

# Prédire l'entrée dans l'assurance invalidité et la dépendance à long terme : Une approche par Machine Learning

\* Ce rapport a été élaboré pour l'INAMI.

Guida Ayza Estopà (guida.ayza.estopa@ulb.be)  
& Ilan Tojerow (ilan.tojerow@ulb.be)  
Novembre 2025

## 1. Introduction

L'incapacité de travail (IT)<sup>1</sup> constitue un pilier essentiel de la protection sociale dans les économies développées. Au cours des dernières décennies, le nombre de ses bénéficiaires a augmenté de manière substantielle dans les pays de l'OCDE, suscitant des préoccupations quant aux facteurs induisant cette croissance. Alors que certains pays sont parvenus à stabiliser ou même à réduire le nombre de bénéficiaires ces dernières années, la Belgique a quant à elle connu une hausse marquante. La part de la population en âge de travailler percevant des prestations d'invalidité (c'est-à-dire, étant en incapacité de travail depuis plus d'un an) a presque doublé entre 2005 et 2020, passant de 3,5 % à 6,8 % (INAMI, 2022). Parallèlement, la Belgique a également connu une augmentation marquée des flux annuels vers l'incapacité primaire, particulièrement à partir de 2015. Cette double évolution a placé le pays parmi ceux de l'OCDE où les dépenses publiques liées à l'incapacité sont les plus élevées. La Belgique affiche en effet le deuxième niveau le plus élevé de dépenses en espèces pour l'incapacité de travail. En tenant compte à la fois des prestations en espèces et en nature, ses dépenses se situent juste derrière celles de pays nordiques comme la Norvège, le Danemark et la Suède. En 2020, elles représentaient 3,5 % du PIB (OCDE, 2025) (voir Figure A1 dans l'Annexe<sup>2</sup> pour une illustration de ces tendances).

Dans la mesure où l'IT joue un rôle essentiel en protégeant les individus confrontés à des problèmes de santé de longue durée, la compréhension des mécanismes à l'origine de la croissance rapide du nombre de bénéficiaires est devenue un enjeu central au vu de ses répercussions sociétales importantes. Sur le plan de la santé, elle peut en effet indiquer une dégradation du bien-être physique ou mental d'une partie de la population. En termes de marché du travail, elle s'accompagne d'une moindre participation et d'un éloignement accru

---

<sup>1</sup> Tout au long de l'article, les périodes d'incapacité de travail de moins d'un an sont qualifiées d'« incapacité de travail primaire », tandis que le terme « invalidité » est utilisé pour désigner les périodes dépassant un an. Par souci de simplification, le terme « Incapacité de travail (IT) » est souvent utilisé dans l'introduction pour englober ces deux dimensions.

<sup>2</sup> Ce document est accompagné d'un Annexe Technique rédigée en anglais, qui fournit des informations méthodologiques détaillées, des résultats supplémentaires ainsi qu'une description complète des données et du cadre analytique utilisés dans l'étude. (Un lien vers cette annexe sera ajouté dès qu'il sera disponible.)

de l'emploi. Enfin, du point de vue financier, la dépendance prolongée à l'IT exerce une pression croissante sur les dépenses publiques et interroge la soutenabilité à long terme du système. Dans ce contexte, les débats politiques se sont de plus en plus orientés vers des stratégies préventives, visant à réduire l'incidence des nouvelles entrées dans l'IT et à mieux comprendre leurs déterminants. Certains travaux ont ainsi déjà cherché à expliquer la hausse soutenue du nombre de bénéficiaires d'une indemnité d'invalidité observée ces dernières années. Autor et Duggan (2006) soulignent que cette croissance ne peut être imputée uniquement à l'évolution de l'état de santé de la population ou à des facteurs démographiques, mais qu'elle reflète également des changements institutionnels, économiques et incitatifs. Dans la même logique, De Brouwer, Leduc et Tojerow (2019) montrent que le renforcement des obligations de recherche d'emploi a pu conduire certains individus à se tourner vers les dispositifs d'invalidité, contribuant ainsi à une partie de l'augmentation observée. Malgré ces contributions, une part importante de la dynamique reste toutefois encore insuffisamment comprise, comme le suggèrent également Saks (2017) ou De Brouwer & Tojerow (2023).

D'autres travaux mettent quant à eux en évidence des mécanismes plus spécifiques, pour expliquer l'entrée en incapacité de travail ou à la transition vers l'invalidité. Plusieurs études soulignent notamment le rôle des incitations financières (Maestas et al., 2013 ; Kostol & Mogstad, 2021 ; Marie & Vall-Castelló, 2023 ; Ayza et al., 2025), l'impact de réformes institutionnelles ou de l'interaction avec d'autres programmes sociaux (Campolieti & Riddell, 2012 ; Fontenay & Tojerow, 2025 ; De Brouwer et al., 2023), ainsi que — dans une moindre mesure — des différences de genre (Low & Pistaferri, 2019), des délais administratifs (Autor et al., 2015) ou des frictions informationnelles (Kostol & Myhre, 2021). Plus récemment, la qualité de l'emploi et les risques psychosociaux ont également été identifiés comme des déterminants émergents de l'incapacité de travail. Le cas du burnout illustre bien cette tendance : les demandes liées à ce motif ont fortement augmenté dans les dernières années, et plusieurs études confirment par ailleurs que des conditions de travail dégradées ou un stress professionnel élevé constituent des prédicteurs importants de l'entrée dans l'IT (Clumeck et al., 2009 ; Holmgren et al., 2012 ; Nekoei et al., 2025).

En se concentrant sur le cas belge, Saks (2017) suggère que les récentes augmentations du taux de bénéficiaires proviennent en partie d'une participation accrue au marché du travail des femmes et des travailleurs plus âgés. Cependant, De Brouwer et Tojerow (2023), à travers une analyse plus complète et prenant en compte de multiples dimensions, montrent que les changements dans les caractéristiques observables, telles que le genre, la structure par âge ou les caractéristiques des emplois, n'expliquent qu'une part marginale de l'augmentation des entrées dans l'invalidité.

Une étude récente pour la Belgique (INAMI & IMA, 2023) applique des modèles prédictifs afin d'identifier les facteurs sociodémographiques, médicaux et liés aux soins de santé susceptibles d'indiquer le risque d'invalidité chez les individus présentant des troubles de santé mentale.

Leurs résultats indiquent que l'utilisation des services médicaux, en particulier les consultations psychiatriques et les médicaments visant le système nerveux, aide à prédire les transitions vers l'invalidité, bien que le pouvoir discriminant du modèle soit limité. S'il s'agit d'une avancée prometteuse, leur approche repose sur une régression logistique, se limitant dès lors aux individus atteints de troubles mentaux et utilisant un ensemble plus restreint de prédicteurs. À l'inverse, le présent article applique des méthodes de Machine Learning (ML) plus flexibles, construit des modèles prédictifs plus performants et couvre l'ensemble de la population en âge de travailler.

Cet article poursuit deux objectifs principaux. Premièrement, il cherche à comprendre quels sont les déterminants de l'augmentation du nombre d'entrées en incapacité de travail primaire. Deuxièmement, il vise à éclairer la conception de politiques plus efficaces en matière de prévention et d'accompagnement : d'une part, en identifiant en amont les individus les plus susceptibles d'entrer dans l'incapacité, et d'autre part, parmi ceux déjà dans le dispositif, en repérant ceux qui sont davantage susceptibles d'y rester ou, au contraire, d'en sortir. Une meilleure identification de ces profils permettrait potentiellement de concevoir des politiques plus finement ciblées sur les caractéristiques et les besoins des personnes concernées.

Pour cela, le présent article exploite des techniques de ML pour identifier comment différents ensembles de facteurs observés (historiques professionnels, déterminants liés à la santé, caractéristiques socio-économiques) et non observés interagissent pour façonner les trajectoires dans l'IT, identifiant ainsi les prédicteurs les plus pertinents du risque d'IT. L'utilisation du ML nous permet de capturer ces interactions complexes, d'améliorer la fiabilité des prédictions et d'adopter une approche flexible et fondée sur les données pour comprendre les dynamiques de l'IT. Notre analyse s'appuie sur le cadre de Mueller & Spinnewijn (2023), qui examinent les dynamiques de l'assurance chômage sous l'angle de l'hétérogénéité, de la sélection dynamique et de la dépendance à la durée. Nous étendons leur cadre à l'IT, en explorant comment les différences individuelles préexistantes, les évolutions du groupe de risque au fil du temps et les effets de durée influencent à la fois l'entrée dans l'incapacité primaire et la transition vers l'invalidité.

### **1.1. Données, méthodologie et contribution :**

Dans cet article, trois modèles complémentaires sont développés pour analyser les dynamiques d'incapacité de travail. Le premier modèle prédit la probabilité d'entrer en incapacité de travail primaire au cours d'une année donnée. Le deuxième s'intéresse quant à lui aux individus déjà en incapacité primaire et estime leur probabilité de basculer vers l'invalidité. Le troisième modèle étudie finalement la probabilité de sortie du système durant la première année d'incapacité, afin de mesurer l'importance de la sélection dynamique et de

la dépendance à la durée.<sup>3</sup> Ces trois exercices permettent d'examiner plusieurs questions centrales : dans quelle mesure existe-t-il une hétérogénéité dans les risques d'entrée en incapacité primaire, de transition vers l'invalidité, et de sortie ? Quelles dimensions (santé, historique professionnel, caractéristiques socio-économiques) sont les plus prédictives à ces égards ? Les variables liées à la santé, telles que les hospitalisations, la consommation de médicaments ou la fréquence des consultations médicales, dominent-elles, ou bien les trajectoires professionnelles jouent-elles un rôle comparable ? Enfin, pour ce qui est de la transition vers l'invalidité, certaines caractéristiques permettent-elles d'identifier les individus particulièrement susceptibles de rester durablement dans le régime ?

Comprendre ces dynamiques est essentiel, tant pour concevoir des systèmes d'assurance sociale plus efficaces que pour améliorer les outils de prédiction du risque d'incapacité primaire et d'invalidité.

Notre analyse repose sur un jeu de données riche, combinant des données administratives relatives à la carrière professionnelle, à l'utilisation des soins de santé et aux caractéristiques sociodémographiques provenant de deux sources principales : la Banque Carrefour de la Sécurité Sociale (BCSS) et l'Agence Intermutualiste (AIM) en Belgique. Le jeu de données couvre la période 2006–2019 et représente 10 % de la population belge, soit 735.000 individus avec des observations trimestrielles et totalise ainsi près de 40 millions d'observations. Le Tableau 1 présente les différents ensembles de variables utilisées pour l'analyse, et le Tableau 2 fournit les statistiques descriptives pour l'ensemble de l'échantillon ainsi que pour le sous-échantillon d'IT, constitué des individus de notre échantillon ayant connu une période d'IT<sup>4</sup>.

L'article s'appuie sur un cadre conceptuel illustrant comment différentes sources d'hétérogénéité, observées et non observées, peuvent influencer les taux d'entrée dans l'IT ainsi que la probabilité de rester longtemps dans le régime. L'entrée dans l'IT et la dépendance de longue durée sont façonnées par des dynamiques similaires à celles observées dans l'assurance chômage, notamment par l'hétérogénéité du risque initial, la sélection dynamique et la dépendance à la durée. L'hétérogénéité du risque d'IT découle de différences individuelles dans les caractéristiques observables telles que l'âge, le genre, le revenu, les conditions de travail ou l'état de santé (représenté à travers différentes dimensions de dépenses de santé), ainsi que de caractéristiques non observées telles que la résilience ou la propension à rechercher des soins. La sélection dynamique apparaît car les individus présentant des risques de santé plus élevés ou une plus faible adhésion au marché du travail sont moins susceptibles de sortir de l'incapacité de travail, modifiant progressivement la

---

<sup>3</sup> La sélection dynamique renvoie aux changements dans la composition d'un groupe d'individus restant dans un même état donné au fil du temps, ceux ayant des risques plus faibles ou des probabilités de sortie plus élevées quittant plus tôt et concentrant progressivement les individus à risque plus élevé dans le groupe restant. La dépendance à la durée décrit la manière dont la probabilité de sortir d'un état (comme l'incapacité de travail) évolue en fonction du temps déjà passé dans cet état, indépendamment des changements de composition.

<sup>4</sup> L'Annexe Technique B fournit des détails supplémentaires sur les données et les caractéristiques institutionnelles.

composition du groupe de ceux restant en IT. Ce processus est comparable aux dynamiques observées dans l'assurance chômage, où les personnes les plus employables sortent plus tôt du régime, bien qu'il puisse être plus complexe ici. L'Annexe Technique C fournit une discussion détaillée de ce cadre. Alors qu'une littérature abondante a déjà documenté une hétérogénéité substantielle dans les taux de retour à l'emploi des travailleurs au chômage (Álvarez & Shimer, 2011 ; Cockx et al., 2023), peu d'informations existent quant à l'ampleur de l'hétérogénéité et à ses implications pour les résultats d'IT.

**Tableau 1** : Variables incluses dans chaque modèle

Sociodémographique	Marché du travail	Hospitalisation	Médecins	Pharma
Genre	A été au chômage	Jours d'hospitalisation	Consultations avec un médecin généraliste	Médicaments liés à un trouble musculosquelettique – achetés en pharmacie
Âge	Temps de travail	Nombre d'hospitalisation	Consultations avec un médecin spécialiste	Médicaments liés à un trouble musculosquelettique – administrés dans un hôpital public
Marié	Type de travailleur	Remboursement lié à l'hospitalisation	Consultations avec un psychiatre ou un psychologue	Médicaments liés au système nerveux – achetés en pharmacie
Enfants	A été indépendant	Passages aux urgences	Consultations avec un kiné	Médicaments liés au système nerveux – administrés dans un hôpital public
Nationalité Étrangère	Épisodes antérieurs d'incapacité du travail		Remboursements des consultations avec un médecin spécialiste	Antidépresseurs achetés en pharmacie
Région	Salaire			Antidépresseurs administrés dans un hôpital Remboursements des dépenses pharmaceutiques

*Note : Pour les informations historiques, nous utilisons les deux années précédentes. Le revenu du travail renvoie au revenu perçu l'année précédente. Les consultations médicales et les variables pharmaceutiques sont incluses sous différentes formes : une variable indicatrice signalant au moins une observation et le nombre total d'occurrences. Le temps de travail peut être à temps plein ou à temps partiel ; le type de travailleur peut être travailleur du secteur public, ouvrier (« col-bleu ») ou employé (« col-blanc »).*

L'ensemble de ces éléments soutient l'utilisation des modèles de ML, capables de capturer des interactions complexes et non linéaires, et montre leurs avantages par rapport aux approches

économétriques traditionnelles pour prédire les résultats d'IT. Pour la mise en œuvre empirique, nous employons des techniques standard de ML. Nous entraînons les modèles prédictifs sur un échantillon d'entraînement et évaluons leur performance sur un échantillon de validation afin d'éviter le surapprentissage. Le principal défi dans tous les exercices de prédiction réside dans le compromis entre l'amélioration du modèle et le risque de surapprentissage lorsque trop de variables sont incluses. Les méthodes de ML et la séparation des échantillons contribuent à optimiser la sélection des variables et à gérer ce compromis entre précision prédictive et surapprentissage dans un environnement riche en données. Nous nous concentrons sur trois résultats principaux : (i) la probabilité d'entrer en incapacité de travail primaire, (ii) la probabilité de basculer de l'incapacité de travail primaire vers l'invalidité et (iii) la probabilité de sortir de l'incapacité de travail durant la première année. Nous définissons ces probabilités comme les variables de risque de notre modèle. Nous utilisons deux modèles d'apprentissage automatique : le *Random Forest* et le *Gradient Boosted Regression Trees*, respectivement fondés sur l'agrégation d'arbres indépendants et sur une construction séquentielle corrigeant les erreurs successives, ce qui permet de capter des interactions et non-linéarités complexes. Ils sont ensuite agrégés dans un modèle d'Ensemble, qui est une combinaison linéaire pondérée de ceux-ci. Après avoir ajusté les modèles sur l'échantillon d'entraînement, nous les estimons pour obtenir les prédictions du modèle d'Ensemble et les probabilités calibrées pour chaque résultat, puis appliquons le modèle entraîné à un échantillon témoin (« Hold-out sample ») non utilisé au préalable<sup>5</sup>.

**Tableau 2** : Statistiques descriptives

Échantillon de population vs. Échantillon IT		
	L'ensemble de l'échantillon	Échantillon IT
Âge (moyenne)	40.9	44.8
Femme	50.1%	40.5%
Étranger	22.2%	22.5%
Médicament "Group N" en pharmacie	34.4%	49.9%
Médicament "Group N" en hôpital	7.5%	13.7%
Médecin généraliste	95.1%	99.6%
Hospitalisation	51.6%	78.0%
Épisode de chômage	6.8%	11.4%
Épisode d'incapacité travail antérieur	6.5%	100%
Épisode d'invalidité antérieur	2.7%	7.4%

Note : Statistiques descriptives pour l'échantillon de base couvrant les années 2006-2018 et la tranche d'âge 16-65 ans. L'échantillon IT inclut les individus en incapacité de travail de courte ou de longue durée.

Cet article contribue à la littérature de trois manières principales. Premièrement, il fournit la première documentation systématique de l'hétérogénéité des risques d'IT à l'entrée et dans les transitions vers la dépendance de longue durée. Deuxièmement, en appliquant le ML à des

<sup>5</sup> L'Annexe Technique D fournit les détails méthodologiques.

données administratives riches, il quantifie la part d'hétérogénéité attribuable aux facteurs observables et capture des interactions complexes. Troisièmement, il apporte de nouvelles preuves sur la dimension dynamique du risque d'IT, montrant comment la prévisibilité évolue avec la durée des épisodes et les horizons de prévision. Ensemble, ces conclusions ouvrent de nouvelles perspectives pour la conception de politiques préventives en identifiant les profils les plus à risque d'entrer dans l'IT et d'y rester à long terme.

## 2. Résultats

Nous présentons à présent les résultats des trois modèles prédictifs. Premièrement, nous examinons le modèle de base estimant la probabilité d'entrer en incapacité de travail primaire. Deuxièmement, nous analysons le modèle prédisant les transitions vers l'invalidité. Troisièmement, nous étudions les dynamiques au cours de la première année d'incapacité de travail.

### 2.1. Probabilité de commencer une période d'incapacité de travail

Le premier modèle prédit la probabilité qu'un individu entre en incapacité de travail primaire au cours d'une année donnée, en utilisant les informations des deux années précédentes. La spécification de base se concentre sur les prédictions pour 2018 et est estimée sur un échantillon aléatoire de 10 % de la population belge en âge de travailler n'ayant connu aucun épisode d'incapacité de travail au cours des deux années précédentes. Pour chaque individu, le modèle attribue une probabilité prédite de commencer un nouvel épisode d'incapacité de travail primaire.

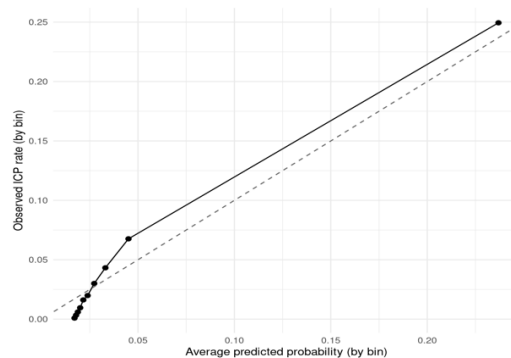
Pour évaluer la qualité d'un modèle prédictif, deux mesures principales sont utilisées : l'AUC et le  $R^2$ . L'AUC (Area Under the Curve) mesure la capacité du modèle à distinguer correctement les individus qui entreront en incapacité de ceux qui n'y entreront pas ; une valeur de 0,5 correspond au hasard, tandis qu'une valeur proche de 1 reflète une discrimination élevée. Dans notre cas, l'AUC atteint 0,86, ce qui indique que le modèle différencie très efficacement les individus à risque élevé de ceux à faible risque, un résultat remarquable pour un phénomène rare et multifactoriel. Le  $R^2$ , pour sa part, mesure la part de la variation des entrées en incapacité pouvant être expliquée par les caractéristiques observées. Avec un  $R^2$  de 18,9 %, environ un cinquième de la variation est attribuable à des facteurs tels que la santé, les revenus ou l'historique d'emploi. Bien qu'un tel niveau puisse sembler modeste en valeur absolue, il est élevé pour ce type d'exercice prédictif. À titre de comparaison, Mueller et Spinnewijn (2023) obtiennent un  $R^2$  d'environ 15 % lorsqu'ils prédisent les sorties de l'assurance chômage à l'aide de données administratives similaires.

Le panneau A de la Figure 1 compare les probabilités prédites aux entrées effectives en incapacité de travail primaire en 2018, montrant un alignement clair : les groupes présentant

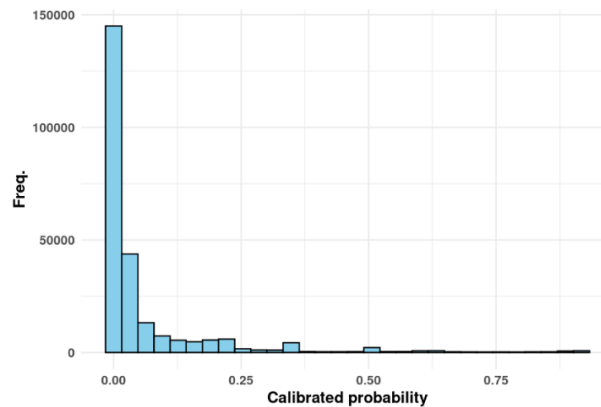
un risque prédit plus élevé affichent des taux d'entrée observés plus élevés. Cette correspondance étroite entre prédictions et réalisations illustre la forte capacité discriminante du modèle et souligne la qualité de la prédiction obtenue. Le panneau B présente la distribution des probabilités prédites dans la population active. Comme attendu, la plupart des individus se concentrent autour de zéro, puisque la majorité ne connaît jamais d'épisode d'incapacité, tandis qu'une petite minorité se distingue nettement avec des risques prédits plus élevés.

**Figure 1** : Probabilité de commencer une incapacité de travail en 2018

Panneau A : Comparaison entre les prédictions du modèle et les résultats observés



Panneau B : Distribution de la probabilité prédite d'entrée en incapacité de travail



*Note : Le panneau A présente un graphique de calibration comparant les probabilités prédites et observées d'entrée en incapacité de travail dans l'échantillon de validation pour l'année 2018. Les individus sont regroupés en 10 déciles selon leur probabilité prédite d'entrée en incapacité. Pour chaque groupe, nous représentons la probabilité prédite moyenne et le taux observé d'entrée en incapacité. La ligne pointillée représente la calibration parfaite (c.-à-d. prédiction = observation). Comme le graphique utilise des moyennes par décile, les valeurs ne couvrent pas l'ensemble de l'intervalle 0–1, bien que les prédictions individuelles le fassent. Le panneau B montre la distribution de la probabilité prédite d'entrée en incapacité de travail dans l'échantillon de validation pour l'année 2018.*

La Figure A3 en annexe trace la distribution des probabilités prédites après exclusion de toutes celles inférieures à 0,05 pour rendre la partie supérieure de la distribution plus visible. Même dans cet échantillon tronqué, la plupart des individus restent concentrés sur des niveaux de risque relativement faibles, reflétant la rareté des entrées en incapacité de travail primaire. Cependant, le modèle attribue des probabilités sur toute la plage jusqu'à des valeurs proches

d'un. Bien que ces cas à haut risque soient très rares, ils montrent que le modèle parvient à identifier un petit groupe d'individus avec des risques prédits très élevés. La distribution reste fortement asymétrique, ce qui révèle qu'un petit groupe concentre l'essentiel du risque prévisible. Ce schéma met en lumière le potentiel pour une identification précoce et une prévention ciblée.

Pour comprendre quels facteurs contribuent le plus à la précision prédictive, le Tableau 3 présente la performance de modèles basés sur différents groupes de variables. Le panneau A ajoute progressivement les informations du marché du travail et de la santé à un modèle de base ne contenant que les caractéristiques sociodémographiques. Les seules caractéristiques sociodémographiques n'expliquent qu'une faible partie de la variation dans le risque d'entrée. L'ajout des informations du marché du travail accroît la capacité explicative de 167 % par rapport au modèle de base. L'inclusion des variables liées à la santé produit l'amélioration la plus importante : la puissance explicative augmente alors de 278 %. Ceci confirme que les indicateurs liés à la santé sont de loin les prédicteurs les plus puissants du risque d'incapacité de travail. Le panneau B examine la contribution marginale des différents sous-groupes de variables de santé : les hospitalisations contribuent le plus, suivies par la consommation pharmaceutique, tandis que les consultations avec spécialistes ou généralistes ajoutent des améliorations plus modestes mais significatives.

**Tableau 3** : R<sup>2</sup> pour différents modèles en 2018

A. Sous-modèles du modèle de référence – Séquentiel

	c (1)	c (2)	c (3)
R <sup>2</sup>	0.0187	0.0499	0.1887
Variation (c) vs (c-1)		+167%	+278%
Sociodémographique	X	X	X
Variables du marché du travail		X	X
Variables liées à la santé			X

B. Décomposition de la contribution des différentes dimensions de santé – Marginal

	c (1)	c (2)	c (4)	c (5)	c (3)	c (6)
R <sup>2</sup>	0.0187	0.0238	0.0281	0.0335	0.1128	0.1301
Variation (c) vs (1)		+27%	+50%	+79%	+503%	+596%
Sociodémographique	X	X	X	X	X	X
Médecin généraliste		X				X
Spécialistes			X			X
Consommation des médicaments				X		X
Hospitalisations					X	X

Note : Le tableau présente le R<sup>2</sup> obtenu en régressant la probabilité prédite d'entrée en incapacité de travail sur une variable indicatrice d'entrée effective dans l'échantillon de validation pour l'année 2018. Le panneau A, commence avec le modèle de base en c (1) et ajoute les groupes de variables séquentiellement jusqu'à inclure l'ensemble des variables du modèle de référence en c (3). Le panneau B utilise les mêmes groupes de variables mais décompose l'information liée à la santé en différents sous-groupes, ajoutés d'abord séparément, puis simultanément dans la colonne c (6).

Bien que le modèle prédictif n'estime pas des effets causaux, il permet d'évaluer la contribution de différents groupes de variables à la performance prédictive. Parmi les prédicteurs liés à la santé, le nombre de jours d'hospitalisation et les admissions ressortent comme les plus importants, suivis des consultations chez le généraliste et le spécialiste. Parmi les variables du marché du travail, la classe salariale de l'année précédente est le prédicteur le plus fort. L'âge est la caractéristique sociodémographique la plus pertinente une fois les autres informations prises en compte<sup>6</sup>.

Pour illustrer de manière intuitive la pertinence des variables, nous avons estimé une régression linéaire standardisée des probabilités calibrées sur l'ensemble des prédicteurs, suivant Mueller et Spinnewijn (2023). Les coefficients sont exprimés en unités d'écart-type, montrant comment chaque caractéristique est associée (positivement ou négativement) au risque prédit d'entrer en incapacité de travail. La Figure 2 présente ces résultats. Comme attendu, les variables liées à la santé (en vert) dominent, en particulier les consultations médicales, les hospitalisations, les remboursements pharmaceutiques et les consultations de spécialistes. Fait intéressant, les consultations psychiatriques affichent une association négative, suggérant qu'un traitement psychiatrique précoce pourrait réduire la probabilité d'entrer en incapacité. Les caractéristiques du marché du travail (en orange), comme la classe salariale et le type d'emploi, jouent également un rôle important. Les coefficients positifs pour les salaires plus élevés et les professions « col-blanc<sup>7</sup> » peuvent refléter que les individus occupant des emplois plus stables ou dotés d'une plus grande protection sont moins contraints lorsqu'ils ont besoin d'un congé maladie. En revanche, les travailleurs du secteur public et les indépendants présentent des associations négatives, indiquant qu'ils sont moins susceptibles de commencer une période d'incapacité. Les variables sociodémographiques (en bleu) jouent un rôle plus limité une fois les autres dimensions contrôlées : être une femme et avoir des enfants sont associés à un risque plus élevé, tandis que l'âge présente une association négative.

Il est important de souligner que cette régression interprétative ne reproduit pas la manière dont les variables contribuent réellement aux prédictions dans le modèle de machine learning. En effet, dans un modèle non linéaire, la contribution d'une variable dépend à la fois de ses interactions avec d'autres caractéristiques et de relations complexes que la régression linéaire ne peut pas capturer. Les coefficients OLS résument uniquement des associations moyennes et linéaires, tandis que le modèle de ML tient compte de relations beaucoup plus riches. Pour cette raison, certaines variables catégorielles, comme le fait d'être travailleur public, peuvent afficher des coefficients linéaires élevés malgré une contribution moindre à la performance globale du modèle non linéaire.

Nous avons également estimé le modèle incluant les individus ayant déjà connu un épisode d'incapacité. Celui-ci affiche des performances encore supérieures, avec une AUC de 0,88 et

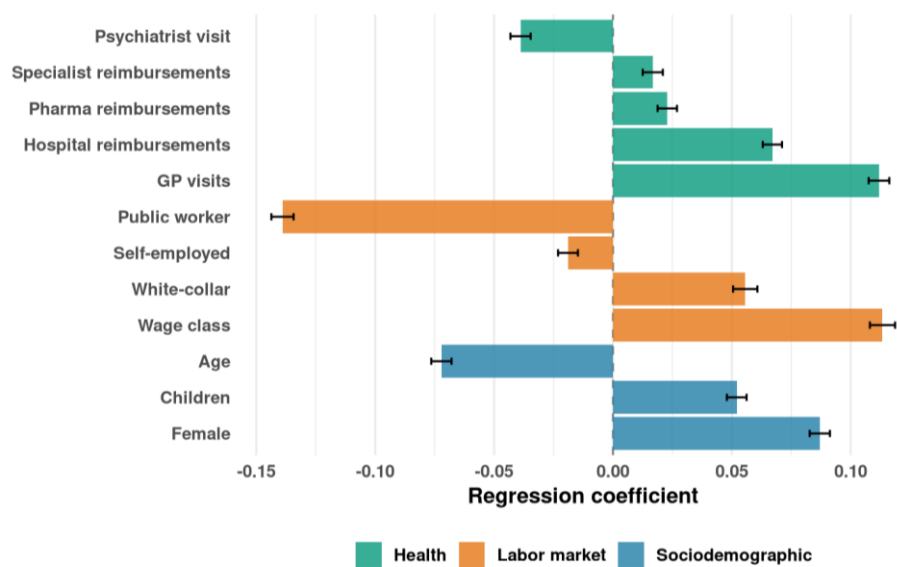
---

<sup>6</sup> La Figure A5 dans l'Annexe Technique présente l'ensemble des prédicteurs et leur importance relative.

<sup>7</sup> Pour « col-blanc », nous faisons référence aux employés. Pour « col-bleu », nous faisons référence aux ouvriers.

un  $R^2$  de 24,8 %. Sans surprise, le prédicteur le plus important dans cette spécification est le fait d'avoir connu un épisode antérieur d'incapacité de travail. Cette variable capture à elle seule une grande partie du risque prévisible, puisque les individus en incapacité au cours des deux années précédentes sont beaucoup plus susceptibles de connaître un nouvel épisode. Bien que cette persistance soit informative sur la récurrence et la dépendance à long terme, elle limite notre capacité à étudier les premières entrées. Les résultats de ce modèle apparaissent dans la Figure A4 en Annexe Technique.

**Figure 2 : Hétérogénéité du risque d'incapacité de travail**



*Note : Cette figure présente les résultats de régressions linéaires des prédictions sur un sous-ensemble de variables observables. Les variables sont standardisées en soustrayant la moyenne de l'échantillon et en divisant par l'écart-type, de sorte que les coefficients peuvent être interprétés comme la variation (en écarts-types) du résultat associé à une variation d'un écart-type de la covariable. Elle montre les coefficients OLS d'une régression de la probabilité prédite d'entrée en incapacité (modèle de base 2018) sur les variables figurant en ordonnée. Des intervalles de confiance à 95 % entourent les estimations ponctuelles. Les ouvriers (col-bleu) constituent la catégorie de référence pour le type de travailleur ; les coefficients pour les employés (col-blanc) et les travailleurs du secteur public s'interprètent donc relativement à cette catégorie. Pour les variables indicatrices — femme, indépendant, avoir des enfants, visite chez un psychiatre — la catégorie omise est « ne pas avoir la caractéristique ». Les variables continues (âge, salaire, remboursements médicaux) entrent sous forme standardisée.*

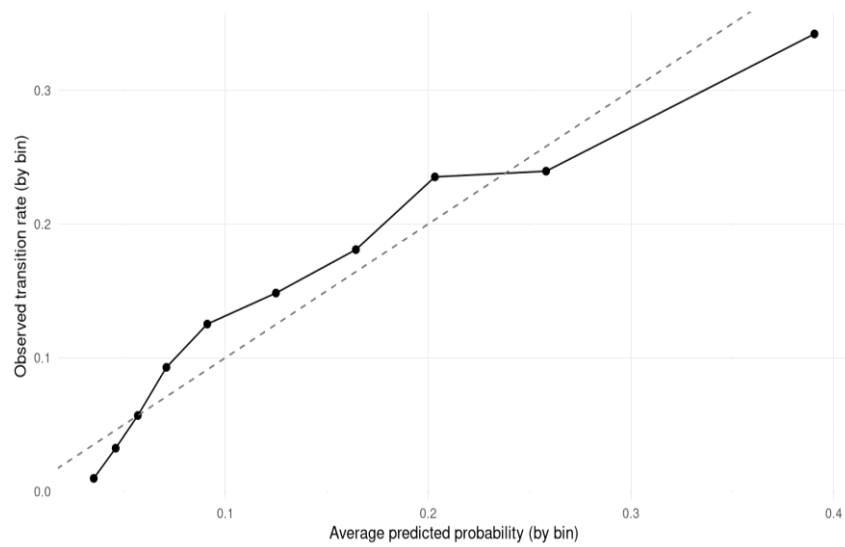
Globalement, ces résultats montrent que le modèle prédisant la probabilité d'entrer en incapacité primaire présente une forte capacité prédictive et une hétérogénéité observable marquée. L'état de santé et les antécédents médicaux sont les principaux déterminants du risque d'incapacité, suivis par les facteurs liés au marché du travail, tandis que les caractéristiques sociodémographiques contribuent relativement peu. Dans le même temps, environ deux tiers de la variance restent inexpliqués, suggérant que l'hétérogénéité non observée — telles que des conditions de santé non mesurées, la résilience ou des comportements individuels — continue de jouer un rôle important.

En somme, les entrées en incapacité de travail sont fortement concentrées parmi un groupe relativement restreint de travailleurs dont les antécédents récents en matière de santé et

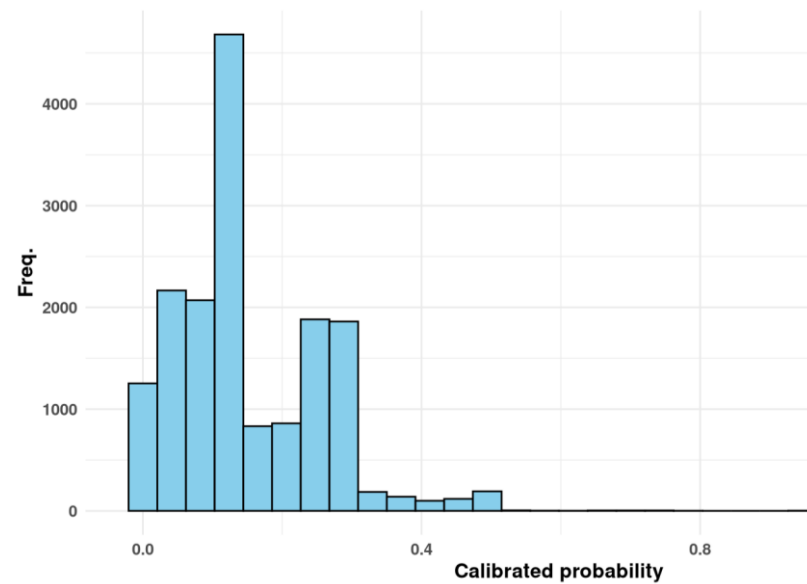
d'emploi signalent déjà un risque élevé. Cette concentration offre des opportunités importantes pour des politiques de prévention ciblées et d'intervention précoce.

**Figure 3 :** Probabilité de transition de l'incapacité de travail vers l'invalidité en 2018

Panneau A : Comparaison entre les prédictions du modèle et les résultats observés



Panneau B : Distribution de la probabilité prédite de transition



Note : Le panneau A présente un graphique de calibration comparant les probabilités prédites et observées de transition de l'incapacité de travail vers l'invalidité dans l'échantillon de validation pour 2018. Les individus sont regroupés en 10 déciles selon leur probabilité prédite d'entrée en incapacité. Pour chaque groupe, nous traçons la probabilité prédite moyenne et le taux observé de transition. La ligne pointillée représente la calibration parfaite. Comme pour la Figure 1, les valeurs ne couvrent pas tout l'intervalle 0–1 parce qu'elles se basent sur des moyennes de décile. Le panneau B présente la distribution de la probabilité prédite de transition vers l'invalidité dans l'échantillon de validation pour 2018.

## 2.2. Probabilité de transition de l'incapacité de travail vers l'invalidité

Notre second modèle se concentre sur les individus déjà en incapacité de travail et prédit ceux qui entreront en invalidité au cours de l'année suivante. Après un an d'incapacité, les individus peuvent devenir éligibles au statut d'invalidité, à la suite d'une nouvelle évaluation médicale. Ce passage les fait entrer dans le régime de longue durée, qui prévoit moins de révisions médicales et applique des règles spécifiques de calcul des prestations. Celles-ci reposent toujours sur le salaire antérieur et la composition du ménage, mais avec des pourcentages de remplacement et des seuils minimum et maximum différents, de sorte que le montant peut augmenter, diminuer ou rester similaire selon la situation familiale et le niveau de salaire. Comprendre cette transition est crucial, car elle capture la persistance de l'incapacité et les facteurs menant à une dépendance de longue durée.

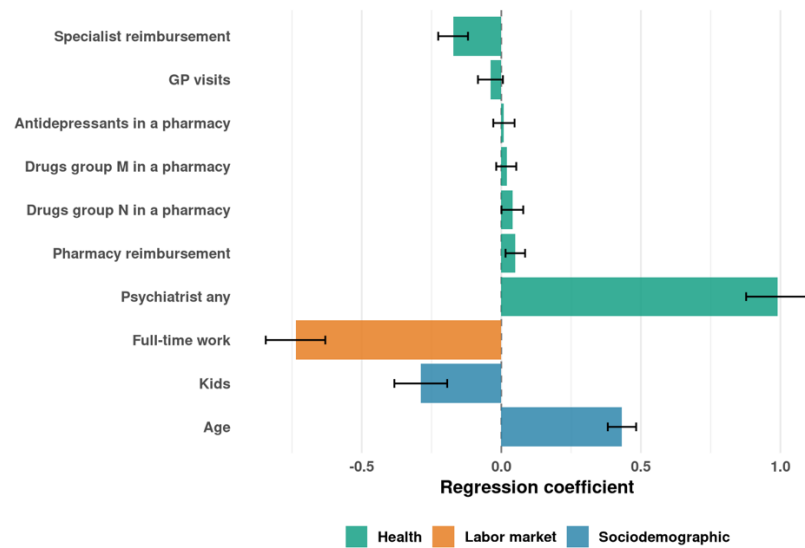
Le modèle atteint une AUC de 0,72 et un  $R^2$  de 0,073, indiquant que, bien que les prédictions différencient raisonnablement entre ceux qui feront et ne feront pas la transition vers l'invalidité, seule une part modeste de la variation totale est expliquée par les caractéristiques observables. Autrement dit, le modèle est capable d'identifier des individus avec un risque relativement plus élevé ou plus faible, mais une grande partie de la persistance de l'incapacité semble être déterminée par des facteurs non observés, tels que des trajectoires de santé sous-jacentes, des processus de récupération ou des comportements liés à des aléas moraux.

Le panneau A de la Figure 3 compare les probabilités de transition observées et prédites, montrant que les risques prédits s'alignent avec les résultats réels, bien que la dispersion soit plus large que dans le modèle d'entrée. Le panneau B illustre la distribution des probabilités prédites. Comme l'analyse est limitée aux individus déjà en incapacité de travail, l'événement de transition vers l'invalidité est moins rare que l'entrée initiale, produisant une distribution plus étalée des risques prédits. Si la plupart des individus se situent dans des niveaux moyens, une part non négligeable dépasse des probabilités prédites de 0,5, ce qui indique que le modèle peut identifier un sous-groupe d'individus à risque élevé. Dans l'ensemble, la probabilité de basculer de l'incapacité primaire vers l'invalidité est beaucoup moins prévisible que la probabilité d'une première entrée ; néanmoins, malgré la plus faible puissance explicative, l'analyse de l'importance des prédicteurs reste instructive.

La Figure A5 dans l'Annexe Technique présente les scores d'importance des variables pour le modèle complet, et la Figure 4 affiche les coefficients de la régression linéaire standardisée, analogue à l'exercice mené pour le modèle d'entrée. Le classement des prédicteurs diffère nettement de celui du modèle d'entrée. L'âge apparaît comme le facteur dominant et est associé à un coefficient positif, cohérent avec le fait que les travailleurs plus âgés ont des perspectives de récupération plus faibles et sont plus susceptibles de passer vers l'invalidité. Les indicateurs de santé mentale, tels que les consultations psychiatriques, figurent également parmi les principaux prédicteurs, ce qui suggère que les troubles mentaux jouent un rôle

central dans l'incapacité de longue durée. La dépense pharmaceutique est un autre prédicteur majeur, reflétant la charge de santé globale plutôt que des conditions médicales spécifiques. Enfin, l'intensité de l'emploi avant l'incapacité (par exemple travailler à temps plein ou à temps partiel avant l'entrée dans le régime) importe également, ce qui suggère que les individus avec une attache plus faible au marché du travail ont une probabilité plus élevée de rester hors de l'emploi.

**Figure 4 : Hétérogénéité du risque de transition vers l'invalidité**



*Note : Cette figure présente les résultats de régressions linéaires des prédictions sur un sous-ensemble de variables observables. Les variables sont standardisées, ce qui permet d'interpréter les coefficients comme l'effet d'un écart-type de variation dans la covariable. Elle montre les coefficients OLS de la probabilité prédite de transition vers l'invalidité (modèle 2018) sur les variables listées en ordonnée. Des intervalles de confiance à 95 % entourent les estimations. Pour les variables indicatrices — travail à temps plein, enfants, visite chez un psychiatre — la catégorie omise est « ne pas avoir la caractéristique ». Les variables continues (âge, consommation de médicaments, remboursements médicaux) entrent sous forme standardisée.*

Globalement, ces résultats mettent en évidence une asymétrie claire entre les déterminants de l'entrée en incapacité de travail et ceux de la persistance vers l'invalidité. L'entrée est davantage prévisible à partir des conditions de santé préexistantes et des trajectoires professionnelles, tandis que la persistance au-delà d'un an l'est beaucoup moins. Une fois les individus déjà en incapacité, les dynamiques de santé ou de travail non observées jouent un rôle croissant.

Deux mécanismes complémentaires peuvent aider à expliquer ce schéma : la sélection dynamique — par laquelle les individus ayant les meilleures perspectives de récupération sortent tôt, modifiant ainsi la composition du groupe restant — et la dépendance à la durée — par laquelle la probabilité de basculer de l'incapacité primaire vers l'invalidité augmente avec le temps passé dans le régime, en raison d'une détérioration réelle ou d'une adaptation au dispositif. Nous examinons l'importance relative de ces mécanismes dans la section suivante

en analysant l'évolution de la performance prédictive au fil du temps dans les épisodes d'incapacité primaire.

### 2.3. Dynamiques de prédictibilité au sein des épisodes d'incapacité de travail

Nous étudions à présent l'évolution de la performance prédictive au cours des épisodes d'incapacité de travail, en suivant l'approche dynamique proposée par Mueller et Spinnewijn (2023). Cette analyse examine si la capacité du modèle à prédire la persistance varie selon le temps passé en incapacité et selon l'horizon de prévision. Comprendre ces dynamiques est essentiel pour déterminer si la persistance est principalement due à la sélection dynamique — lorsque les individus présentant des probabilités de sortie plus élevées (meilleure santé, meilleure employabilité, plus grande résilience, ...) quittent le régime tôt, laissant un groupe résiduel plus homogène et caractérisé par des probabilités de sortie plus faible — ou à la dépendance à la durée, lorsque les risques évoluent d'une manière non capturée par les variables observables, possiblement liée à des enjeux d'aléa moraux ou à une familiarisation accrue avec le dispositif à mesure que le temps s'écoule.

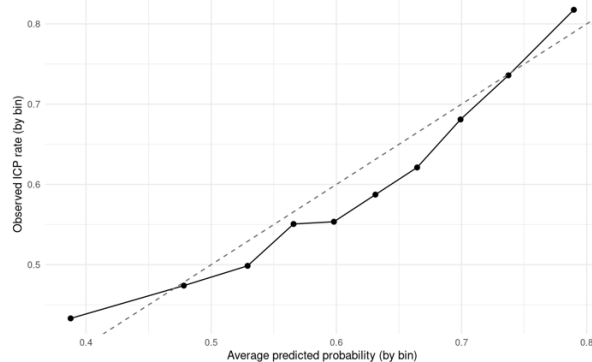
Nous commençons par un modèle dynamique de référence qui prédit si les individus qui viennent d'entrer en incapacité de travail restent dans le régime pendant au moins deux trimestres consécutifs ou en sortent plus tôt, en utilisant uniquement l'information disponible au début de la période, autrement dit, si les personnes sortent avant ou après deux trimestres. Ce modèle de base capture le degré auquel la persistance à court terme peut être anticipée ex ante. Comme pour les deux résultats précédents, le panneau A de la Figure 5 compare les probabilités observées et prédites, montrant que les risques prédits s'alignent raisonnablement bien avec les résultats réels, même si ce modèle est clairement plus faible que les précédents ( $AUC = 0,64$  ;  $R^2 = 6\%$ ). Une AUC de 0,64 indique que le modèle distingue les sortants des non-sortants mieux que le hasard, mais avec une capacité discriminante limitée. De même, un  $R^2$  de 6 % montre que seule une petite part de la variation observée dans les sorties est expliquée par les variables disponibles, ce qui reflète la difficulté intrinsèque de prédire des trajectoires très courtes au début de l'incapacité. Le panneau B trace la distribution des probabilités prédites, plus dispersée, reflétant la plus grande difficulté à discriminer la persistance à court terme parmi des individus venant juste d'entrer en incapacité.

Nous estimons ensuite un modèle conditionnel restreint aux individus restant en incapacité de travail pendant deux trimestres, et prédisons s'ils continueront encore deux trimestres supplémentaires ou, au contraire, s'ils sortiront après quatre trimestres. La comparaison de ces deux exercices fournit une évaluation directe de l'évolution du risque une fois les sorties précoces réalisées. Le panneau A de la Figure 6 compare les distributions de risques prédites au début de l'épisode et après deux trimestres. La distribution au deuxième trimestre est plus concentrée et décalée vers des probabilités de sortie plus faibles, reflétant que les individus restant en incapacité après les premiers mois présentent des risques de sortie durablement

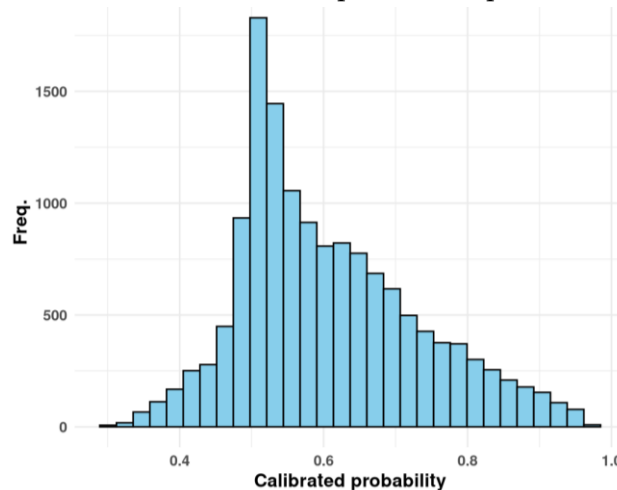
faibles. Cette réduction de la dispersion est cohérente avec la sélection dynamique : les individus à risque élevé sortent tôt, laissant un groupe plus homogène avec des probabilités systématiquement plus faibles de quitter l'incapacité. En conséquence, la performance prédictive s'améliore légèrement, l'AUC et le  $R^2$  augmentant respectivement à 0,65 et 8,5 %. Ces valeurs restent modestes, mais montrent que, une fois les sorties précoces éliminées, les différences observables entre individus expliquent légèrement mieux les probabilités de sortie. Si ces mesures avaient diminué et si la distribution avait convergé vers une forte persistance indépendamment des caractéristiques individuelles, cela aurait indiqué une dépendance à la durée.

**Figure 5 :** Probabilité de sortir de l'incapacité de travail dans les deux premiers trimestres de l'épisode (2018)

Panneau A : Comparaison entre les prédictions du modèle et les résultats observés



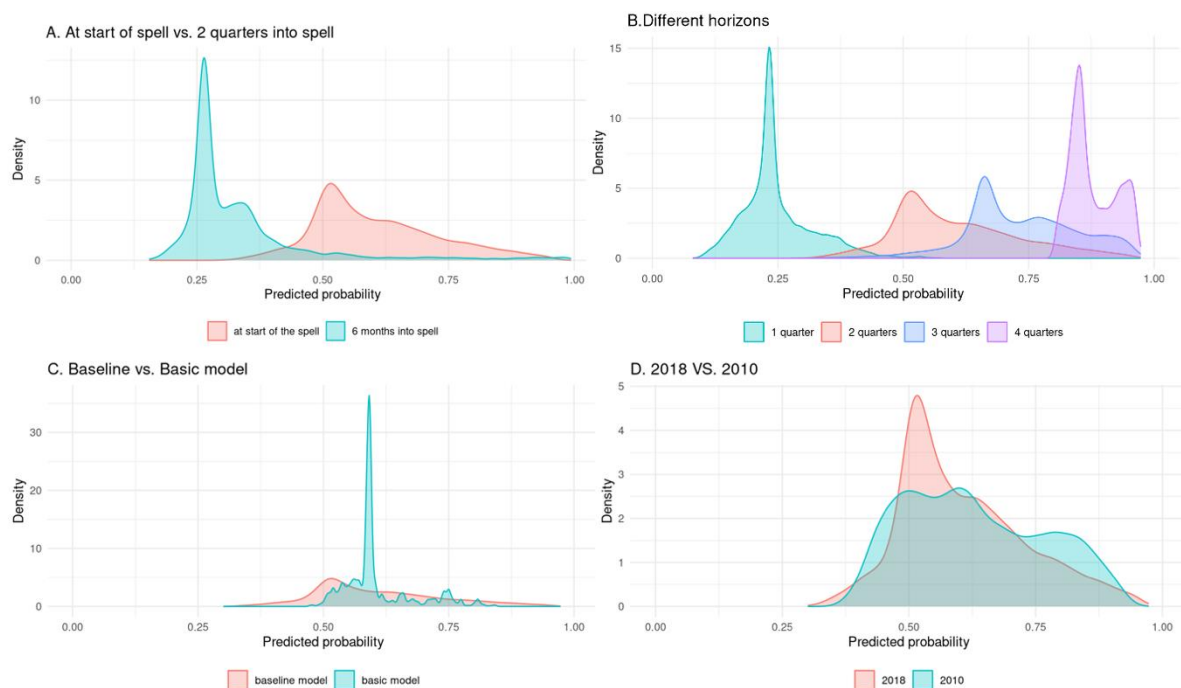
Panneau B : Distribution de la probabilité prédite de sortie



*Note :* Le panneau A montre un graphique de calibration comparant les probabilités prédites et observées de sortie de l'incapacité de travail au cours des deux premiers trimestres dans l'échantillon de validation pour 2018. Les individus sont regroupés en 10 déciles selon leur probabilité prédite d'entrée en incapacité. Pour chaque groupe, nous traçons la probabilité prédite moyenne et le taux observé de sortie. La ligne pointillée représente la calibration parfaite. Le panneau B montre la distribution de la probabilité prédite de sortie dans l'échantillon de validation pour 2018.

Nous examinons ensuite comment la prédictibilité varie selon la longueur de l'horizon de prévision, en maintenant constant l'ensemble d'information. Le panneau B de la Figure 6 trace les distributions des risques de sortie prédites pour des horizons d'un, deux, trois et quatre trimestres. Les prédictions à court terme sont concentrées sur des probabilités de sortie faibles, reflétant qu'une faible part d'individus quittent l'incapacité au trimestre suivant. À mesure que l'horizon s'étend à deux ou trois trimestres, les distributions se déplacent vers la droite et deviennent plus dispersées, ce qui correspond à l'accumulation du risque de sortie dans le temps et à une hétérogénéité observable plus grande à moyen terme. La précision prédictive s'améliore en conséquence : l'AUC augmente de 0,58 à 0,64 puis 0,66, tandis que le  $R^2$  passe de 0,015 à 0,056 puis 0,060. Cependant, lorsque l'horizon atteint quatre trimestres, les deux mesures diminuent, suggérant que les trajectoires à long terme sont de plus en plus influencées par l'hétérogénéité non observée et deviennent intrinsèquement moins prévisibles. Le Tableau A1 (panneau B) en annexe résume les valeurs correspondantes d'AUC et de  $R^2$ .

**Figure 6 :** Distribution des probabilités prédites de sortie de l'incapacité de travail



*Note :* Cette figure présente la distribution de différentes probabilités prédites de sortie de l'incapacité de travail. Dans les quatre panneaux, la distribution de référence correspond aux probabilités de sortie sur deux trimestres prédits au début de l'épisode. Le panneau A compare cette distribution de référence avec la distribution prédite pour les deux trimestres suivants, conditionnellement au fait que l'individu soit resté au moins deux trimestres en incapacité. Le panneau B présente les probabilités de sortie prédites pour différents horizons — un, deux, trois et quatre trimestres — en utilisant le même ensemble d'information. Le panneau C compare les prédictions du modèle complet (incluant les variables détaillées de santé, marché du travail et sociodémographiques) avec celles d'un modèle réduit utilisant seulement un ensemble limité de prédicteurs. Le panneau D compare les distributions de risque prédites pour deux cohortes, 2010 et 2018, afin d'évaluer la stabilité de l'hétérogénéité observable dans le temps.

Le panneau C évalue à quel point la performance prédictive dépend de la richesse de l'ensemble d'information. Un modèle limité à un petit nombre de variables de base génère une

distribution avec des pics marqués, reflétant que, avec peu de prédicteurs, le modèle attribue des probabilités très similaires à de nombreux individus. À l'inverse, le modèle complet, qui incorpore des variables détaillées liées à l'utilisation des soins de santé, aux trajectoires professionnelles et aux caractéristiques sociodémographiques, produit des distributions plus lisses et différenciées. Cette comparaison montre que l'information administrative riche est essentielle pour une prédiction pertinente, et que la majeure partie de la valeur prédictive documentée précédemment provient directement de l'inclusion des variables détaillées liées à la santé et au marché du travail.

Enfin, nous comparons les distributions de risques prédites pour deux cohortes, 2010 et 2018, afin d'examiner si la distribution du risque a évolué au cours d'une période marquée par une hausse importante des entrées en incapacité de travail. L'analyse de ces deux années permet d'appréhender dans quelle mesure la structure du risque a changé, ou non, au fil du temps. Les deux cohortes présentent une forme générale similaire, suggérant une relative stabilité de l'hétérogénéité observable sous-jacente à la persistance en DI au cours de la dernière décennie.

Comme le montre le panneau D de la Figure 6, la cohorte de 2018 présente toutefois un pic nettement plus prononcé autour d'une probabilité d'environ 0,5, tandis que la distribution de 2010 est plus lisse et légèrement plus dispersée. Cette concentration plus marquée en 2018 indique qu'une part plus importante des individus se situent autour d'un risque intermédiaire de persistance, ce qui pourrait refléter une évolution progressive des profils des personnes entrant dans l'incapacité au fil du temps. Malgré cette différence visible dans la forme de la distribution, la structure globale reste proche entre les deux cohortes, ce qui suggère que les mécanismes d'hétérogénéité capturés par le modèle demeurent en grande partie stables dans le temps.

Dans l'ensemble, ces analyses montrent que la persistance en DI évolue de manière systématique au fil des épisodes. La prédictibilité tend à augmenter après les premiers mois, conformément à la sélection dynamique parmi les sortants précoces, tandis que les horizons de prévision plus longs révèlent une hétérogénéité non observée plus importante. Des ensembles d'information riches améliorent substantiellement la discrimination, et les profils de risque semblent stables entre cohortes. Ensemble, ces résultats soulignent l'importance d'une information rapide et détaillée pour suivre les épisodes d'incapacité et identifier les individus dont le risque de dépendance à long terme devient plus prononcé au fil du temps.

### 3. Conclusion

Cet article offre une analyse approfondie des déterminants et de la prédictibilité des trajectoires d'incapacité de travail en Belgique, en utilisant des données administratives riches et des techniques modernes de machine learning. Trois enseignements principaux en ressortent.

Premièrement, la probabilité d'entrer en incapacité de travail de courte durée est fortement prévisible. Le modèle de base présente une forte capacité discriminante et explique près d'un cinquième de la variation observée des risques d'entrée, une part substantielle compte tenu de la rareté et de la complexité de l'événement. Les indicateurs liés à la santé — en particulier les hospitalisations et la consommation pharmaceutique — sont de loin les prédicteurs les plus informatifs, les trajectoires professionnelles jouant un rôle secondaire mais important. Les caractéristiques sociodémographiques contribuent peu une fois les informations de santé et d'emploi prises en compte. La distribution des risques prédits est fortement asymétrique : la plupart des individus présentent un risque quasi nul, tandis qu'un petit groupe concentre l'essentiel du risque prévisible. Ce schéma confirme le potentiel des interventions préventives ciblées.

Deuxièmement, la transition de l'incapacité primaire vers l'invalidité est nettement moins prévisible. Bien que le modèle différencie correctement les individus qui effectuent ou non la transition, seule une fraction modeste de la variance est expliquée par les caractéristiques observables. Le classement des prédicteurs change également : l'âge devient le facteur le plus important, suivi des indicateurs de santé mentale et de la dépense pharmaceutique, tandis que l'attache au marché du travail avant l'incapacité conserve un pouvoir prédictif. Ces résultats mettent en évidence une asymétrie importante : si l'entrée dans l'incapacité est largement structurée par des profils de santé et d'emploi observables, la persistance au-delà de la première année dépend davantage de facteurs difficiles à mesurer à travers des données administratives, tels que l'évolution clinique, les trajectoires de récupération ou une familiarisation croissante avec le système d'incapacité de travail.

Troisièmement, l'analyse dynamique au sein des épisodes d'incapacité révèle que la sélection dynamique joue un rôle central dans le maintien initial dans le régime. Les individus restant en IT au cours des premiers trimestres présentent des risques de sortie systématiquement plus faibles et deviennent un groupe plus homogène en termes de prédicteurs observables. En conséquence, la précision prédictive s'améliore entre le premier et le troisième trimestre. En revanche, les horizons de prévision plus longs voient leur prédictibilité diminuer, ce qui reflète l'importance croissante de l'hétérogénéité non observée à mesure que le temps passe. Les comparaisons entre différents ensembles d'information confirment que les données détaillées sur les soins de santé et les trajectoires professionnelles sont essentielles pour une prédiction pertinente, tandis que des modèles simplifiés échouent à discriminer efficacement. Enfin, la stabilité des distributions de risques entre les cohortes de 2010 et 2018 suggère que l'hétérogénéité observable sous-jacente à la persistance au sein de l'IT est remarquablement stable au fil du temps et peu influencée par les conditions macroéconomiques ou les réformes institutionnelles.

Pris ensemble, ces résultats démontrent la valeur des approches de machine learning pour caractériser les profils de risque et orienter la conception de politiques préventives. Ils

montrent que les risques d'incapacité sont fortement concentrés et, dans une large mesure, prévisibles avant même le début d'un épisode, ce qui offre des opportunités claires pour une identification précoce et un soutien ciblé. Dans le même temps, les dynamiques de persistance soulignent l'importance d'un suivi attentif au cours des premiers mois d'incapacité, lorsque les trajectoires individuelles commencent à diverger. Ces enseignements peuvent contribuer à des interventions plus proactives et mieux ciblées, améliorant in fine la gestion des risques d'incapacité de travail et la soutenabilité du système d'assurance invalidité.

## Références

- Alvarez, F., & Schimer, R. (2011). Search and rest unemployment. *Econometrica*, 79(1):75-122.
- Autor, D., & Duggan, M. (2006). The rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics*, 118(1), 157–205.
- Autor, D. H., Maestas, N., Mullen, K. J., & Strand, A. (2015). Does delay cause decay? The effect of administrative decision time on the labor force participation and earnings of disability applicants. NBER Working Paper No. 20840, National Bureau of Economic Research.
- Bruyneel, L., Rygaert, X., Oslejova, J., Avalosse, H., Fabri, V., Noirhomme, C., Willaert, D., Vrancken, J., Meeus, A., Leclercq, A., Karakaya, G., Brunois, T., Di Zinno, T., & Roelants, E. (2024). Incapacité de travail de longue durée et invalidité dues à des troubles psychosociaux – Profil socio-démographique, médical et de consommation de soins (Rapport). Agence Intermutualiste – INAMI.
- Campolieti, Michele & Riddell, Chris, 2012. "Disability policy and the labor market: Evidence from a natural experiment in Canada, 1998–2006," *Journal of Public Economics*, Elsevier, vol. 96(3), pages 306-316.
- Carey, C., Miller, N.H., & Molitor, D. (2022). Why does disability increase during recessions? Evidence from Medicare. NBER Working Papers 29988, National Bureau of Economic Research, Inc.
- Charles, K.K., Li, Y., & Stephens, M. (2018). Disability benefit take-up and local labor market conditions. *The Review of Economics and Statistics*, MIT Press, vol. 100(3), 416-423.
- Clumeck, N., Kempenaers, C., Godin, I., Dramaix, M., Kornitzer, M., Linkowski, P., & Kittel, F. (2009). Working conditions predict incidence of long-term spells of sick leave due to depression: results from the Belstress I prospective study. *Journal of Epidemiology & Community Health*, 63(4), 286–292.
- Cockx, B., Lechner, M., & Bollens, J. (2023). Priority to unemployed immigrants. A causal Machine Learning evaluation of training in Belgium". *Labour Economics*, 80, 102306.
- De Brouwer, O., Leduc, E., & Tojerow, I. (2023). The unexpected consequences of job search monitoring. *Journal of Public Economics*. 224: 104929
- De Brouwer, O., & Tojerow, I. (2023). The growth of disability insurance in Belgium: Determinants and policy implications (IZA Discussion Paper No. 16376). IZA – Institute of Labor Economics.
- French, E., & Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, 6(2), 291–337.
- Fontenay, S., & Tojerow, I. (2025). Is supported employment effective for Disability Insurance recipients with mental health conditions? Evidence from a randomized experiment in Belgium. *Journal of Health Economics*, 91, 103103.
- Holmgren, K., Fjällström-Lundgren, M., & Hensing, G. (2013). Early identification of work-related stress predicted sickness absence in employed women with musculoskeletal or mental disorders: A prospective, longitudinal study in a primary health care setting. *Disability and Rehabilitation*, 35(5), 418–426.

Institut national d'assurance maladie-invalidité (INAMI). (2018). Facteurs explicatifs relatifs à l'augmentation du nombre d'invalides : Régime des salariés et régime des indépendants. Période 2007–2016 [Study]. Brussels: INAMI.

Institut national d'assurance maladie-invalidité (INAMI). (2023). Incapacités de travail en 2023 : combien d'invalidités en raison d'une dépression ou d'un burnout ? Quel coût pour l'assurance indemnités ? INAMI. <https://www.inami.fgov.be/fr/statistiques/statistiques-indemnites/statistiques-sur-les-incapacites-de-travail-decoulant-d-un-burnout-ou-d-une-depression/incapacites-de-travail-en-2023-combien-d-invalidites-en-raison-d-une-depression-ou-d-un-burnout-quel-cout-pour-l-assurance-indemnites>

Institut national d'assurance maladie-invalidité & Agence InterMutualiste (INAMI & AIM). (2024). Incapacité de travail de longue durée et invalidité dues à des troubles psychosociaux: Profil socio-démographique, médical et de consommation de soins. INAMI. [https://aim-ima.be/IMG/pdf/rapport\\_aim\\_-\\_incapacite\\_de\\_travail\\_de\\_longue\\_duree\\_et\\_invalidite\\_dues\\_a\\_des\\_troubles\\_psychosociaux.pdf](https://aim-ima.be/IMG/pdf/rapport_aim_-_incapacite_de_travail_de_longue_duree_et_invalidite_dues_a_des_troubles_psychosociaux.pdf)

Kostøl, A. R., & Mogstad, M. (2014). How financial incentives induce disability insurance recipients to return to work. *American Economic Review*, 104(2), 624–655.

Kostøl, A. R., & Myhre, A. S. (2021). Labor supply responses to learning the tax and benefit schedule. *American Economic Review*, 111(11), 3733–3766.

Low, H., & Pistaferri, L. (2019). Disability insurance: Error rates and gender differences. *Journal of Political Economy*, 127(6), 2839–2892.

Maestas, N., Mullen, K. J., & Strand, A. (2013). Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *American Economic Review*, 103(5), 1797–1829.

Marie, O., & Vall Castelló, J. (2023). "Sick Leave Cuts and (Unhealthy) Returns to Work," *Journal of Labor Economics*, University of Chicago Press, vol. 41(4), pages 923-956.

Moreau, M., Valente, F., Mak, R., Pelfrene, E., De Smet, P., De Backer, G., & Kornitzer, M. (2004). Occupational stress and incidence of sick leave in the Belgian workforce: The Belstress study. *Journal of Epidemiology & Community Health*, 58(6), 507–516.

Nekoei, A., Salines, M., Schmieder, J., & von Wachter, T. (2025). Labor Market Effects of Social Insurance Reforms. CESifo Working Paper No. 11128.

Saks, Y. (2017). Better understanding the upward trend in the number of disability insurance claimants. *National Bank of Belgium*, issue ii, pages 55-68.

Toppinen-Tanner, S., Ojajarvi, A., Väänänen, A., Kalimo, R., & Jäppinen, P. (2005). Burnout as a predictor of medically certified sick-leave absences and their diagnosed causes. *Behavioral Medicine*, 31(1), 18–32.

**Annexe technique en anglais accompagnant le résumé  
« Prédire l'entrée dans l'assurance invalidité et la dépendance  
à long terme : Une approche par Machine Learning »**

Guida Ayza Estopà & Ilan Tojerow (DULBEA, ULB)

Novembre 2025

## Technical appendix

This Technical Appendix accompanies the main French report for the study “Predicting Disability Insurance Entry and Long-Term Dependence: A Machine Learning Approach” and provides additional methodological detail, supplementary results, and a comprehensive description of the data and analytical framework used throughout the study. Its purpose is to offer full transparency on the empirical approach, the construction of variables, and the statistical models underlying the findings presented in the main document.

Using rich Belgian administrative data (2006–2019) and ensemble machine-learning models, we predict (i) entry into work incapacity, (ii) transitions to long-term disability, and (iii) early-spell exit dynamics. Results show that work-incapacity entry is highly predictable, mainly driven by health-related indicators, while transitions to long-term disability are less so and reflect greater unobserved heterogeneity. Dynamic analyses reveal strong early selection effects and broadly stable risk distributions across cohorts.

The appendix is structured into four parts.

**Section A** presents additional figures and tables that complement the descriptive evidence and empirical results discussed in the report.

**Section B** provides detailed information on the data sources, sample construction, variable definitions, and the institutional context of work incapacity and disability insurance in Belgium.

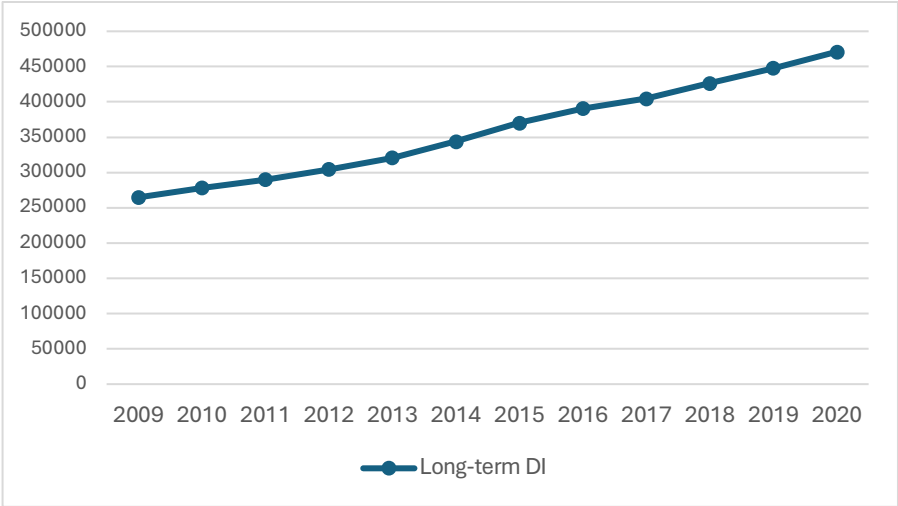
**Section C** outlines the conceptual framework used to interpret the dynamics of work incapacity, clarifying the role of heterogeneity, dynamic selection, and duration dependence.

**Section D** documents the prediction methodology in detail, including model training, variable selection, calibration procedures, and performance evaluation.

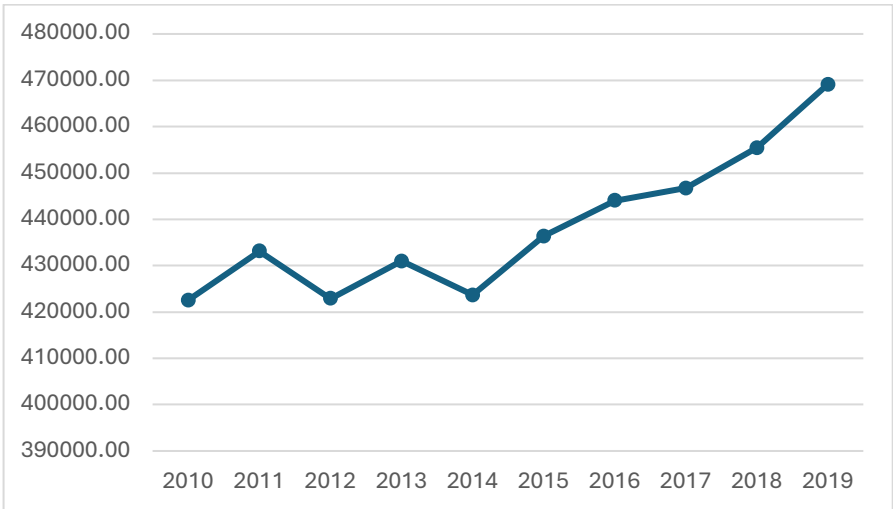
# A. Additional Figures and Tables

**Figure A1: Long and short-term DI evolution in Belgium**

**Panel A: Long-term DI cases**

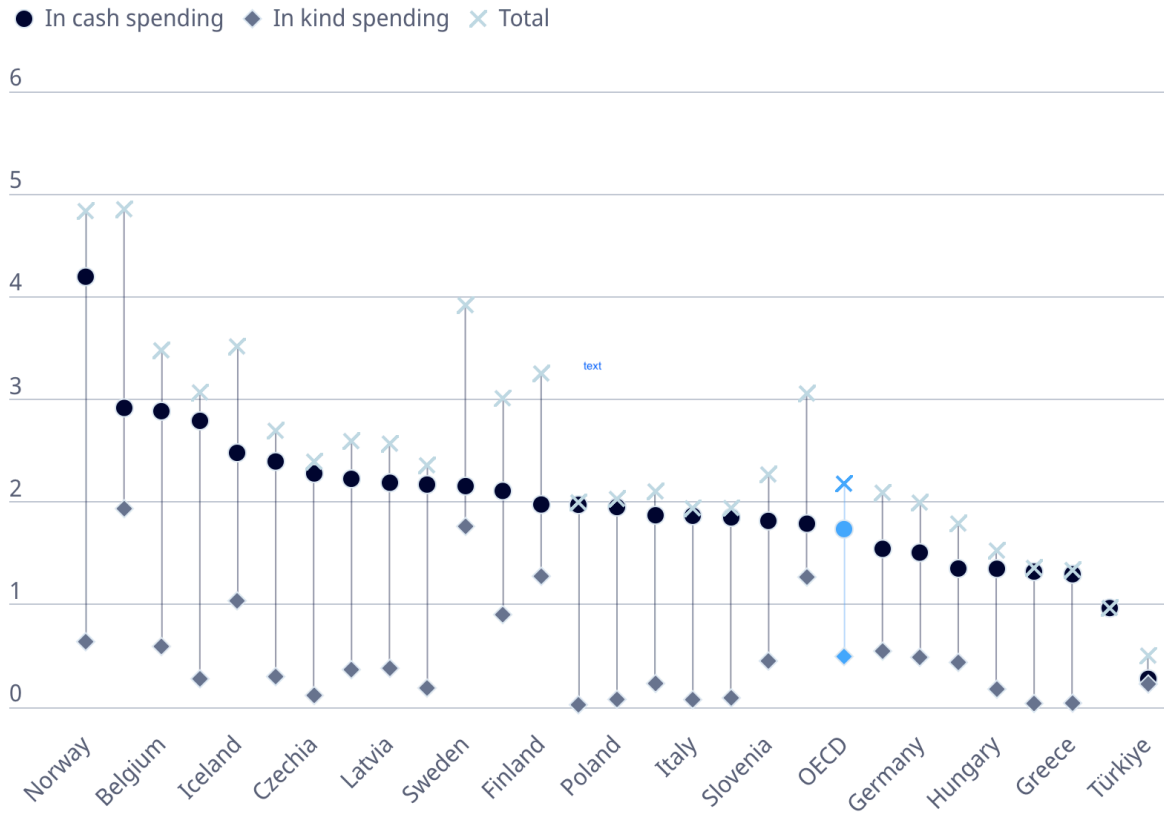


**Panel B: Net entries on short-term DI**



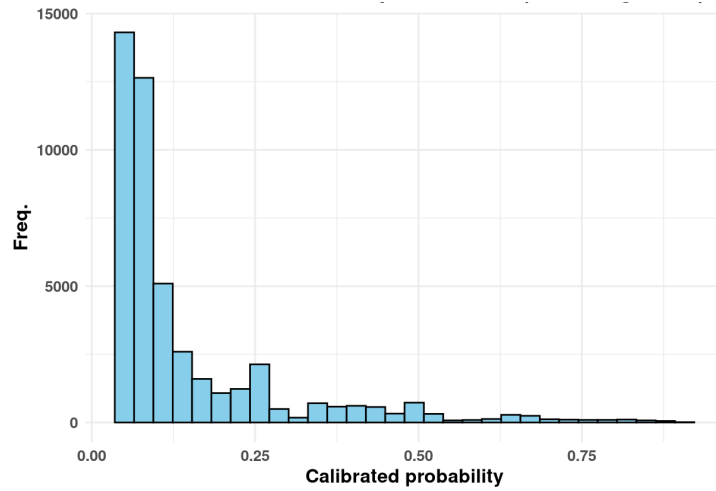
*Note: self-constructed graphics with data from the National Institute for Health and Disability Insurance (NIHDI; INAMI in French). Panel A represent the total number of individuals on the long-term disability insurance program (more than 1 year) while panel B represents the net entries; the number of people starting a DI spell on that year.*

**Figure A2: Public spending on incapacity in 2020**



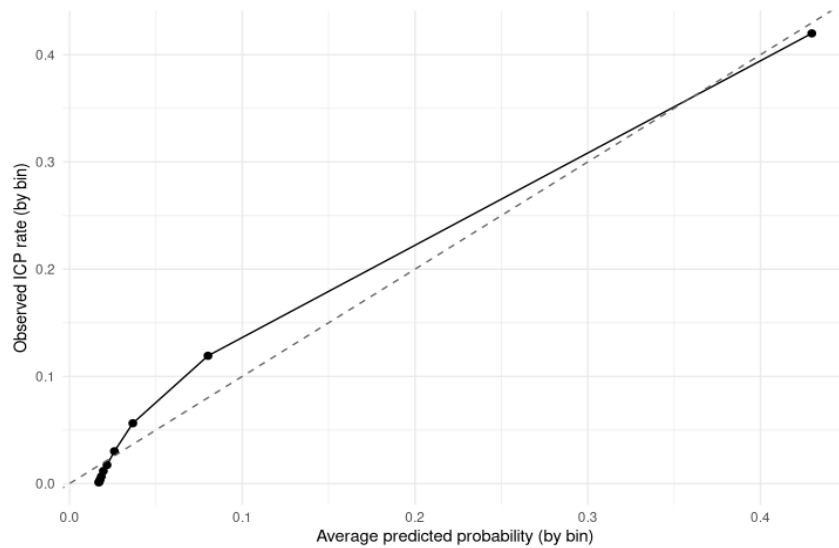
Note: Public spending on incapacity in % of GDP in the OCDE countries in 2020, it includes disability payments in cash, in kind and the sum of both. Data comes from OCDE statistics, 2020.

**Figure A3:** Distribution of predicted transition probability (excluding probabilities <0.05)



*Note: The plot presents the distribution of the predicted work incapacity entry probability in the hold-out sample for the year 2018, excluding all the values lower than 0.05, to improve visual interpretation of the rest of the distribution.*

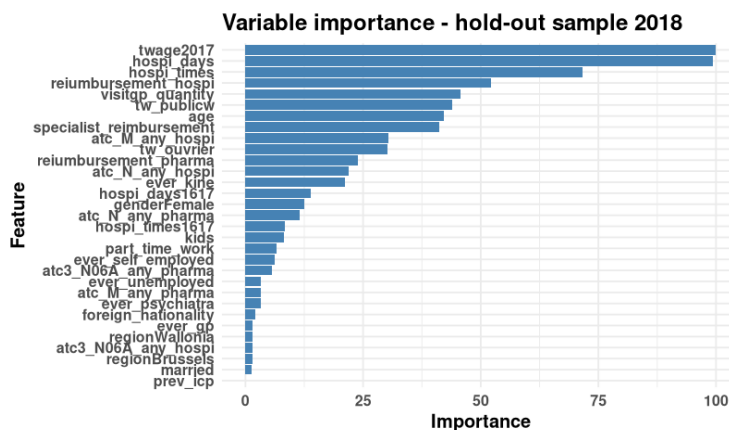
**Figure A4:** Comparing predictions from the baseline model to outcomes in 2018



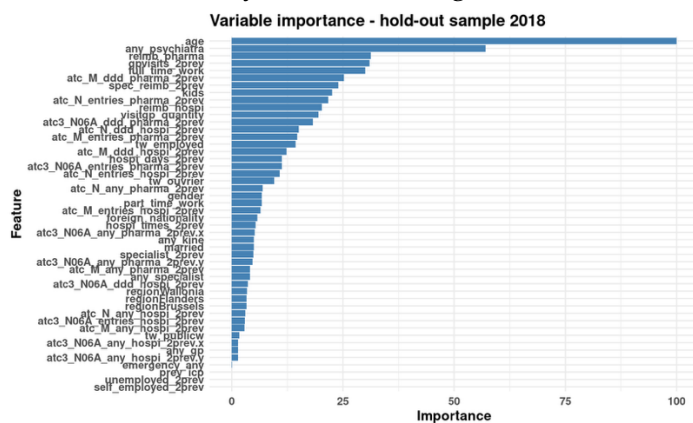
*Note: Panel A shows a calibration plot comparing predicted and observed work incapacity entry in the hold-out sample for 2018, using all the individuals in our sample, without excluding the ones with previous work incapacity spells. Individuals are grouped into 10 deciles based on their predicted work incapacity entry probability. For each bin, we plot the average predicted probability against the observed work incapacity entry rate. The dashed line represents perfect calibration (i.e., predicted = observed). Because the plot uses decile averages, the values do not span the full 0–1 range even though individual predictions do.*

**Figure A5: Variables importance**

Panel A: Work incapacity entry risk (excluding individuals with a previous DI spell)



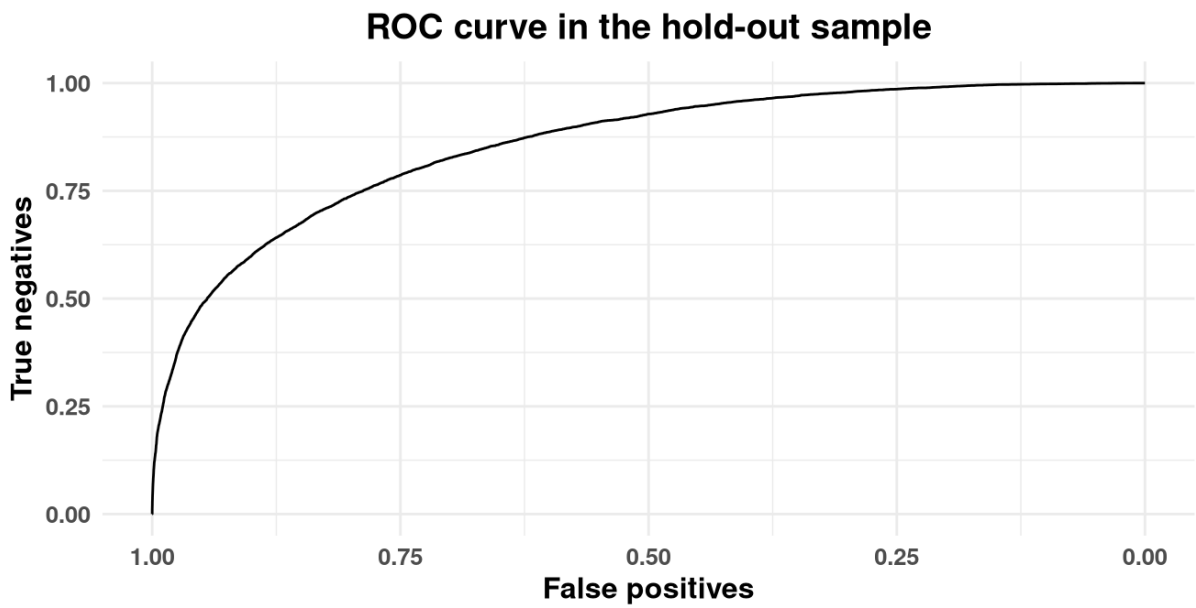
Panel B: Probability of transitioning from ST- to LT-DI



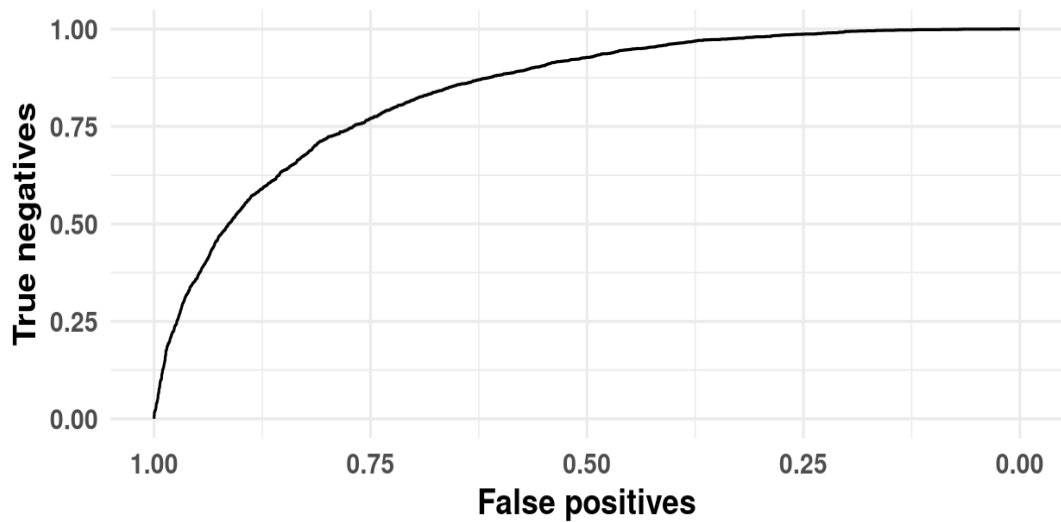
Note: Contribution of each variable to the predictive power of the three different ensemble models for 2018.

**Figure A6: ROC curves**

Panel A: Outcome 1 (probability of entering DI)



Panel B: Outcome 2 (probability of transitioning from ST- to LT-DI)



*Note: Receiver operating characteristic (ROC) curves of the ensemble model for outcomes 1 & 2 evaluated in the hold-out sample for 2018.*

**Table A1:** AUC and R<sup>2</sup> for the models in Figure 6

Panel / Model	Specification	AUC	R <sup>2</sup>
<b>Panel A – Dynamic selection</b>	0 → 2 quarters (baseline)	0.6410	0.0558
	2 → 4 quarters (conditional on staying)	0.6502	0.0855
<b>Panel B – Forecasting horizons</b>	1-quarter horizon	0.5765	0.0150
	2-quarter horizon	0.6410	0.0558
	3-quarter horizon	0.6565	0.0600
	4-quarter horizon	0.6265	0.0209
<b>Panel C – Information sets</b>	Reduced model (basic variables only)	0.5798	0.0209
	Full baseline model	0.6410	0.0558
<b>Panel D – Cohorts</b>	Cohort 2010	0.6524	0.0653
	Cohort 2018	0.6410	0.0558

*Note: These metrics correspond to the predictive performance of the models illustrated in Figure 6. For each panel, the baseline case—representing the predicted two-quarter exit probabilities at the start of the spell—is reported again to facilitate comparison across specifications; this baseline distribution is also shaded in grey. The AUC quantifies the model’s ability to discriminate between individuals who exit and those who remain in work incapacity, while the R<sup>2</sup> captures the proportion of variation in predicted exit risks explained by observable characteristics.*

## B. Data and Institutional context

The present analysis relies on a rich administrative dataset combining longitudinal records on labor market history, healthcare utilization, and sociodemographic characteristics from two main sources: the Data Warehouse of the Social Security (BCSS-DWSS) and the Inter-Mutualist Agency (IMA-AIM) in Belgium. The data span the period from 2006 to 2019 and cover 10% of the Belgian working-age population, which consist of 735,000 individuals with quarterly observations, totaling around 40 million records. Data from different sources are linked using an anonymized social security identifier. Our unit of observation is the individual-quarter, which allows us to construct quarterly spells of short-term DI and to observe transitions to long-term DI at the 12-month threshold.

We use several datasets from the BCSS, which is a central data system managed by the Belgian government that integrates individual-level administrative records across social security, healthcare, and employment domains. Our main variables of interest are those related to DI. In this context, we have data on short- and long-term disability spells, as well as related information such as the type of benefits, household composition, and, in the case of long-term DI, the pathology that led to the disability. In Belgium, workers who have contributed sufficiently to the social insurance system are eligible for DI if they are unable to work for health-related reasons, regardless of their employment status at the onset (employed or unemployed). During the first month of sickness, white-collar workers receive full salary financed by the employer, whereas for blue-collar workers, the employer covers part of the salary, and the rest is paid by the National Institute for Health and Disability Insurance (NIHDI). From the second month onwards, all benefits are paid by the individual's health insurance fund (mutuality). The replacement rate depends on prior employment status and declines with the duration of the sickness spell, starting at 60% of gross salary for most workers, subject to income-dependent caps and floors.

Sick leave can be prescribed by any treating medical practitioner from the first day of illness. To qualify, three conditions must be met: (i) the individual must cease all productive activity; (ii) the cessation must be due to a deterioration in health unrelated to professional activity; and (iii) work capacity must be reduced by at least 66% relative to their previous occupation. After one month, an advisory physician evaluates eligibility for short-term DI. After one year, the mutuality doctor may propose a transition to the long-term DI scheme, which requires a reassessment by a certified medical advisor. Long-term DI benefits are also financed by the NIHDI and differ mainly in how residual work capacity is assessed and in the replacement rate, which increases to 65%, adjusted for household composition. We define two main outcome variables: starting a DI spell and transitioning from Short-term DI to LT-DI. We identify an individual as starting a DI spell when they start receiving benefits from their mutuality, so just after the first month which is covered by the employer, and transition to a

LT-DI when they start receiving the other type of benefits because they already spent one year disabled. In addition to modelling entry and transition probabilities, we later exploit the longitudinal nature of the dataset to track how individuals' status evolves within short-term DI spells. This allows us to study the probability of remaining in short-term DI after one, two, three, or four quarters, which forms the basis of the dynamic analysis. In this case, the outcome is defined as a dichotomous indicator equal to one if the individual remains in short-term DI for the corresponding number of consecutive quarters.

As predictor variables, we first include those related to the labour market, which are also drawn from the BCSS. These include worker type (blue- or white-collar; public or private sector), working time (part- or full-time), self-employment status, unemployment spells, and income data categorized into normalized wage brackets. In Belgium, salaried workers are eligible for unemployment insurance (UI) following involuntary job loss, conditional on a sufficient employment history. Uniquely, UI benefits can be received for an indefinite period, provided the claimant remains available for the labour market and complies with job search obligations and reintegration plans. The benefit amount decreases progressively over time, depending on past earnings, family situation, and unemployment duration. A minimum benefit is guaranteed, but non-compliance may lead to benefit suspension or reduction. Self-employed individuals are subject to different eligibility rules for both DI and UI. The BCSS data also include sociodemographic information such as age, gender, nationality, household composition, and region of residence, that are also included as predictor variables.

The second main data source is the IMA-AIM, which collects and harmonizes individual-level healthcare and reimbursement data from all Belgian mutual insurance funds. In Belgium, all residents must be affiliated with a mutuality, which acts as an intermediary between individuals and the compulsory public health insurance system. Mutualities, apart from the payment of health-related benefits, also manage the reimbursement of medical expenses. Although they are private entities with voluntary affiliation, they operate under a public and regulated framework.

From the IMA, we obtained data on reimbursements—hence consumption—of prescription drugs (whether purchased in public pharmacies or administered in hospitals) and other health expenditures such as general practitioner and specialist visits, as well as hospital stays. Drug information is categorized using the ATC (Anatomical Therapeutic Chemical) classification. At its first level, the ATC identifies the anatomical system targeted by the drug. We focus in particular on drugs affecting the nervous and musculoskeletal systems, which include medications for mental health and musculoskeletal disorders respectively, the two main conditions leading to DI spells. For these two categories, we have more granular data up to ATC level 3, which allows us to distinguish, for instance, between antidepressants and antipsychotics within the nervous system category.

Table 1 presents the full set of variables used in the baseline prediction model, which are generally available for all quarters in the sample. All predictor variables are measured prior to the predicted event—that is, before the onset of the DI spell or the observed transition to LT-DI spell. For historical variables, we use a two-year lookback window from the year of analysis (e.g., the 2018 unemployment history variable indicates whether the individual was unemployed at any point since 2016).

The model includes three main groups of predictors. First, sociodemographic variables: gender, age, marital status, number of children, nationality of origin, region, and, where available, district of residence. Second, health-related variables, such as whether the individual had any visits to a general practitioner or specialist in the previous two years, as well as the number of such visits. Given the prevalence of mental health and musculoskeletal disorders among DI recipients, we also include visits to psychologists, psychiatrists, and physiotherapists. In addition, we consider the number of hospitalization spells and total days spent in hospital, allowing us to distinguish between short recurrent hospitalizations and prolonged stays. Third, we incorporate pharmaceutical variables, focusing on drug consumption related to mental health and musculoskeletal conditions. Specifically, we include indicators for any drug belonging to the ATC level 1 categories corresponding to the nervous system and musculoskeletal system, respectively. In addition, we separately identify the use of antidepressants (ATC level 3). For each of these drug groups, we distinguish whether the drug was purchased in a public pharmacy or administered in a hospital. For all variables, we record both whether the drug was taken at least once and the total quantity consumed. Finally, we include labor market variables to assess the role of employability on the probability of being on DI. These include unemployment or self-employment status in the two years prior to the disability spell, working time (full- or part-time), type of occupation (blue- or white-collar), employment sector (public or private), and labor income. Income is reported in normalized brackets and refers to total earnings in the previous completed calendar year; while less precise than raw earnings, it allows for meaningful comparisons across individuals.

We also include an indicator for whether the individual had any prior disability spells. This variable is used both as a predictor and, in an alternative specification, to restrict the sample to individuals with no previous DI episodes, in order to analyse the determinants of first-time entries into DI.

Table 2 presents descriptive statistics for the overall working-age population and for the subpopulation of individuals currently on DI. In the full sample, the gender, nationality, and age distributions are balanced, with 50.1% women, 22.2% foreign nationals, and a mean age of 40.9 years. Regarding mental health–related healthcare use, 34.4% of individuals have at some point purchased medication in a public pharmacy and 7.5% have received such medication in a hospital. General practitioner visits are nearly universal (95.1%), and more than half of the sample (51.6%) has experienced at least one hospital stay. In terms of labour market histories,

6.8% of individuals had an unemployment spell, 6.5% ever entered DI, and 2.7% transitioned to long-term disability.

The DI subpopulation displays broadly similar demographic characteristics, though with slightly higher mean age (44.8 years) and a lower share of women (40.5%). Differences are more pronounced in healthcare use: 49.9% have purchased medication in a public pharmacy and 13.7% have received it in a hospital, with almost all individuals having visited a general practitioner—reflecting their central role in issuing DI certificates—and a large majority having experienced a hospital stay (78.0%). Labour market differences are also notable: unemployment spells are nearly twice as frequent among individuals on DI (11.4% compared to 6.8% in the overall population). 7.4% of this group transitioned to long-term disability.

## C. Conceptual framework

We aim to explain why the number of people on DI has increased so much on the last decade. With the models used in this paper, we are going to study which family of factors is better at predicting the entrance into DI. In a second exercise we are going also to understand the predicting factors for being long term disabled (to transition from short term DI to long term DI). Identifying these micro-level determinants may help us understand the mechanisms behind the aggregate trend.

Following Mueller & Spinnewijn, we present a conceptual framework to account for heterogeneity in DI risk, the dynamics of the probabilities of being on DI, and duration dependence. Their work builds on the unemployment benchmark, where job-finding dynamics are well established. In our case, we need to build the link between individual characteristics—socioeconomic, labor-related, and health-related—and the probability of transitioning into DI.

We first describe heterogeneity in the initial DI risk and then turn to dynamic selection, duration dependence, and other factors that may influence DI hazards.

### C.1. Heterogeneity on the initial DI risk:

The first step is to define how different sources of heterogeneity, both observed and unobserved, may influence DI entry rates as well as the probability of remaining in DI for a long period. De Brouwer & Tojerow (2023) offer an extensive analysis of DI determinants in Belgium, showing that changes in observable characteristics such as age and work type only marginally account for the increase in the long term. Heterogeneity in DI risk arises from differences in observable characteristics that affect baseline DI risk—such as age, gender, income, labor market position, and health status (which we proxy through dimensions of healthcare use)—as well as unobservable traits such as resilience, health-seeking behavior, and moral hazard.

It is widely accepted that age, gender, and income play an important role in the probability of starting a DI spell (INAMI, 2028; De Brouwer & Tojerow, 2023). Health is by definition negatively correlated with DI entry, since eligibility requires reduced work capacity due to a health condition. At the same time, certain work characteristics can make continued employment feasible despite health limitations, while others may increase the likelihood of health deterioration and DI entry—for example, stress or burnout (Moreau et al., 2004; Toppinen-Tanner et al., 2005; Holmgren et al., 2013; INAMI 2023, INAMI-AIM, 2024).

The literature has focused strongly on ageing and the increase in women's employment, but we consider a broader set of observables grouped into three domains: sociodemographic factors, labor-related factors, and health-related factors. Whereas health is often proxied simply by age, we use healthcare consumption to provide a more direct measure of health status.

However, we are also aware of the existence of unobservable heterogeneity; individual resilience, health-seeking behavior, moral hazard, and work preferences affect DI risk beyond what can be explained by observables. Moral hazard here refers to the behavioral response to the incentives created by the DI system, such as a greater tendency to apply for benefits or a lower probability of returning to work when benefits reduce the financial cost of non-employment (Autor & Dugan, 2003; Maestas et al., 2013; French & Song, 2014; Kostøl & Mogstad, 2014).

**Formalization:**

Formally, we write the individual DI entry probability as

$$D_{i,t} = D_t(X_i) + \varepsilon_{i,t}$$

where  $D_t(X_i) = E_t(D_{i,t}|X_i)$  is the individual DI entry probability based on observable characteristics  $X_i$  in time  $t$ , and  $\varepsilon_{i,t}$  captures unobservable characteristics, assumed orthogonal to the observables:

$$E_t(\varepsilon_{i,t}|X_i) = 0$$

The set of observable characteristics is:

$$X_i = \{S_i, L_i, H_i\}$$

where  $S_i$  refers to sociodemographic variables,  $L_i$  labour market characteristics, and  $H_i$  health-related factors.

We cannot observe individual DI entry probabilities directly, but we do observe whether an entry occurs. The realization of the probability is:

$$R_{i,t} = \begin{cases} 1 & \text{with probability } D_{i,t}, \\ 0 & \text{otherwise,} \end{cases} \quad D_{i,t} \in (0,1)$$

Thus, our prediction model estimates the probability of  $R_{i,t} = 1$  based on  $X_i$ :

$$R_{i,t} = R_t(X_i) + e_{i,t}$$

where  $e_{i,t}$  is the prediction error. If the prediction model is unbiased, then

$$E_t(\hat{R}_{i,t}|X_i) = D_t(X_i)$$

**Proposition 1. Lower bound on heterogeneity**

We are first interested in quantifying the extent of observable heterogeneity in DI entry. In practice, the explanatory power of the prediction model is summarized by the  $R^2$ , which measures the share of the variance in realizations  $R_{i,t}$  that can be explained by observables  $X_i = \{S_i, L_i, H_i\}$ .

Formally, for a hold-out sample,

$$R^2 = 1 - \frac{\sum_i (R_{i,t} - \hat{R}_{i,t})^2}{\sum_i (R_{i,t} - \bar{R}_{i,t})^2}$$

where  $\hat{R}_{i,t} = R_t(X_i)$  is the predicted probability of entry and  $\bar{R}_{i,t}$  the sample mean.

This measure represents a lower bound on heterogeneity because it captures only the variation explained by the included observables. Unobservable factors  $\varepsilon_{i,t}$  remain outside the model, so the  $R^2$  cannot capture the full extent of heterogeneity in DI entry.

By comparing  $R^2$  across different sets of variables (e.g. only sociodemographics, then adding labour, then adding health-related variables), we can document how each block contributes to explaining DI risk. In particular, healthcare consumption is expected to be especially informative, as it proxies underlying health status much more directly than age or other demographic controls.

**Proposition 2. Persistent heterogeneity**

We next ask whether heterogeneity is persistent (time-invariant) or transitory (time-varying). Persistence would indicate that certain risk factors (e.g. gender, education, permanent health conditions) produce stable differences in DI risk across years, while transitory components would capture factors that change over time, such as health shocks, business-cycle conditions, or short-term institutional effects.

Formally, let the variance of the individual DI risk at time  $t$  be:

$$Var_t(D_{i,t}) = Cov_t(D_{i,t}, D_{i,t'}) + Cov_t(D_{i,t}, D_{i,t} - D_{i,t'}),$$

where  $t'$  refers to another period.

- The first term on the right side of the equation captures the persistent component, i.e. the part of the variance that remains stable across periods.

- The second term captures the transitory component, i.e. the part of the variance that changes between periods.

A high covariance indicates that the same individuals are consistently predicted at high (or low) risk across periods, i.e. persistent heterogeneity. A low covariance would suggest that heterogeneity is mainly transitory.

Equivalently, we can compute the cross-year  $R^2$ : the explanatory power of predictions from year  $t'$  when applied to realizations in year  $t$ . This corresponds to:

$$R_{t,t'}^2 = \text{Corr}(D_t(X_i), D_{t'}(X_i))^2$$

which measures how stable predicted risks are over time.

In practice, this involves estimating prediction models year by year and then examining how much predictive power transfers across years.

## C.2. Selection on the probability to be on Long-Term DI

We next focus on the second outcome of interest: the probability that individuals in short-term DI transition to long-term DI after twelve months. Formally, for the set  $S_t$  of individuals in ST-DI at year  $t$ , we define:

$$E_{i,t+1} = \mathbf{1}\{\text{transition to LT-DI at } t + 1\}, \quad i \in S_t$$

Where  $\mathbf{1}\{\cdot\}$  denotes the indicator function, equal to 1 if the condition is satisfied and 0 otherwise, and estimate the transition probability conditional on observables at  $t$ :

$$D_{i,t+1} = D_{t+1}(X_{i,t}) + \varepsilon_{i,t+1}, \quad \mathbb{E}[\varepsilon_{i,t+1}|X_{i,t}] = 0$$

Our prediction model provides  $\hat{p}_{i,t+1} = \hat{D}_{t+1}(X_{i,t})$ , and predictive performance in a hold-out sample is summarized by

$$R_{t \rightarrow t+1}^2 = 1 - \frac{\sum_{i \in S_t} (E_{i,t+1} - \hat{p}_{i,t+1})^2}{\sum_{i \in S_t} (E_{i,t+1} - \bar{E}_{t+1})^2}$$

This captures how much of the heterogeneity in the ST-DI to LT-DI transition can be explained by observables. A higher  $R^2$  indicates that the pool of individuals reaching the 12-month threshold is increasingly composed of systematically high-risk individuals (individuals with poorer health, weaker labor market attachment, or stronger attachment to the DI scheme), consistent with dynamic selection.

## C.3. Duration dependence vs. dynamic selection

Finally, we distinguish between true duration dependence and dynamic selection in the evolution of DI spells. Duration dependence refers to changes in an individual's hazard of exit

or transition as the spell lengthens, while dynamic selection arises because, over time, those who remain in DI are disproportionately individuals with lower exit probabilities.

Formally, let  $h_i(d)$  be the exit probability of individual  $i$  from ST-DI at elapsed duration  $d$ . The observed hazard at duration  $d$  is the average across those still at risk,

$$\bar{h}(d) = \mathbb{E}[h_i(d)|i \in S_d]$$

The change in the observed hazard between  $d$  and  $d+1$  can be decomposed into:

$$\bar{h}(d+1) - \bar{h}(d) = \mathbb{E}[\bar{h}_i(d+1) - \bar{h}(d) | i \in S_{d+1}] + (\mathbb{E}[h_i(d)|i \in S_{d+1}] - \mathbb{E}[h_i(d)|i \in S_d])$$

The first term is the duration dependence, and it captures genuine changes in individual hazards as the spell progresses—for instance, because health deteriorates, labor market attachment weakens, or adaptation to DI increases. The second term reflects the dynamic selection: individuals with higher exit probabilities tend to leave earlier, so those who remain are increasingly concentrated among the hardest-to-exit cases.

Dynamic selection and duration dependence represent two distinct channels through which persistence in DI spells may arise. Dynamic selection reflects underlying heterogeneity: individuals differ in their exit hazards from the moment they enter DI, and those with higher hazards tend to leave early, progressively concentrating the remaining population among those with intrinsically lower recovery prospects. Duration dependence instead captures a causal effect of time spent in DI on subsequent exit probabilities, whereby remaining longer in the program reduces the likelihood of exit even for individuals with similar observables, potentially due to health deterioration, learning about the DI system, or increasing distance from employment. These mechanisms have opposite empirical implications: dynamic selection leads to narrower risk distributions and improved predictive accuracy as spells lengthen, while duration dependence generates patterns that observables cannot explain and therefore tends to reduce predictability.

Empirically, we distinguish duration dependence from dynamic selection by examining how predicted exit hazards and predictive accuracy evolve as DI spells progress. If duration dependence is present, individual exit hazards would decline with elapsed duration even after conditioning on observables, reflecting mechanisms such as clinical deterioration, psychological adaptation, or increasing detachment from the labor market. Such dynamics would make exit risks evolve in ways that the model cannot capture, thereby reducing predictive accuracy. In contrast, if dynamic selection dominates, individuals with higher exit hazards leave DI early, and those who remain over time become a more homogeneous group with systematically lower hazards. Under this mechanism, the distribution of predicted risks becomes more concentrated as spells lengthen, and predictive performance tends to improve

because observable characteristics become more informative for those who remain. To assess the relative importance of these two forces, we study how predicted exit risks and standard predictive metrics such as the AUC and  $R^2$  behave at different elapsed durations and forecasting horizons.

## D. Prediction Model: details on the methodology

For the empirical analysis we employ standard Machine Learning (ML) techniques, training a prediction model on a training sample and then evaluating the predictive power in a hold-out sample. The main problem in all prediction exercises is the trade-off between improving the prediction model and overfitting it when including too many variables. ML methods and the separation of the two samples help to optimize variable selection and to deal with the overfitting problem in a data-rich environment. We focus on two outcomes: (1) the probability of entering DI, and (2) the likelihood of transitioning from short-term to long-term DI. We define these probabilities as the risk variables for our model. To further understand the dynamics for the first entry on DI, we exclude from the baseline sample individuals who had an ICP episode in the previous two years. We start using three ML models: Random Forest, Gradient Boosted Regression Trees and LASSO in the baseline model for the year 2018, after analyzing their discriminatory power and overall performance in the training sample, we decide to keep only the two first models and combine them in an Ensemble Model, which is a linear weighted combination of them. These models take different approaches for the selection of variables but also allow differently for nonlinearities and interactions between these variables. Random Forest tends to be more robust to noise and provides stable predictions across many weakly correlated trees, while Gradient Boosted Trees sequentially focus on harder-to-predict cases and typically achieve higher accuracy by minimizing residual errors. Combining both leverages the strengths of each method.

We divide the sample in a 60% training sample and a 40% hold-out sample. The first step of the prediction process, after preparing the data and selecting the variables, is tuning key parameters for all the prediction models, we follow standard practice in machine learning and do it by 3-fold cross-validation. We then estimate the different models separately, obtain the Ensemble Model and calibrate the probabilities for each outcome. Later on, we apply this to different years to analyze time differences.

For the tuning process, we use the 15% of the sample to optimize, among other features, the minimal node size and the number of variables used at each node for the Random Forest model and the learning rate for the Boosted Regression Trees. We run and compare different alternatives, and we finally choose the one that optimizes the area under the receiver operating characteristic curve (ROC-AUC), which is also a standard practice in ML. Within this process, we run the models several times adapting the parameters to the ones that result in a better performance, this includes changing the hyperparameters of the algorithm, the sampling technique, the validation process and the set of predictor variables. It also involves a deep study of the contribution of the variables and their possible interactions between them. Once decided the best tuning parameters, the three models are estimated using a 30% of the sample not used before. As a third step in the prediction model, we use 7.5% of the sample to obtain the Ensemble Model. Instead of a simple weighted combination, we apply a stacking approach, where a logistic regression model learns to optimally combine the predictions from

the random forest and the gradient boosting regression trees. The probability we get from the Ensemble Model can be defined as:

$$p_{EnsembleModel} = \hat{\beta}_{RF}\hat{p}_{RF} + \hat{\beta}_{GB}\hat{p}_{GB}$$

where  $\hat{p}_x$  is the prediction from algorithm  $x$  and  $\hat{\beta}_x$  is the associated weight. Finally, we calibrate the raw predictions get from the ensemble model to the actual observed probabilities by estimating a linear spline in a different 7.5% of the sample. This flexible functional form allows for piecewise linear adjustments to better align predicted and observed risks. After these steps, we evaluate the final model on a hold-out sample, which represents 60% of the data. This sample has not been used in any previous step, ensuring an unbiased assessment of the model's performance. The main results we present correspond to this hold-out evaluation for the year 2018. These prediction models are not designed for causal inference but to quantify the share of systematic variation in DI outcomes that can be captured by observables (i.e., the degree of observable heterogeneity).

### **Assessing the model:**

To evaluate the accuracy of our prediction model, we compare predictions and outcomes in the hold-out sample for the year 2018 in Figure 1 of the main French report. Panel A displays results for outcome 1, the probability of entering DI. Individuals are grouped into 10 equally sized bins based on predicted risk. For each bin, we plot the average predicted probability against the observed DI entry rate. The dashed 45-degree line indicates perfect calibration. The points lie close to this line, indicating that the model's predictions are well aligned with actual outcomes. Panel A of Figure 3 shows analogous results for outcome 2, the probability of transitioning from short-term to long-term DI. As with outcome 1, predictions track observed rates reasonably well.

We use our prediction model to assess the probability of entering DI and, conditional on entry, the probability of transitioning to long-term DI. Panel B of both figures display the distribution of calibrated predicted probabilities in the 2018 hold-out sample. In both panels, most of the predicted probabilities remain below 0.5. This upper bound reflects the underlying incidence of DI events, which is relatively low in the population: only a limited share of workers ever enter DI. Figure 1 shows the distribution for DI entry among the full population. The probabilities are highly concentrated near zero, reflecting the fact that only a small share of individuals actually enters DI. This concentration illustrates the challenge of predicting a relatively rare event. Consequently, even those at highest predicted risk face probabilities well below one. Far from indicating poor performance, this pattern highlights the model's ability to capture meaningful variation in risk within the empirically relevant range. The close alignment between predicted and observed rates suggests that the model is well calibrated and provides reliable risk stratification despite the inherently low baseline probabilities. Within this empirically relevant range, the model captures substantial variation in risk and shows close alignment between predicted and observed rates, indicating good calibration and

reliable predictive performance. Figure 3 shows the corresponding distribution for transitions to long-term DI among those who already entered DI. Again, the mass of the distribution lies at the lower end, consistent with the relatively low incidence of long-term transitions.

Despite the rarity of both outcomes, the model performs well when evaluated on predictive accuracy. Figure A6 in Appendix A reports the Receiver Operating Characteristic (ROC) curves. The ROC curve contrasts the true-positive rate with the false-positive rate at different thresholds for classifying predicted probabilities into binary outcomes. The area under the curve (AUC) equals 0.86 for DI entry and 0.72 for long-term transitions, compared to a benchmark of 0.5 for random guessing and 1 for perfect prediction. These values indicate excellent discriminatory power for DI entry and solid performance for transitions to long-term DI.

As an additional measure, we compute the R-squared, which equals 0.189 for DI entry and 0.073 for long-term transitions. While these values may seem low, this is expected in models with binary outcomes, where the dependent variable is a random realization of an underlying probability. In this context, the reported values still indicate that the model captures a substantial share of the systematic variation in DI risks.

#### **Dynamic prediction models:**

Building on the baseline setup, we estimate a set of dynamic prediction models to study how predictive performance evolves as short-term work-incapacity spells unfold. Following the intuition in Mueller and Spinnewijn (2023), we examine how exit risks and their predictability change with elapsed duration and with different forecasting horizons. In this part of the analysis, the outcome of interest is not entry into DI but the probability of exiting short-term DI within a specified horizon.

We first estimate models at different durations of the spell (e.g., at the start of the spell and after two quarters) to assess whether predictive accuracy improves or deteriorates as individuals remain longer in DI. This allows us to infer the relative importance of dynamic selection versus duration dependence: increasing predictability with elapsed time is indicative of stronger dynamic selection, as individuals with higher exit hazards tend to leave early, leaving a more homogeneous population with respect to observed characteristics.

We then evaluate models with alternative forecasting horizons; predicting the probability of exit within one, two, three, or four quarters, while holding constant the information set available at the start of the horizon. Comparing predictive accuracy across horizons provides insight into how informative observables remain when forecasting persistence over different time frames.

All dynamic models use the same set of predictors as the baseline model, measured up to the relevant point in time, and are evaluated on independent hold-out samples. Together, these exercises characterise how the predictability of DI persistence evolves both with spell duration

and with the length of the prediction window, shedding light on the mechanisms shaping early patterns of return to work.

### **Predicting variables:**

To understand how different groups of variables contribute to predictive performance, we estimate a sequence of nested models that progressively incorporate additional information. Prior work has shown that sociodemographic characteristics such as age or gender correlate with DI risks, yet these factors explain only a small share of the overall variation in disability reciprocity (De Brouwer and Tojerow, 2023). This suggests that broader and more detailed sources of heterogeneity must be taken into account when analysing DI entry and persistence. A key strength of our setting is the richness of the administrative data, which allows us to combine sociodemographics with labour-market histories and multiple dimensions of healthcare utilisation. Table 1 summarises the variable groups used across the models, and Table 3 reports the predictive accuracy of the nested specifications. Consistent with the literature, sociodemographics alone provide limited predictive power, while adding labour-market information leads to a substantial improvement. The largest gains arise when health-related variables are included, underscoring their central role in structuring work-incapacity risks.

In the machine-learning models, we use variable-importance measures to assess the marginal contribution of predictors to model performance. These measures, shown in Figure A5 of the Appendix, capture both non-linearities and interaction effects, which are often substantial for health-related variables (e.g. hospitalisations or pharmaceutical expenditures). However, they do not have a causal interpretation, nor do they provide easily interpretable linear associations. For this reason, and following the approach of Mueller and Spinnewijn (2023), we complement the ML variable-importance results with a standardized OLS regression of the calibrated predicted probabilities on the underlying predictors. This regression, reported in Figures 2 and 4 for the first two outcomes, expresses coefficients in standard-deviation units and offers a transparent measure of how each observable characteristic is linearly associated with the predicted risk. Unlike the ML importance metrics, the OLS coefficients summarise only linear relationships and do not reflect interactions or non-linear effects. As a result, the two approaches naturally yield different rankings of predictors, but together they provide a comprehensive and interpretable picture of the observable factors shaping work-incapacity risks.