles données (projet DRPP, Northwestern University) pour validation finale de l'algorithme.

Apports potentiels de ces résultats

Notre identification de la démence probable pourrait permettre à d'autres équipes de recherche d'étudier les déterminants et les conséquences de la démence probable en Europe grâce à l'analyse de la cohorte SHARE.

Notre algorithme de prédiction de la démence à partir du modèle illness-death pourrait devenir un outil majeur pour la prévention des troubles cognitifs et des démences en Europe. Cet outil a la capacité d'identifier les sujets à risque dès l'âge de 50 ans et offre ainsi l'opportunité de leur proposer des stratégies de prévention des démences alors qu'ils sont encore en activité professionnelle.

Conditions de transférabilité

Ce projet de recherche a été bâti d'emblée avec des données internationales Européennes. La validation finale de l'algorithme passera par une confrontation à des cohortes longitudinales de différents pays. La transférabilité de ces travaux de recherche est attendue au moins au niveau Européen.

Conclusion et Perspectives de recherches

La création de l'algorithme de prédiction de la démence est la première étape du développement d'un outil pour promouvoir le bien vieillir au niveau cognitif. Des travaux d'enquête et des focus groupes avec les usagers potentiels seront encore nécessaire pour déterminer la meilleure méthode pour faire de cet algorithme un outil acceptable et efficace pour la prévention des troubles cognitifs. L'identification des personnes à risque pourrait leur permettre d'être orientées vers des programmes de prévention ciblée visant à réduire leurs facteurs de risque de démence et ainsi diminuer leur risque futur.

REFERENCES

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5, 5th edition. Washington, DC, 2013.
- Haut Conseil de la Santé Publique (HCSP). Prévention de la maladie d'Alzheimer et des maladies apparentées. 2017.
- Global action plan on the public health response to dementia 2017–2025. <https://www.who.int/publications-detail-redirect/9789241513487> (accessed Sept 1, 2023).
- Dementia. <https://www.who.int/news-room/fact-sheets/detail/dementia> (accessed Sept 4, 2023).
- Maladie d'Alzheimer et autres démences. <https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-neurodegeneratives/maladie-d-alzheimer-et-autres-demences> (accessed Sept 4, 2023).
- Lang L, Clifford A, Wei L, *et al.* Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. *BMJ Open* 2017; **7**: e011146.
- Livingston G, Huntley J, Sommerlad A, *et al.* Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet* 2020; **396**: 413–46.

Ngandu T, Lehtisalo J, Solomon A, *et al.* A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet* 2015; **385**: 2255–63.

Cleret de Langavant L, Bayen E, Yaffe K. Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study. *J Med Internet Res* 2018; **20**: e10493.

Gharbi-Meliani A, Husson F, Vandendriessche H, *et al.* Identification of High Likelihood of Dementia in Population-Based Surveys using Unsupervised Clustering: a Longitudinal Analysis. *Alzheimer research & therapy, in press.*

Thomas A. Gerds, Michael Kattan. *Medical Risk Prediction Models: With Ties to Machine Learning.* 2022.