



CIRANO

Allier savoir et décision

TRAJECTOIRES DES GRANDS UTILISATEURS DE SOINS DE SANTÉ AU QUÉBEC

MAUDE LABERGE
BILE YACOUBA DJEDOU
THOMAS PODER
ANAÏS LACASSE
CATHERINE HUDON



RP

2025RP-18
RAPPORT DE PROJET

Les rapports de projet sont destinés plus spécifiquement aux partenaires et à un public informé. Ils ne sont ni écrits à des fins de publication dans des revues scientifiques ni destinés à un public spécialisé, mais constituent un médium d'échange entre le monde de la recherche et le monde de la pratique.

Project Reports are specifically targeted to our partners and an informed readership. They are not destined for publication in academic journals nor aimed at a specialized readership, but are rather conceived as a medium of exchange between the research and practice worlds.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO – CIRANO Partners

Partenaires Corporatifs - Corporate Partners

Autorité des marchés financiers
Banque de développement du
Canada
Banque du Canada
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du
Québec
Énergir
Hydro-Québec
Intact Corporation Financière
Investissements PSP
Manuvie
Mouvement Desjardins
Power Corporation du Canada
Pratt & Whitney Canada
VIA Rail Canada

Partenaires gouvernementaux - Governmental partners

Ministère des Finances du Québec
Ministère de l'Économie, de
l'Innovation et de l'Énergie
Innovation, Sciences et
Développement Économique
Canada
Ville de Montréal

Partenaires universitaires - University Partners

École de technologie supérieure
École nationale d'administration
publique
de Montréal
HEC Montreal
Institut national de la recherche
scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.
CIRANO collaborates with many centers and university research chairs; list available on its website.

© Juillet 2025. Maude Laberge, Bile Yacouba Djedou, Thomas Poder, Anaïs Lacasse et Catherine Hudon. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©. *Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*

Les idées et les opinions émises dans cette publication sont sous l'unique responsabilité des auteurs et ne représentent pas les positions du CIRANO ou de ses partenaires. *The observations and viewpoints expressed in this publication are the sole responsibility of the authors; they do not represent the positions of CIRANO or its partners.*

Trajectoires des grands utilisateurs de soins de santé au Québec

Maude Laberge^{*}, Bile Yacouba Djedou[†], Thomas Poder[‡], Anaïs Lacasse[§],
Catherine Hudon^{**}

Résumé/Abstract

Cette étude analyse les trajectoires de soins des grands utilisateurs (GU) de services de santé au Québec, en exploitant la cohorte TorSaDe, qui combine des données administratives de santé et des réponses d'enquête sur une période de dix ans (2007–2016). Elle visait à : (1) caractériser les GU selon leurs profils démographique, clinique et socio-économique; (2) classifier leurs trajectoires d'utilisation des soins par analyse en classes latentes (ACL); et à (3) identifier les facteurs associés à l'appartenance aux différentes classes en utilisant une régression logistique multinomiale. Nous avons défini les GU comme appartenant aux 5 % des utilisateurs ayant les coûts annuels de soins les plus élevés. Les résultats montrent que les GU sont majoritairement plus âgés, socialement défavorisés, porteurs de comorbidités multiples et qu'ils présentent des besoins élevés en soins. Les forêts aléatoires ont permis d'identifier des prédicteurs clés du statut de GU, tels que la multimorbidité, un mauvais état de santé perçu, un faible revenu et la polymédication. L'ACL a permis d'identifier six classes de trajectoires distinctes dont une classe comportant majoritairement des GU persistants, c'est-à-dire des utilisateurs ayant le statut de GU au moins cinq années sur dix. La régression multinomiale a révélé que les utilisateurs persistants étaient plus susceptibles de présenter de graves comorbidités et d'avoir une mauvaise santé mentale, tandis que les utilisateurs transitoires se caractérisaient par des besoins plus aigus et épisodiques en matière de soins de santé. Les résultats soulignent l'importance de l'identification précoce et de la mise en œuvre d'interventions ciblées visant à mieux gérer l'utilisation intensive des soins, dans le but d'améliorer l'efficacité du système et la continuité des services.

This study investigates the care trajectories of high-cost users (HCU) in Quebec, using the TorSaDe cohort, which links administrative health data and survey responses over a ten year period (2007–2016). The objectives were threefold: (1) to characterize HCU based on demographic, clinical, and socio-economic profiles; (2) to classify their care trajectories through latent class analysis (LCA); and (3) to identify factors associated with trajectory class membership via a multinomial logistic regression. HCU were defined as the top 5% of healthcare users by annual cost. The analysis revealed that HCU are disproportionately older, socioeconomically disadvantaged, burdened with multiple chronic conditions, and demonstrate high healthcare needs. Random forest models identified key predictors of HCU status, including multimorbidity, poor self-rated health, low income, and high medication use. LCA identified six distinct classes of

* Université Laval, CIRANO

† Université Laval

‡ Université de Montréal, CIRANO

§ Université du Québec en Abitibi Témiscamingue

** Université de Sherbrooke

trajectories among HCU with one class in which the majority were persistent high-cost users, i.e, HCU at least 5 out of 10 years. Multinomial regression analyses showed that persistent HCU were more likely to have severe comorbidities and poor mental health, whereas transient HCU were characterized by acute, episodic health needs. Findings highlight the need for early identification and tailored interventions to manage healthcare utilization among at-risk populations. This work advances understanding of the dynamic patterns of healthcare use among GU, providing insights for policy-makers aiming to enhance system efficiency and care continuity.

Mots-clés/Keywords : grands utilisateurs; coûts de santé; forêts aléatoires; analyse de classes latentes / high-cost users; healthcare costs; random forest; latent class analysis .

Pour citer ce document / To quote this document

Laberge, M., Djedou, B. Y., Poder, T. G., Lacasse, A., & Hudon, C. (2025). Trajectoires de soins des grands utilisateurs : exploitation des données de la cohorte torsade (2025RP-18, Rapports de projets, CIRANO.) <https://doi.org/10.54932/OBKC2252>

Table des matières

LISTE DES TABLEAUX.....	3
LISTE DES FIGURES	4
SIGLES ET ABRÉVIATIONS.....	5
1. Introduction.....	7
2. Méthodes.....	17
2.1 Source de données : cohorte TorSaDe	17
2.2 Variables	18
2.2.1 Variables d'intérêt.....	18
2.2.2 Variables explicatives et sources des variables	21
2.3 Objectif 1	29
2.3.1 Population et période d'étude	29
2.3.2 Approche analytique	29
a. Analyse descriptive	29
b. Analyse empirique	30
2.4 Objectifs 2 et 3	34
2.4.1 Population et période d'étude	34
2.4.2 Approche analytique	34
a. Analyse en classes latentes (ACL)	34
b. Régression logistique multinomiale	38
3. Résultats.....	40
3.1 Objectif 1	40
3.1.1 Description de l'échantillon.....	40
3.1.2 Distribution des coûts	41
3.1.3 Description des données de l'échantillon des grands utilisateurs et échantillon de la cohorte TorSaDe	42
3.1.4 Caractéristiques des grands utilisateurs de soins	45
3.2 Objectifs 2 et 3	48
3.2.1 Description de l'échantillon utilisé	48
3.2.2 Résultat analyse de classes latentes	49
3.2.3 Résultat du modèle logistique multinomial	53
4. Discussion.....	56
5. Conclusion	67

Références.....	68
Annexes.....	82

LISTE DES TABLEAUX

Tableau 1 : Description des données des GU, Non GU et échantillon de la cohorte TorSaDe	43
Tableau 2 : Critères statistiques post-estimation	50
Tableau 3 : Caractérisation des classes parmi les GU	52
Tableau 4 : Facteurs liés à l'appartenance aux différentes classes de GU.....	54

LISTE DES FIGURES

Figure 1 : Dépenses de soins de santé au Canada.....	7
Figure 2 : Taux de croissance du PIB et des dépenses de soins de santé au Canada.....	8
Figure 3 : Diagramme de flux des participants.....	40
Figure 4 : Évolution des coûts moyens des GU 2012 sur la période de 2007 à 2016.....	41
Figure 5 : Résultats de la prédiction des facteurs associées aux GU	46
Figure 6 : Flux des participants pour classifier les trajectoires.....	48
Figure 7 : Nombre d'années pour lesquelles les individus sont des GU	51
Figure 8 : Représentation graphiques des classes de GU	52

SIGLES ET ABRÉVIATIONS

\$: dollars

AIC : Akaike Information Criterion (critère d'information d'Akaike)

ACL : analyse de classes latentes

Bac : baccalauréat

BIC : Bayesian Information Criterion (critère d'information bayésien)

BPCO : bronchopneumopathie chronique obstructive

CÉGEP : collège d'enseignement général et professionnel

CH-CIM-9 : système combiné de comorbidité Charlson-Elixhauser pour la CIM-9 (Classification internationale des maladies, 9^e révision)

COVID-19 : maladie à coronavirus 2019

EM : espérance maximisation

ESCC : enquête sur la santé des collectivités canadiennes

GU : grands utilisateurs

HUI3 : indice de l'état de santé (Health Utilities Index)

IC : intervalle de confiance

ICIS : institut canadien d'information sur la santé

IMC : indice de masse corporelle

IPC : indice de prix à la consommation

kg : kilogramme

m² : mètre carré

MC : modèle complet

ML : maximum de vraisemblance

MR : modèle réduit

MSSS : ministère de la Santé et des Services sociaux

NIRRU : niveau d'intensité relative des ressources utilisées

OCDE : Organisation de coopération et de développement économiques

PIB : produit intérieur brut

PoLCA : Polynomial Latent Class Analysis (régression polynomiale de classe latente)

Post-bac : post-baccalauréat

RAMQ : Régie de l'assurance maladie du Québec

RC : rapport de cotes

RPAM : régime public d'assurance médicaments

SRAP : système de recherche appliquée en santé (plateforme de données jumelées)

TorSaDe : Trajectoire Santé - Données Enrichies

VIF : Variance Inflation Factor (facteur d'inflation de la variance)

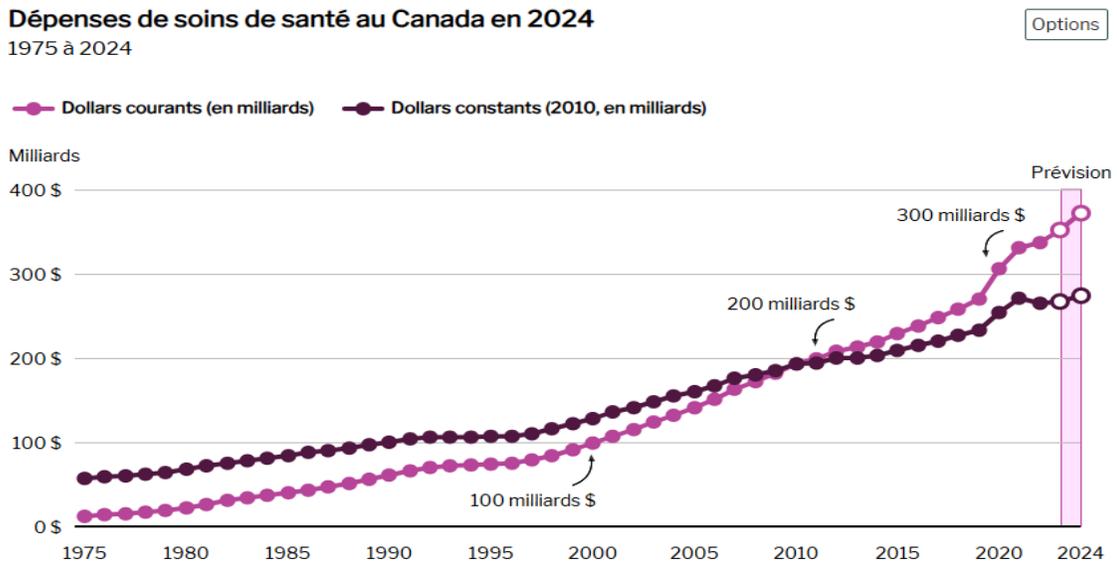
1. Introduction

Contexte

La gestion de l'utilisation des services de santé est un enjeu majeur pour les systèmes de santé à travers le monde, particulièrement face à la croissance constante des coûts et à la demande accrue en soins de santé (OCDE, 2019).

Au Canada, les dépenses totales de santé au Canada étaient estimées à 372 milliards de dollars en 2024, soit 9 054 \$ par habitant, représentant une augmentation¹ de 5,7 % par rapport à 2023 (Institut canadien d'information sur la santé ci-après ICIS, 2024). Ce taux de croissance, supérieur à la moyenne annuelle observée entre 2011 et 2019 (3,9 %), contraste avec les hausses exceptionnelles de 13,2 % en 2020 et de 7,8 % en 2021, attribuables à la pandémie de COVID-19 (figure 1). Cette progression serait en partie liée à l'augmentation rapide de la population et à l'évolution de l'économie, qui influencent les coûts des services de santé (ICIS, 2024).

Figure 1: Dépenses de soins de santé au Canada



Graphique linéaire illustrant les dépenses totales de santé de 1975 à 2024, en dollars courants et en dollars constants de 2010. Il présente les tendances en matière de dépenses de santé observées depuis plus de 40 ans.

Remarque

PIB : produit intérieur brut.

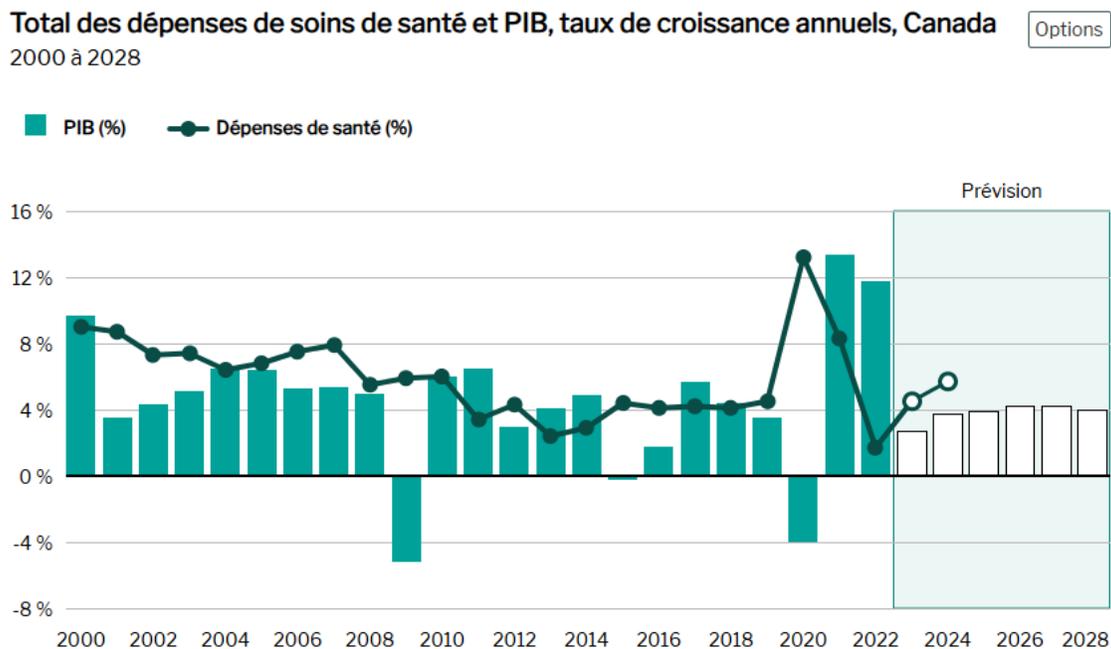
Source

Base de données sur les dépenses nationales de santé, Institut canadien d'information sur la santé.

¹ D'après l'ICIS, la hausse des dépenses de santé pourrait se prolonger, notamment en raison de l'accélération récente de la croissance démographique. Par ailleurs, l'évolution de la conjoncture économique mondiale pourrait influencer les coûts du secteur de la santé, notamment à travers la renégociation des ententes avec les prestataires de services.

Au Canada, les dépenses en santé progressent au même rythme, voire plus rapidement que l'économie dans son ensemble depuis plusieurs années (ICIS, 2023, 2024). Au cours des deux dernières années, elles ont connu une croissance plus soutenue que celle de l'économie, entraînant une augmentation du ratio des dépenses de santé par rapport au produit intérieur brut (PIB). Les dépenses de santé auraient ainsi augmenté de 5,7 % en 2024, tandis que la croissance économique se serait limitée à 3,7 %, portant la part des dépenses de santé à 12,4 % du PIB² (figure 2) (ICIS, 2024). En excluant les années de pandémie (2020 et 2021), il s'agit du niveau le plus élevé jamais atteint.

Figure 2: Taux de croissance du PIB et des dépenses de soins de santé au Canada



Graphique à barres illustrant les taux de croissance annuels du PIB (2000 à 2028) combiné à un graphique linéaire des dépenses totales de santé (2000 à 2024). Cette combinaison permet de voir la relation entre les dépenses de santé et la croissance économique.

Remarque

PIB : produit intérieur brut.

Sources

Base de données sur les dépenses nationales de santé, Institut canadien d'information sur la santé. Ministère des Finances Canada. *Budget de 2024 : Une chance équitable pour chaque génération*. 2024.

Dans un contexte marqué par une demande croissante en soins de santé due au vieillissement démographique, à la pénurie de personnel médical et à l'augmentation des coûts des traitements, le gouvernement fédéral a conclu en février 2023 un accord avec les

² Fait référence au PIB nominal aux prix courants, non corrigé en fonction de l'inflation.

provinces et territoires, afin de renforcer le financement du système de santé (ICIS, 2023). Cet engagement prévoit un investissement supplémentaire de 196,1 milliards de dollars sur une période de dix ans. Cet accord s'inscrit dans la continuité des ententes précédentes, notamment celles de 2004 et de 2017, et marque une intensification des efforts fédéraux pour soutenir la viabilité des soins.

Au Québec, le budget alloué à la santé et aux services sociaux a connu une expansion au fil des ans. Le budget³ du Québec pour l'exercice 2024-2025 évaluait les dépenses publiques en santé à 61,9 milliards de dollars (G\$), soit 49,81 % des dépenses publiques totales du gouvernement du Québec, marquant une progression par rapport aux 59,4 milliards de dollars de l'année précédente. Cette augmentation de 2,5 milliards de dollars, soit une croissance de 4,2 %, s'inscrit dans une tendance à long terme visant à stabiliser et à améliorer l'accès aux soins. En termes relatifs, les dépenses en santé représentaient 10,8 % du PIB du Québec, en 2023-2024.

Une observation récurrente dans la littérature révèle que les dépenses de santé sont concentrées dans une minorité d'utilisateurs qualifiés de « grands utilisateurs » (GU). Une étude de la population ontarienne au Canada rapporte que 5 % des usagers est responsable de près de 50 % des dépenses totales de santé (Wodchis et al., 2016). Ce phénomène soulève des questions fondamentales sur la répartition des ressources et la capacité des systèmes de santé à répondre efficacement aux besoins complexes de cette sous-population (Berk et Monheit, 2001; Rosella et al., 2014).

Comment définir les grands utilisateurs de soins?

L'étude de l'intensité d'utilisation des soins par certains individus a conduit à l'émergence du concept de GU de soins (Krieg et al., 2016; Moe et al., 2013). L'un des premiers travaux scientifiques ayant documenté la fréquence de consultations des patients remonte à 1975 (Ullman et al., 1975). Cette recherche a mis en évidence les caractéristiques démographiques et cliniques d'un sous-groupe représentant 7,5 % des patients, mais générant à lui seul 22,3 % des visites aux services d'urgence (Ullman et al., 1975).

³ Le Budget de dépenses 2024-2025 du gouvernement du Québec s'élève à 124,3 milliards de dollars. Les dépenses consolidées, incluant le service de la dette, atteignent 157,6 milliards de dollars, soit une augmentation de 4,4 % par rapport à l'exercice précédent.

À ce jour, aucune définition universelle du concept de grand utilisateur des soins n'a été arrêtée (Moe et al., 2013). En termes de nature des soins, les GU sollicitent aussi bien des consultations ambulatoires en médecine générale et spécialisée que des visites aux services d'urgence ou des hospitalisations (Doupe et al., 2012; Hansagi et al., 2001; Huynh et al., 2016; O'Toole et al., 2007; Rask et al., 1998). Mais la littérature est plus riche au sujet de l'analyse des GU des soins d'urgence (Andrén et Rosenqvist, 1987; Doupe et al., 2012; Fuda et Immekus, 2006; Hansagi et al., 2001; Hudon et al., 2017; Moe et al., 2013; Sun et al., 2003). Plusieurs raisons expliquent ce choix. Tout d'abord, une fréquentation élevée des services d'urgence est souvent corrélée à une consommation accrue de soins ambulatoires et hospitaliers (Doupe et al., 2012; Hansagi et al., 2001; O'Toole et al., 2007). Par ailleurs, ces visites répétées peuvent être le reflet d'un besoin de soins non satisfait par le reste du système de santé (Malone, 1995). La continuité des soins primaires et leur accessibilité semblent jouer un rôle clé dans la baisse de fréquentation des services d'urgence (Hudon et al., 2016; Krieg et al., 2016), suggérant qu'une meilleure prise en charge en amont pourrait optimiser la pertinence des services et mieux répondre aux besoins des patients (Baker et Baker, 1994). D'autres études ont défini les GU en tenant compte des coûts engendrés (Nghiem et al., 2023; Sowa et al., 2021; Wong et al., 2024; M. Zhang et al., 2023), de la fréquence d'admission (Chiu et al., 2022), de la durée des séjours à l'hôpital (Karunayawong et al., 2024) et des conditions spécifiques de santé (Anderson et al., 2018; Armoon et al., 2022; Hudon et al., 2017; Moe et al., 2021; Sowa et al., 2021). Malgré la diversité des modes d'identification des GU dans la littérature, ils reposent tous sur un point commun : le recours intensif aux services de santé. Certains chercheurs les ont définis au moyen de mesures absolues d'utilisation, comme le fait de visiter les services d'urgence au moins trois fois par an (Armoon et al., 2022; Doupe et al., 2012; Hudon et al., 2017; Moe et al., 2021) ou de consulter un médecin généraliste ou un spécialiste au moins dix fois par an (Tille et al., 2017). D'autres chercheurs ont plutôt choisi une mesure relative comme les 10 % de patients comptabilisant le plus de consultations à l'hôpital (Chiu et al., 2022; Moe et al., 2021). Ce critère permet de détecter précocement les comportements d'hyper-utilisation, qui peuvent refléter une mauvaise orientation dans le parcours de soins, un défaut de suivi ambulatoire ou une mauvaise couverture des besoins.

Au-delà de cette fréquence élevée de consultations, la prise en charge des GU représente un poids financier considérable pour les systèmes de santé. Les GU sont aussi définis comme les patients faisant partie des 5 à 10 % qui engendrent les coûts médicaux les plus élevés (Wong et al., 2024; M. Zhang et al., 2023) ou comme les patients responsables de 40 à 60 % des dépenses hospitalières totales (Nghiem et al., 2023; Sowa et al., 2021). Ces coûts élevés s'expliquent par des hospitalisations prolongées, des traitements spécialisés et un recours fréquent aux soins ambulatoires ou aux services d'urgence pour des motifs non urgents. Cette approche par les coûts met en évidence la disproportion entre les ressources consommées et le nombre de patients concernés.

Chiu et al. (2022) définissent les GU comme les patients qui cumulent au moins cinq hospitalisations par an. Ces patients se caractérisent par une durée moyenne d'hospitalisation largement supérieure à celle des patients non-GU (Karunayawong et al., 2024). Ils présentent aussi un risque élevé de réadmission dans les trente jours suivant leur congé de l'hôpital (Sowa et al., 2021), ce qui traduit une instabilité de leur état de santé et soulève des questions quant à la qualité et à la continuité du suivi post-hospitalisation. Ce critère est particulièrement utile pour identifier le profil des patients les plus instables.

Dans le cadre de cette étude, nous définissons les GU comme les patients faisant partie des 5 % des patients générant les coûts annuels les plus élevés. Ce seuil de 5 % repose sur la concentration disproportionnée des coûts de santé, où cette petite fraction de la population représente environ 50 % des dépenses hospitalières (Wong et al., 2024; M. Zhang et al., 2023). Cette approche, largement utilisée dans la littérature après l'approche basée sur la fréquence des consultations, assure une comparabilité méthodologique (Nghiem et al., 2023) et permet d'optimiser la gestion des ressources en ciblant les interventions sur les patients à fort impact budgétaire (Sowa et al., 2021). L'utilisation de données administratives objectives évite le biais lié au simple comptage des visites médicales et identifie les véritables patients à coûts élevés (Moe et al., 2021). Enfin, cette approche distingue fréquence de consultations et coût réel, puisque certains patients sont très coûteux même s'ils consultent peu (M. Zhang et al., 2023).

L'utilisation de données administratives objectives présente plusieurs avantages. Contrairement aux simples comptages déclaratifs, ces données permettent de capter de manière plus fiable et exhaustive les contacts réels avec le système de santé, ce qui améliore

l'exactitude des analyses (Moe et al., 2021). De plus, leur richesse en informations détaillées (diagnostics, services reçus, coûts associés) et leur nature longitudinale en font une source précieuse pour identifier les patients à coûts élevés sur une période donnée.

Qui sont les grands utilisateurs de soins?

Plusieurs études suggèrent que les GU ne se distinguent pas uniquement par leur état de santé ou leurs besoins cliniques, mais aussi par des caractéristiques socio-économiques. Au Canada, des facteurs tels qu'un faible revenu, un faible niveau de scolarité, l'insécurité alimentaire ou encore le fait de ne pas être propriétaire de son logement sont positivement associés aux GU (Fitzpatrick et al., 2015; Rosella et al., 2014; Zhang et al., 2023). Les personnes âgées vivant dans des quartiers à statut socio-économique faible recourent plus souvent aux services de santé de manière intensive (Andreyeva et al., 2023). Wong et al. (2024) confirment également une interaction marquée entre précarité sociale et besoins non couverts qui contribue à une utilisation disproportionnée des services de santé. Ces constats indiquent l'importance de prendre en compte ces facteurs dans la caractérisation des GU.

Le modèle comportemental d'utilisation des soins d'Andersen (1995) constitue un cadre conceptuel pour l'analyse des déterminants du recours aux soins, que l'on peut appliquer aux GU. Ce modèle repose sur trois catégories de déterminants : les facteurs prédisposants, les facteurs facilitants ou inhibants, et les facteurs de besoin. Les facteurs prédisposants correspondent aux caractéristiques sociodémographiques et individuelles influençant la probabilité d'un recours aux soins, notamment l'âge, le sexe, l'appartenance ethnique et le niveau d'éducation. Les facteurs facilitants ou inhibants renvoient aux ressources matérielles et organisationnelles, telles que la couverture d'assurance-maladie ou la disponibilité des infrastructures de soins. Les facteurs de besoin se rapportent à l'état de santé perçu ou objectif de l'individu et sont souvent associés à des pathologies chroniques ou à des troubles psychiatriques (Andersen, 1995; LaCalle et al., 2013).

Le rôle du sexe et de l'âge a été étudié mais les consensus sont limités. Certaines études rapportent une probabilité plus élevée d'être un GU chez les hommes (Hwang et al., 2015; Paul et al., 2010), et d'autres chez les femmes (Hudon et al., 2017; Palmer et al., 2014). Certaines recherches suggèrent que les personnes âgées sont plus enclines à être de GU (Huynh et al., 2016; Hwang et al., 2015; Palmer et al., 2014; Paul et al., 2010), alors que

d'autres identifient les jeunes adultes comme étant plus représentés parmi les GU (Hudon et al., 2017). Un faible niveau d'éducation, une précarité économique et l'absence de domicile fixe sont associés à une consommation élevée de soins (Bodenmann et al., 2017; Hudon et al., 2016; Ronksley et al., 2016; Sun et al., 2003) tandis qu'être marié semble avoir un effet protecteur (Bieler et al., 2012; Sun et al., 2003). Certaines habitudes de vie, telles que le tabagisme, la sédentarité, l'obésité ainsi que la consommation et l'abus d'alcool ou de drogues, ainsi que la surconsommation de certains médicaments, notamment les opioïdes et les benzodiazépines, sont plus fréquemment observées chez les GU (Chiu et al., 2022; Kurteva et al., 2023; Moe et al., 2021; Wong et al., 2024). L'adhésion aux traitements médicaux joue également un rôle clé : une mauvaise observance augmente le risque de complications et donc le recours aux soins (Moe et al., 2021).

Si les facteurs prédisposants modulent la probabilité d'un recours fréquent aux soins, les facteurs facilitants ou inhibants influencent quant à eux la capacité réelle des individus à accéder aux services de santé. Parmi ces facteurs, l'antécédent d'utilisation des soins est l'un des principaux indicateurs d'être un GU. Les personnes ayant déjà eu un recours intensif aux soins de santé ont une forte probabilité de maintenir cette tendance dans le temps (Krieg et al., 2016). L'accès aux soins primaires constitue un autre élément clé. Certaines études indiquent que les patients inscrits auprès d'un médecin de famille présentent une probabilité plus élevée d'être des GU (Bieler et al., 2012; Krieg et al., 2016; Palmer et al., 2014; Rask et al., 1998; Sun et al., 2003). Ce résultat peut paraître paradoxal, mais il s'explique par le fait que la majorité des GU ont des besoins de santé complexes, souvent liés à la présence de maladies chroniques multiples, de troubles mentaux ou de conditions sociales précaires. Ces patients, en raison de leur profil, ont davantage de contacts avec le système de santé. Cependant, avoir un médecin de famille pourrait permettre de réduire la fréquence d'utilisation des soins (Hudon et al., 2016, 2017). L'inscription auprès d'un médecin de famille agirait comme levier de réduction de l'utilisation, en améliorant la continuité des soins, la prévention des décompensations et la coordination interprofessionnelle. Une accessibilité perçue comme faible aux soins de première ligne et des difficultés de coordination des soins sont des obstacles souvent rapportés par les GU (Hudon et al., 2016). Aucune relation nette n'a été démontrée entre le fait d'habiter en milieu rural ou urbain et la probabilité d'être un grand utilisateur (Bieler

et al., 2012; Doran et al., 2013; Krieg et al., 2016). Toutefois, des disparités importantes existent entre les pays. Par exemple, aux États-Unis, les patients ont tendance à se tourner plus fréquemment vers les services hospitaliers (Lasser et al., 2006). Parmi les trois dimensions du modèle d'Andersen, les facteurs de besoin constituent les déterminants les plus directs de la grande utilisation des soins. La présence de maladies chroniques multiples est un élément central. Les GU souffrent fréquemment de maladies chroniques, notamment cardiovasculaires (Sowa et al., 2021), métaboliques (Wong et al., 2024), rénales, respiratoires ou musculosquelettiques (Sowa et al., 2021). En plus des maladies physiques, les maladies mentales comme la schizophrénie, les troubles psychotiques, les troubles bipolaires, la dépression sévère et l'anxiété sont associés à un recours accru aux services hospitaliers et au statut de GU (Anderson et al., 2018; Armoon et al., 2022; Chiu et al., 2022).

En combinant ces différentes dimensions, le modèle d'Andersen met en évidence l'interaction entre les facteurs prédisposants, facilitants et de besoin, ce qui explique pourquoi certains individus deviennent des GU des soins.

Un aspect clé pour mieux comprendre les GU est l'analyse de leurs trajectoires de soins. Cette analyse offre une lecture dynamique du parcours de soins, en intégrant les séquences, transitions et ruptures dans la prise en charge (Corbin et Strauss, 1988; Wodchis et al., 2016). Une trajectoire de soins désigne ainsi l'enchaînement d'événements vécus par un individu au fil du temps, telles que les consultations en soins primaires, les hospitalisations, les visites aux services d'urgence, les transitions ou les ruptures dans la prise en charge (Krieg et al., 2016; Moe et al., 2013; Wodchis et al., 2016). Chez les GU, ces trajectoires sont souvent complexes, instables, et ponctuées d'interruptions révélant des failles dans la coordination des soins (Chiu et al., 2022; Moe et al., 2021; Wodchis et al., 2016). Ces dernières décennies, l'intérêt pour l'analyse des trajectoires dans le domaine de la santé s'intensifie (Nguena Nguetack et al., 2020), notamment pour étudier les GU. L'analyse longitudinale permet d'identifier les parcours singuliers des GU, souvent masqués dans les approches transversales classiques. Elle éclaire les processus qui mènent certains patients à devenir ou à rester GU, là où d'autres, aux profils cliniques similaires, n'adoptent pas la même trajectoire (Nguena Nguetack et al., 2020; Wodchis et al., 2016).

Plusieurs études ont mis en évidence des profils distincts, selon la régularité des soins, la chronicité des besoins ou la présence de ruptures de suivi (Chiu et al., 2022; Moe et al., 2021; Wodchis et al., 2016). Ces profils reflètent l'intensité, la stabilité et la structure du recours aux soins. Gruneir et al. (2011) ont montré que l'absence de suivi structuré après une hospitalisation chez les personnes âgées peut engendrer des cycles de réadmission. Moe et al. (2021), via des analyses de classes latentes, identifient plusieurs profils selon les types de recours aux services d'urgence, tandis que Chiu et al. (2022) soulignent l'importance de la coordination des soins pour éviter les trajectoires de surutilisation, même en présence de maladies chroniques. Ces trajectoires peuvent être décrites selon plusieurs dimensions : la durée (chronique ou temporaire), l'intensité (modérée ou élevée), la stabilité (régulière ou erratique) et la structure (coordonnée ou fragmentée). Cette diversité appelle une analyse plus fine des facteurs contextuels et organisationnels (Anderson et al., 2018; Hudon et al., 2016).

Historiquement, les recherches sur les GU s'appuyaient sur des définitions statiques de seuils de consommation ou de coûts. Or, ces approches, bien que descriptives, peinent à saisir la complexité des dynamiques à l'œuvre (Moe et al., 2021; Wodchis et al., 2016). Pour dépasser ces limites, l'analyse des trajectoires de soins propose une lecture plus dynamique, permettant d'identifier les transitions, les ruptures et les moments charnières dans les parcours de soins. Cette approche repose sur l'exploitation de données médico-administratives longitudinales permettant de suivre les patients sur plusieurs mois ou années (Gruneir et al., 2011; Wodchis et al., 2016). Certaines études, comme celles de Wodchis et al. (2016) et Hudon et al. (2016), ont analysé des suivis sur cinq ans pour identifier des trajectoires stables, transitoires ou chaotiques. D'autres travaux, comme ceux de Moe et al. (2021) et de Chiu et al. (2022), utilisent des analyses de classes latentes sur des périodes plus courtes (12 à 24 mois), pour identifier des profils types d'utilisation.

Sur le plan méthodologique, plusieurs approches sont utilisées : l'analyse de séquences, qui permet de visualiser les enchaînements d'événements; l'analyse de classes latentes, qui regroupe les patients selon leur schéma de recours; ou encore les modèles longitudinaux et les analyses de survie, qui permettent de mesurer les intervalles entre les événements critiques (Chiu et al., 2022; Kurteva et al., 2023; Nguena Nguéfack et al., 2020; Wodchis et al., 2016). Malgré ces avancées, plusieurs limites demeurent, particulièrement dans la

littérature québécoise. Les données sont souvent restreintes, les suivis sont courts et les dimensions sociales sont peu intégrées. Très peu d'études combinent des données médico-administratives avec des enquêtes populationnelles. La littérature manque également d'articulations entre trajectoires et déterminants sociaux, cliniques et organisationnels.

Problématique

Bien que la concentration des dépenses de santé au sein d'une minorité d'utilisateurs dénommée GU soit bien documentée, l'analyse de leurs trajectoires de soins demeure encore marginale dans les travaux empiriques. Ces patients présentent des profils hétérogènes, caractérisés par des comorbidités, une vulnérabilité socio-économique, et une instabilité dans le recours aux soins (Hudon et al., 2016; Moe et al., 2021). La majorité des travaux s'appuie sur des données transversales, qui offrent une vision statique (Gruneir et al., 2011; Moe et al., 2021). Or, plusieurs travaux récents appellent à adopter une approche longitudinale, apte à capter la complexité, l'hétérogénéité et la non-linéarité des trajectoires de soins (Chiu et al., 2022; Nguena Nguefack et al., 2020; Wodchis et al., 2016). Dans ce contexte, le recours à des données médico-administratives enrichies, jumelées à des données d'enquête populationnelle, constitue une occasion de mieux comprendre les mécanismes à l'origine de la grande utilisation des soins.

Cette étude s'inscrit dans cette perspective en cherchant à approfondir la compréhension des GU de soins au moyen d'une analyse longitudinale de leurs trajectoires de soins. En mobilisant le modèle comportemental d'utilisation des soins d'Andersen (1995) comme cadre d'analyse, il s'agira de caractériser les GU selon leurs profils clinique, démographique et social, de classer leurs trajectoires en fonction des archétypes d'utilisation observés, puis d'identifier les facteurs associés à l'appartenance à chacun de ces profils. L'objectif est de mettre en lumière les dynamiques sous-jacentes à la grande utilisation des services, afin de soutenir l'élaboration de stratégies d'intervention ciblées mieux adaptées à la complexité des besoins de cette population, et susceptibles d'améliorer l'efficacité du système de santé québécois.

Les objectifs de l'étude sont les suivants :

1. **Objectif 1** : caractériser les grands utilisateurs de soins;

2. **Objectif 2** : classifier les trajectoires d'utilisation de soins des grands utilisateurs selon les caractéristiques de leur utilisation;
3. **Objectif 3** : identifier les facteurs liés à l'appartenance à l'une ou l'autre des classes d'utilisateurs.

2. Méthodes

Nous utilisons différentes méthodes quantitatives pour atteindre chacun des objectifs.

2.1 Source de données : cohorte TorSaDe

Nous utilisons la cohorte TorSaDe (Trajectoires Santé - Données Enrichies), une base de données longitudinale constituée pour étudier les trajectoires de soins et les coûts associés pour la population québécoise (Samadoulougou et al., 2022; Vanasse et al., 2021). L'ampleur de cette cohorte et la richesse de ses données en font un outil précieux pour la recherche en santé publique. La cohorte TorSaDe regroupe des individus du Québec ayant répondu à l'Enquête sur la santé des collectivités canadiennes (ESCC) entre 2007 et 2016, et ayant autorisé le jumelage⁴ de leurs réponses avec les données administratives de la Régie de l'assurance maladie du Québec (RAMQ) (Vanasse et al., 2021). L'ESCC est une enquête transversale qui recueille des renseignements sur la santé, y compris sur l'utilisation des soins, sur les habitudes de vie et sur l'état de santé général de la population canadienne.

Au total, cinq cycles de l'ESCC générale sont inclus dans cette cohorte TorSaDe : 2007-2008, 2009-2010, 2011-2012, 2013-2014, 2015-2016. Les répondants à l'ESCC ont été sélectionnés selon un plan d'échantillonnage probabiliste stratifié, conçu pour être représentatif de la population canadienne vivant en ménage privé. Il est important de noter que les individus n'étaient pas les mêmes d'un cycle à l'autre. Dans l'éventualité d'une deuxième participation, seule la participation la plus récente d'un individu a été retenue, afin d'éviter la duplication des profils individuels et de préserver l'indépendance des observations.

⁴ Presque tous les répondants ont autorisé le jumelage.

La cohorte TorSaDe inclut plus de 100 000 participants et couvre la période de 1996 à 2016. La cohorte TorSaDe se distingue par plusieurs atouts majeurs. La cohorte suit les participants sur une période de 21 ans, permettant des analyses sur l'évolution des trajectoires de soins dans le temps. Cette durée permet d'analyser les changements dans l'utilisation des soins et les coûts associés sur une longue période, et est essentielle pour comprendre les dynamiques de soins à long terme (Vanasse et al., 2021). Les données combinent les réponses autorapportées de l'ESCC avec les données administratives de la RAMQ, fournissant une vue complète des caractéristiques sociodémographiques, des états de santé perçus et de l'utilisation réelle des services de santé. Elles incluent des informations détaillées sur les visites médicales, les hospitalisations et les coûts de soins de santé, permettant ainsi des analyses approfondies (Rosella et al., 2014). La cohorte inclut une population québécoise vaste et diversifiée, reflétant les différentes réalités et besoins en matière de santé dans la province. Cette représentativité élevée permet de généraliser les résultats à l'ensemble de la population québécoise. La cohorte TorSaDe est particulièrement adaptée pour atteindre nos objectifs pour plusieurs raisons. Les données longitudinales permettent de repérer les individus qui sont des GU de soins et d'examiner leur trajectoire sur une longue période, distinguant ainsi les utilisateurs récurrents de ceux dont l'utilisation est ponctuelle (Qureshi et al., 2020). La richesse et la temporalité des données permettent d'appliquer des techniques de classification avancées, telles que les analyses de classes latentes, pour identifier différentes trajectoires de soins. Les informations détaillées sur les caractéristiques sociodémographiques des participants permettent d'examiner les facteurs influençant les trajectoires de soins, comme le niveau de scolarité, le statut matrimonial et d'autres variables pertinentes.

2.2 Variables

2.2.1 Variables d'intérêt

Pour chaque année de la période d'étude, nous avons créé une variable binaire nommée « grand utilisateurs » (GU) qui prend la valeur 1 si la personne fait partie du 5^e percentile de coûts les plus élevés et la valeur 0 dans toute autre situation. Ainsi, pour l'objectif 1 (caractériser les grands utilisateurs de soins), la variable d'intérêt est d'être GU en 2012. En revanche, pour les objectifs 2 (classifier les trajectoires d'utilisation de services) et 3

(identifier les facteurs liés à l'appartenance à une classe d'utilisateurs), la variable d'intérêt correspond au nombre d'années pendant lesquelles une personne a été identifiée comme GU au cours des dix années de la période d'étude. Les coûts utilisés pour définir le statut de GU correspondent aux coûts totaux des services de santé payés au cours de chaque année de la période de l'étude. Nous avons utilisé l'année calendaire, c'est-à-dire du 1^{er} janvier au 31 décembre. Ces coûts sont calculés en utilisant les coûts associés aux hospitalisations, aux chirurgies d'un jour, aux visites au service d'urgence et aux consultations médicales. Les coûts de services pharmaceutiques n'ont pas été inclus, car ces données n'étaient disponibles que pour les personnes couvertes par le Régime d'assurance publique des médicaments (RPAM), soit surtout des personnes de 65 ans et plus.

Les hospitalisations et les chirurgies d'un jour sont associées à des NIRRU (Niveau d'Intensité Relative des Ressources Utilisées). Le NIRRU est un indicateur permettant d'évaluer l'intensité des ressources de santé utilisées par une personne en fonction des services reçus (hospitalisations, consultations, interventions, etc.). Il sert à comparer la consommation des ressources, suivre l'évolution des besoins en soins et optimiser la planification des services de santé. Dans cette étude, nous avons retenu la dernière valeur attribuée au NIRRU par le ministère de la Santé et des Services sociaux (MSSS) qui date de 2019. Par souci de cohérence, cette année a été retenue comme année de référence pour l'estimation du coût des services dont l'utilisation est associée à un NIRRU et pour leur conversion en dollars constants. Nous avons décidé d'utiliser les dollars canadiens de 2019, car cela permet d'assurer la comparabilité des coûts sur l'ensemble des années étudiées en neutralisant l'effet de l'inflation. Nous avons utilisé les coûts réels en dollars courants, puis nous les avons convertis en dollars constants de 2019 en utilisant l'indice des prix à la consommation (IPC) (Kouakou et al., 2025).

Coûts des hospitalisations :

- **NIRRU** : en 2019, la valeur monétaire d'un NIRRU pour une hospitalisation était de 6 026,06 \$.
- **Calcul des coûts** : le coût total de chaque hospitalisation répertoriée dans les données de la RAMQ est calculé en multipliant le NIRRU de l'hospitalisation par la valeur

monétaire d'un NIRRU. Par exemple, si une hospitalisation a un NIRRU de 2, le coût serait de 2 multiplié par 6 026,06 \$, soit 12 052,12 \$.

Coûts des chirurgies d'un jour :

- **NIRRU pour chirurgie d'un jour** : comme pour les hospitalisations, chaque chirurgie d'un jour est associée à un NIRRU. En 2019, la valeur monétaire d'un NIRRU pour une chirurgie d'un jour était de 1 368,30 \$.
- **Calcul des coûts** : le coût total de chaque chirurgie d'un jour est calculé en multipliant le NIRRU par la valeur monétaire correspondante. Par exemple, une chirurgie avec un NIRRU de 1,5 coûterait 1,5 multiplié par 1 368,30 \$, soit 2 052,45 \$.

Coûts des visites au service d'urgence :

- **Estimation des coûts** : le coût des visites au service d'urgence est estimé en utilisant le NIRRU pour chaque visite. Pour chaque visite au service d'urgence, une valeur monétaire de 355,60 \$ est attribuée à un NIRRU.
- **Calcul des coûts** : le coût total de chaque visite au service d'urgence enregistrée est simplement 1 multiplié par 355,60 \$, soit 355,60 \$ par visite.

Coûts des consultations médicales :

- **Facturation à la RAMQ** : le coût des consultations médicales est basé sur les montants facturés pour chaque acte médical par les médecins à la RAMQ. Ces montants étaient enregistrés en dollars courants de l'année où la consultation a eu lieu. Nous avons fait une conversion en dollars constants de 2019 en utilisant l'IPC.

Le coût annuel par individu est égal à la somme des coûts des différents services de santé utilisés au cours de chaque année. Ce coût a été calculé pour chaque année de la période de suivi, soit de 2007 à 2016. Une fois les coûts obtenus pour chaque année, nous avons déterminé le percentile de coût associé à chaque individu pour identifier les GU, soient ceux appartenant au 5^e percentile. Ce qui signifie que 95 % des individus se trouvent en dessous de ce seuil en termes de coûts.

Nous avons réalisé des diagrammes de distribution, incluant des boîtes à moustaches et des histogrammes, afin de visualiser la distribution des coûts. Nous avons également comparé les seuils de coûts d'une année à l'autre pour observer les variations.

2.2.2 Variables explicatives et sources des variables

Nous avons regroupé les facteurs pouvant être associés au statut de GU en quatre catégories principales : démographiques, socio-économiques, cliniques et comportementaux. Le choix de ces facteurs repose à la fois sur le cadre conceptuel du modèle comportemental d'Andersen, qui distingue les facteurs prédisposants, facilitants et de besoin (Andersen, 1995), et sur plusieurs études empiriques ayant démontré leur pertinence dans la prédiction d'une utilisation élevée de services de santé (Armoon et al., 2022; Moe et al., 2021; Wong et al., 2024).

Facteurs démographiques

Les variables démographiques sont essentielles pour comprendre les différences structurelles dans l'utilisation des soins.

- **Sexe** : les hommes et les femmes présentent des profils de consommation des soins distincts. Par exemple, plusieurs études montrent que les femmes ont tendance à consulter plus fréquemment les médecins généralistes, tandis que les hommes sont plus susceptibles d'avoir des hospitalisations non planifiées (Tille et al., 2017). Cette variable provient de l'ESCC.
- **Âge et Âge au carré** : l'âge a été inclus en tant que variable continue, accompagné de son carré ($\hat{\text{Age}}^2$), afin de capter la relation non linéaire entre l'âge et la consommation des soins (Wong et al., 2024). Par exemple, les jeunes adultes peuvent avoir une utilisation plus modérée des soins, tandis que les personnes âgées sont plus susceptibles d'être de GU en raison de maladies chroniques. À noter que la variable provient de l'ESCC. Uniquement pour l'objectif 1, nous avons utilisé l'âge des individus en 2012, bien que les autres données soient des données antérieures à cette date. Nous avons fait ce choix à des fins d'harmonisation de l'âge.

- **Pays de naissance** : les personnes nées à l'étranger peuvent faire face à des barrières linguistiques et culturelles influençant leur accès aux soins de santé. De plus, les immigrants peuvent se distinguer des natifs du pays en termes de comportements de santé et de recours aux services médicaux (M. Zhang et al., 2023). Dans notre base de données, la variable « pays de naissance », issue de l'ESCC, était initialement catégorielle et comprenait les modalités suivantes : *Autres pays de l'Amérique du Nord, Amérique centrale, Antilles et Bermudes, Europe, Afrique, et Canada*. Afin de simplifier les analyses et d'assurer une comparabilité avec la littérature existante, nous avons regroupé ces modalités en deux catégories : *Canada et extérieur du Canada*.
- **Lieu de résidence** : nous avons distingué les milieux ruraux et urbains, car les différences d'accessibilité aux soins sont bien documentées dans la littérature. Les résidents ruraux ont un accès plus limité aux soins spécialisés, ce qui peut augmenter leur recours aux services d'urgence et hospitaliers (Moe et al., 2021). La variable « lieu de résidence » qui provient des données de l'ESCC était une variable catégorielle dont les modalités étaient les suivantes : *Montréal, Laval, Montérégie, Capitale-Nationale, Outaouais, Saguenay-Lac-Saint-Jean, Chaudière-Appalaches, Abitibi-Témiscamingue, Bas-Saint-Laurent, Côte-Nord, Gaspésie / Îles-de-la-Madeleine, Nord-du-Québec, Mauricie / Centre-du-Québec, Estrie, Laurentides, Lanaudière*. Nous avons regroupé les modalités pour la rendre binaire, à savoir Urbain pour les modalités *Montréal, Laval, Montérégie, Capitale-Nationale, Outaouais* et Rural pour les modalités *Saguenay-Lac-Saint-Jean, Chaudière-Appalaches, Abitibi-Témiscamingue, Bas-Saint-Laurent, Côte-Nord, Gaspésie / Îles-de-la-Madeleine, Nord-du-Québec, Mauricie / Centre-du-Québec, Estrie, Laurentides, Lanaudière*. Ce regroupement repose sur des critères de densité de population, d'accessibilité aux services de santé et d'organisation des soins, généralement plus concentrés en milieu urbain. Il reflète également les disparités entre les milieux urbains et ruraux en matière d'offre de soins, de continuité des services et de temps d'attente (Doran et al., 2013; Hudon et al., 2016).

Facteurs socio-économiques

Les inégalités économiques et sociales jouent un rôle clé dans la consommation des soins de santé.

- **Niveau d'éducation** : classé en plusieurs catégories (école secondaire non terminée, diplôme secondaire, CÉGEP ou certificat avant le baccalauréat, baccalauréat ou plus), le niveau d'éducation influence la littératie en santé et la capacité à naviguer dans le système de soins (Armoon et al., 2022). Les études montrent que les personnes moins éduquées consultent plus fréquemment les services d'urgence, faute d'un suivi médical structuré. Cette variable issue de l'ESCC est catégorielle, avec les modalités initiales suivantes : *école secondaire non terminée, diplôme d'études secondaires terminées, CÉGEP, certificat avant baccalauréat, baccalauréat, certificat post-baccalauréat, études supérieures (post-baccalauréat)*. Cependant nous avons regroupé certaines modalités pour simplifier les analyses. Ainsi les nouvelles modalités de la variable « Niveau d'éducation » sont : *école secondaire non terminée, diplôme d'études secondaires terminées, CÉGEP / certificat avant baccalauréat et baccalauréat / certificat après baccalauréat / études supérieures (post-baccalauréat)*.
- **Statut matrimonial** : l'état matrimonial (marié[e], en union libre, séparé[e] / divorcé[e], veuf[-ve], célibataire) est une variable importante, car les personnes vivant seules ont souvent moins de soutien social, ce qui peut conduire à un plus grand recours aux soins hospitaliers (Chiu et al., 2022). Également issues de la base de données de l'ESCC, ces modalités initiales étaient : *marié(e), en union libre, séparé(e), veuf(-ve), divorcé(e), célibataire, jamais marié(e)*. Afin de faciliter l'analyse statistique et de refléter des situations sociales comparables, nous avons regroupé certaines de ces modalités pour former quatre catégories : *Marié(e), En union libre, Séparé(e) / veuf(-ve) / divorcé(e) et Célibataire / jamais marié(e)*. Ce regroupement permet de distinguer les individus en couple (formellement ou de fait) des personnes ayant vécu une rupture ou un deuil, ainsi que des personnes n'ayant jamais été en union. Il s'aligne sur des pratiques courantes dans la littérature (Kouakou et al., 2025; Wiens et al., 2021), où le statut matrimonial est souvent

utilisé comme indicateur de soutien social ou de vulnérabilité psychosociale pouvant influencer les comportements de santé et le recours aux services.

- **Revenu annuel** : cette variable issue de l'ESCC est catégorisée en cinq tranches (moins de 20 000 \$, 20 000-40 000 \$, 40 000-60 000 \$, 60 000-80 000 \$, plus de 80 000 \$). Les personnes à faible revenu sont plus vulnérables aux maladies chroniques et rencontrent davantage de barrières financières pour accéder aux soins primaires, augmentant ainsi leur probabilité de devenir de GU de soins (Wong et al., 2024).

Facteurs cliniques

Les caractéristiques médicales et l'état de santé des individus sont des éléments déterminants de la consommation des soins.

- **État de santé autoévalué** : la variable « état de santé autoévalué », issue de l'ESCC, reposait initialement sur cinq modalités : mauvais, passable, bon, très bon et excellent. Afin de simplifier l'analyse et de regrouper les répondants selon des niveaux comparables de perception de leur santé, ces catégories ont été consolidées en trois groupes : mauvais / passable, bon, et très bon / excellent. Ce regroupement s'appuie sur la littérature, qui montre que les individus s'autoévaluant en mauvaise santé présentent une probabilité significativement plus élevée de recourir de façon intensive aux services de santé, notamment hospitaliers (Moe et al., 2021). L'état de santé perçu constitue ainsi un indicateur pertinent et validé de besoin en soins, tel que reconnu dans le modèle comportemental d'Andersen (Andersen, 1995).
- **État de santé mentale autoévaluée** : la variable « état de santé mentale autoévaluée », également issue de l'ESCC, utilise les mêmes catégories que l'état de santé général, à savoir : mauvais, passable, bon, très bon et excellent. Conformément à la méthode adoptée pour l'état de santé général, ces modalités ont été regroupées en trois catégories : Mauvais / passable, Bon et Très bon / excellent, afin de faciliter l'analyse statistique tout en conservant la pertinence clinique du gradient de perception. Cette variable permet de capter la perception subjective du bien-être psychologique de l'individu. Elle revêt une importance particulière dans le cadre de cette étude, puisque la littérature établit une association entre les

troubles de santé mentale et une utilisation accrue des services de santé (Anderson et al., 2018; Armoon et al., 2022). L'intégration de cette mesure contribue ainsi à une compréhension des trajectoires de soins, en particulier chez les GU.

- **Indice de comorbidité** : l'indice de comorbidité constitue un indicateur quantitatif permettant d'estimer la charge globale des maladies chroniques présentes chez un individu. Il s'agit d'une variable synthétique reflétant le cumul de conditions médicales diagnostiquées, souvent utilisée pour prédire le recours aux services de santé. Plus cet indice est élevé, plus le patient est susceptible de présenter des besoins complexes, de consulter fréquemment ou d'être hospitalisé à répétition. Chiu et al. (2022) ont démontré une forte association entre un niveau élevé de comorbidité et une utilisation intensive des services, en particulier hospitaliers (Chiu et al., 2022). Dans le cadre de cette étude, l'indice de comorbidité a été calculé à partir des données de diagnostic recueillies lors des hospitalisations survenues dans l'année précédant l'entrevue ESCC, en appliquant le système de pondération CH-CIM-9⁵ tel que défini par Simard, Sirois et Candas (2018). Il a été validé pour prédire la mortalité à 30 jours selon les classifications ICD-9 et ICD-10 (Simard et al., 2018). Cet indice issu de la base de données SRAP (base de données construite avec des données de l'ESCC, de la RAMQ et de l'ISQ) est intégré comme un facteur de besoin, conformément au modèle d'Andersen (Andersen, 1995), afin de mieux comprendre son rôle dans la trajectoire et la caractérisation des GU.
- **Nombre de médicaments distincts sur 12 mois** : cette variable est mesurée à partir du nombre de médicaments distincts consommés au cours des 12 derniers mois, tel que rapporté dans l'ESCC. Elle permet d'identifier les patients en situation de polymédication, généralement définie comme la prise concomitante de cinq médicaments ou plus (Hovstadius et Petersson, 2012; Sirois et al., 2019). Un seuil de dix médicaments ou plus est parfois utilisé pour désigner une polymédication

⁵ Le système de pondération CH-CIM-9 est un outil développé pour exploiter les données de diagnostics hospitaliers codés selon la Classification internationale des maladies, 9^e révision, version canadienne (CIM-9-CM). Ce système permet d'attribuer un poids spécifique à chaque diagnostic en fonction de sa gravité et de son impact potentiel sur des issues cliniques comme la mortalité. Il est utilisé notamment dans le cadre du calcul d'indices de comorbidité, comme ceux de Charlson et al. (1987) et Elixhauser et al. (1998), afin de mieux évaluer le fardeau de la maladie chez les patients.

excessive, associée à une charge médicamenteuse importante. La polymédication est reconnue comme un indicateur indirect de complexité clinique et constitue un facteur prédictif d'hospitalisations répétées et d'effets indésirables, notamment chez les personnes âgées ou atteintes de comorbidités multiples (Wong et al., 2024). Dans le cadre du modèle comportemental d'Andersen, cette variable est considérée comme un facteur de besoin (Andersen, 1995).

- **Nombre de médecins consultés** : variable continue issue des données de l'ESCC, elle quantifie le nombre de médecins consultés au cours des 12 derniers mois. Elle est utilisée comme indicateur de fragmentation des soins. Les patients consultant plusieurs médecins différents peuvent avoir un parcours de soins mal coordonné, augmentant leur dépendance aux services hospitaliers (Moe et al., 2021). Dans le cadre du modèle d'Andersen, cette variable est considérée comme un facteur facilitant, dans la mesure où elle reflète non seulement l'accessibilité aux soins, mais aussi la qualité de leur organisation (Andersen, 1995).
- **Présence d'un médecin de famille** : la variable « présence d'un médecin de famille », telle que définie dans cette étude, est dérivée de la variable « fournisseur habituel de soins de santé » de l'ESCC. Cette variable binaire indique si un individu déclare avoir un professionnel de la santé attribué, tel qu'un médecin de famille, qu'il consulte habituellement en cas de besoin de soins. L'inscription auprès d'un médecin de famille est reconnue comme un indicateur d'accès aux soins primaires ainsi que de continuité dans la prise en charge médicale, deux composantes essentielles de la qualité des soins. À l'inverse, l'absence d'un tel fournisseur a été associée à un recours accru aux services d'urgence, souvent en raison d'un manque de suivi régulier, d'une faible coordination des soins, ou d'une difficulté à obtenir des consultations en première ligne (Armoon et al., 2022; Hudon et al., 2016). Cette variable est classée parmi les facteurs facilitants dans le modèle comportemental d'utilisation des soins de santé d'Andersen, puisqu'elle reflète la capacité structurelle de l'individu à accéder à des soins continus et coordonnés dans le système de santé (Andersen, 1995).

- **Satisfaction de vie** : la variable « satisfaction de vie » permet d'évaluer le bien-être subjectif global de l'individu, au-delà des dimensions strictement cliniques ou fonctionnelles. Issue de l'ESCC, cette variable était initialement catégorisée selon les modalités suivantes : très satisfait(e), satisfait(e), ni satisfait(e) ni insatisfait(e), insatisfait(e), très insatisfait(e). Pour faciliter l'analyse, ces catégories ont été regroupées en trois niveaux : Très satisfait(e), satisfait(e), et insatisfait(e) / très insatisfait(e) / ni satisfait(e) ni insatisfait(e). Le niveau de satisfaction de vie constitue un indicateur pertinent dans l'analyse des GU, car un faible bien-être subjectif est associé à une forte utilisation des services de santé, notamment en lien avec une détresse psychologique ou des besoins non comblés (Anderson et al., 2018). Conformément au modèle comportemental d'Andersen, cette variable est considérée comme un facteur prédisposant, dans la mesure où elle influence la perception de la santé et les comportements de recours aux soins (Andersen, 1995).
- **Niveau de stress perçu** : la variable « niveau de stress perçu » permet d'évaluer la perception subjective du stress quotidien, un facteur reconnu pour ses effets sur la santé physique et mentale. Elle est issue de l'ESCC, où elle était initialement mesurée à l'aide des modalités suivantes : pas du tout stressé(e), pas très stressé(e), un peu stressé(e), assez stressé(e), très stressé(e), extrêmement stressé(e). Pour faciliter l'analyse et assurer une meilleure distribution statistique, nous avons regroupé ces modalités en trois catégories : pas du tout stressé(e), un peu/assez stressé(e), et très/extrêmement stressé(e). Ce regroupement est également cohérent avec la littérature (Kouakou et al., 2025), qui distingue généralement les niveaux faible, modéré et élevé de stress perçu. Le stress chronique est associé à une augmentation du recours aux soins, notamment en raison de son lien avec les troubles de santé mentale, la somatisation et une perception accrue des besoins médicaux (Tille et al., 2017). Dans le cadre du modèle comportemental d'Andersen, cette variable est considérée comme un facteur prédisposant, influençant à la fois la perception de la santé et les comportements de recours aux services (Andersen, 1995).

- **Indice de masse corporelle (IMC)** : l'IMC, calculé à partir du poids et de la taille autodéclarés, a été classé selon les seuils standards en vigueur : poids insuffisant (<18), poids normal (18–24,9), surpoids (25–29,9) et obésité (≥ 30). Cette catégorisation permet de capturer les niveaux de risque associés à l'excès de poids, notamment en lien avec les maladies chroniques. L'obésité est associée à une forte prévalence de comorbidités telles que le diabète, l'hypertension, les maladies cardiovasculaires et certains troubles musculosquelettiques. Elle constitue également un facteur associé aux hospitalisations répétées et de recours intensif aux services de santé (Wong et al., 2024). Dans le modèle comportemental d'Andersen, l'IMC est considéré comme un facteur de besoin, car il reflète un état de santé objectif, susceptible d'augmenter la demande en soins médicaux.

Facteurs comportementaux

Les habitudes de vie influencent directement l'état de santé et la consommation des soins.

- **Tabagisme** : cette variable distingue les fumeurs quotidiens, les fumeurs occasionnels et les non-fumeurs. La variable catégorielle « tabagisme » possède trois modalités : fumeurs quotidiens, fumeurs occasionnels et non-fumeurs, conformément à la classification utilisée dans l'ESCC. Le tabagisme est en effet associé à un risque accru de maladies chroniques, telles que les maladies cardiovasculaires, les maladies respiratoires, le cancer. Il constitue ainsi un déterminant important de la consommation de soins de santé, notamment en termes de consultations, d'hospitalisations et de prescriptions (Moe et al., 2021). Dans le cadre du modèle comportemental d'Andersen, le tabagisme est considéré comme un facteur prédisposant, car il reflète un comportement de santé susceptible d'influencer l'état de santé futur (Andersen, 1995).
- **Consommation d'alcool** : variable catégorielle, issue de l'ESCC, mesurée à partir de la fréquence de consommation d'alcool au cours des 12 derniers mois. Les modalités initiales étaient : aucune consommation, moins d'une fois par mois, une fois par mois, 2-3 fois par mois, une fois par semaine, 2-3 fois par semaine, 4-6 fois par semaine et tous les jours. Nous avons reclassé ces modalités en trois catégories, à savoir : Aucune consommation, Consommation occasionnelle (moins d'une fois

par mois / une fois par mois / 2-3 fois par mois / une fois par semaine) et Toutes les semaines (2-3 fois par semaine / 4-6 fois par semaine / tous les jours), afin d'évaluer l'impact de l'alcool sur l'utilisation des services de santé. Cette variable permet d'évaluer l'effet du comportement de consommation d'alcool sur l'utilisation des services de santé. La consommation excessive ou régulière d'alcool est associée à une augmentation du risque de complications médicales, de comorbidités et d'hospitalisations (Chiu et al., 2022). Dans le cadre du modèle d'Andersen, elle est considérée comme un facteur prédisposant.

2.3 Objectif 1

2.3.1 Population et période d'étude

Pour cet objectif, nous définissons la population à l'étude comme les personnes ayant répondu à l'ESCC de 2007-2008 et de 2009-2010. Nous définissons ensuite les GU comme les utilisateurs faisant partie du 5^e percentile de coûts les plus élevés en 2012 et comparons ce groupe aux autres répondants qui n'étaient pas des GU en 2012. Ce choix a permis de suivre l'évolution des coûts associés aux GU et aux non-GU sur une période d'étude de dix ans (2007 à 2016), incluant l'année de référence (2012), les cinq années précédentes (2007-2011) et les quatre années suivantes (2013-2016). L'année 2012 a été choisie comme année de référence, car une année trop proche des données d'entrée (2007-2008, 2009-2010) aurait limité la capacité de prédiction, tandis qu'une année trop éloignée aurait réduit la pertinence des données initiales. Par ailleurs, en utilisant uniquement des données antérieures (2007-2008, 2009-2010) pour la prédiction des GU, on s'assure que les facteurs étudiés sont indépendants du statut de GU en 2012 et non une conséquence de celui-ci.

2.3.2 Approche analytique

a. Analyse descriptive

Nous avons effectué des statistiques descriptives pour obtenir une vue d'ensemble des caractéristiques des deux groupes étudiés, à savoir les GU et les non-GU. Afin d'analyser les différences entre ces groupes, nous avons eu recours à différents tests statistiques,

sélectionnés en fonction de la nature des variables et des conditions d'application propres à chaque méthode.

Pour les comparaisons entre variables catégorielles, nous avons principalement utilisé le test χ^2 de Pearson, lorsque les effectifs observés dans les tableaux de contingence étaient suffisamment élevés. Ce test permet d'évaluer si la distribution des fréquences observées diffère significativement de celle attendue sous l'hypothèse d'indépendance. En cas de faibles effectifs (nombre de cellules inférieur à 5), nous avons appliqué le test exact de Fisher, plus approprié dans ce contexte.

En ce qui concerne les variables continues, la comparaison des moyennes entre deux groupes indépendants s'est faite à l'aide du test t de Student lorsque l'hypothèse de normalité était respectée. Toutefois, lorsque cette hypothèse n'était pas remplie, nous avons privilégié le test de Wilcoxon (ou test de Mann-Whitney), une méthode non paramétrique fondée sur le rang des observations plutôt que sur leur distribution.

b. Analyse empirique

Nous avons utilisé les forêts aléatoires pour caractériser les GU, plus spécifiquement pour prédire les facteurs associés aux GU.

Les forêts aléatoires constituent une méthode d'apprentissage supervisé largement utilisée pour les tâches de classification et de régression (Breiman, 2001). Cette technique repose sur la construction et la combinaison de multiples arbres de décision, ce qui permet d'améliorer la précision des prédictions par rapport aux modèles individuels et de limiter le risque de surapprentissage en réduisant la variance du modèle (*overfitting*).

L'équation mathématique d'une forêt aléatoire repose sur la combinaison de plusieurs arbres de décision, chacun construit à partir d'un sous-échantillon des données. Chaque arbre fait une prédiction, et la forêt aléatoire agrège ces prédictions pour produire le résultat final. L'équation mathématique sur laquelle s'appuient les forêts aléatoires pour la classification est la suivante :

Si nous considérons une forêt aléatoire composée de M arbres de décision, chaque arbre T_m produit une prédiction \hat{y}_m pour une observation donnée x (où x est un vecteur de caractéristiques d'une observation). La prédiction finale \hat{y} pour la forêt aléatoire est obtenue par un vote majoritaire parmi les prédictions des arbres :

$$\hat{y} = \text{Mode}(\{\hat{y}_m(x)\}_{m=1}^M)$$

Où $\hat{y}_m(x)$ est la prédiction de l'arbre T_m pour l'observation x , et le résultat final \hat{y} est la classe la plus fréquente parmi les M arbres.

Chaque arbre de décision T_m est construit à partir d'un sous-échantillon aléatoire des données et utilise un sous-ensemble aléatoire de variables à chaque nœud afin de déterminer la meilleure séparation des données. L'arbre peut être défini comme une fonction récursive qui partitionne l'espace des caractéristiques en sous-régions et assigne une prédiction \hat{y}_m à chaque sous-région. Cette partition est basée sur des critères tels que l'entropie pour la classification.

$T_m : X \rightarrow Y$, où X est l'espace des caractéristiques et Y est l'espace des prédictions.

L'importance des variables dans une forêt aléatoire est calculée en observant la réduction moyenne de l'impureté pour la classification chaque fois qu'une variable est utilisée pour scinder un nœud dans un arbre. Formellement, cela peut être exprimé comme la somme de la réduction d'impureté pondérée par la fréquence à laquelle une variable est sélectionnée.

- **Étapes analytiques**

- **Échantillonnage et sélection de variables :**

- **Échantillonnage d'autoamorçage (*bootstrapping*):** pour chaque arbre de décision, un échantillon d'autoamorçage est tiré aléatoirement à partir des données d'entraînement. L'échantillonnage d'autoamorçage, initialement proposé par Efron (1979), permet de créer plusieurs ensembles de données légèrement différents en répétant aléatoirement certains échantillons, tout en excluant d'autres (Efron, 1979).
- **Sélection aléatoire des variables :** à chaque nœud de décision, un sous-ensemble aléatoire de variables est choisi pour déterminer la meilleure séparation. Cet aspect, connu sous le nom de *feature bagging* (Ho, 1998), réduit la corrélation entre les arbres individuels, augmentant ainsi la diversité au sein de la forêt.

- **Agrégation des résultats :**

- **Classification** : dans les tâches de classification, chaque arbre émet une prédiction pour une classe, et la classe la plus fréquemment prédite est retenue comme la prédiction finale (Breiman, 2001). Cette approche a montré de meilleures performances que les arbres de décision simples, qui sont souvent sujets à l'instabilité et au surajustement (Cutler et al., 2012).
- **Régression** : les prédictions des différents arbres sont agrégées par moyenne afin d'obtenir une valeur finale. Cette méthode renforce la robustesse du modèle par rapport à l'utilisation d'un arbre de décision unique (Hastie et al., 2009).

- **Application dans le cadre de notre étude**

Dans le cadre de notre étude, les forêts aléatoires sont utilisées pour analyser les facteurs socio-économiques, démographiques, cliniques et comportementaux qui différencient les grands utilisateurs (GU) de soins étudiés de la population générale. En raison de leur capacité à gérer de grands ensembles de données avec de nombreuses variables explicatives et à modéliser des interactions complexes entre ces variables, les forêts aléatoires sont particulièrement bien adaptées à cette tâche. Leur capacité à réduire le biais et la variance simultanément (Hastie et al., 2009) en fait un outil essentiel pour explorer les caractéristiques influentes au sein de populations hétérogènes.

En outre, la mesure de l'importance des variables permet de hiérarchiser les facteurs les plus déterminants dans la différenciation entre les GU et la population générale. Cette capacité à extraire des variables significatives, combinée avec l'absence de nécessité de spécifier une forme fonctionnelle a priori (Breiman, 2001), renforce la pertinence des forêts aléatoires pour notre analyse. Nous avons réalisé des tests de robustesse en modifiant le seuil de classification des GU à 1 % et 10 %, en plus du seuil principal utilisé (5 %).

- **Avantages et limites des forêts aléatoires**

Les forêts aléatoires présentent plusieurs avantages documentés dans la littérature scientifique :

- **Robustesse et précision** : l'agrégation de nombreux arbres réduit la variance et améliore la précision globale du modèle. Cela a été démontré par Breiman (2001),

qui a indiqué que les forêts aléatoires surpassaient souvent d'autres techniques d'apprentissage supervisé dans diverses applications.

- **Réduction du surapprentissage** : grâce à la combinaison des arbres, les forêts aléatoires atténuent les risques de surapprentissage, même lorsque chaque arbre individuel est très complexe. Cette réduction de variance, tout en conservant un faible biais, est l'un des principaux atouts des forêts aléatoires (Biau, 2012).
- **Gestion des données manquantes et des variables non linéaires** : les forêts aléatoires sont capables de traiter les données manquantes et d'identifier des relations non linéaires entre les variables (Hastie et al., 2009).
- **Importance des variables** : un autre avantage clé des forêts aléatoires est la possibilité d'évaluer l'importance relative des variables prédictives. Cette mesure de l'importance des variables peut fournir des informations utiles pour la sélection des variables ou la compréhension du modèle (Genuer et al., 2010).

Bien que les forêts aléatoires présentent de nombreux avantages, certaines limitations doivent être prises en compte :

- **Complexité computationnelle** : construire et combiner un grand nombre d'arbres peut être coûteux en temps de calcul, en particulier pour de grands ensembles de données (Hastie et al., 2009). Cependant, l'optimisation algorithmique et les capacités accrues des machines actuelles atténuent cet inconvénient.
- **Interprétabilité** : les forêts aléatoires sont souvent considérées comme des « boîtes noires » car elles ne fournissent pas de règles claires ou facilement interprétables. Cela contraste avec les arbres de décision simples, qui produisent des règles explicites et faciles à comprendre (Louppe, 2014).
- **Sensibilité aux données déséquilibrées** : comme pour d'autres algorithmes, les forêts aléatoires peuvent avoir du mal à traiter les ensembles de données où les classes sont fortement déséquilibrées. Dans ces situations, il peut être nécessaire d'utiliser des techniques de rééchantillonnage ou d'ajustement du modèle (More et Rana, 2020).

2.4 Objectifs 2 et 3

2.4.1 Population et période d'étude

Pour l'analyse des trajectoires, nous avons considéré tous les répondants ayant participé à l'ESCC entre 2007 et 2016 et ayant, au moins une fois, fait partie des 5 % de la population de la cohorte ayant les coûts d'utilisation des ressources les plus élevés. En intégrant toutes les enquêtes de l'ESCC, nous maximisons à la fois la taille de l'échantillon et la représentativité de l'étude. Par conséquent, cette approche permet d'inclure des individus aux profils encore plus variés et d'améliorer la robustesse des analyses.

De plus, afin de mieux comprendre l'évolution des coûts de santé, nous avons suivi les individus de cette cohorte sur une période de 10 ans (2007-2016). Ce suivi prolongé a été retenu car, d'une part, une période trop courte n'aurait pas permis de détecter des tendances à long terme; d'autre part, ce choix permet d'évaluer si les GU restent dans cette catégorie au fil des années ou, au contraire, s'ils reviennent à une consommation plus modérée après une période donnée.

2.4.2 Approche analytique

Nous avons réalisé une analyse de classes latentes (ACL) pour classifier les trajectoires des personnes ayant été identifiées comme GU au moins une fois au cours de la période d'étude. Un modèle de régression logistique multinomiale a ensuite été estimé pour déterminer les facteurs influençant l'appartenance à ces classes.

a. Analyse en classes latentes (ACL)

L'ACL, ou *Latent Class Analysis (LCA)* en anglais, est une méthode statistique utilisée pour identifier des sous-groupes latents au sein d'une population à partir de variables observées. Elle permet de modéliser des groupes d'individus partageant des caractéristiques similaires en fonction de leurs réponses à des questions ou de leurs comportements observés (Collins et Lanza, 2009; Lanza et Rhoades, 2013; Nguena Nguenfack et al., 2020). Initialement développée en sciences sociales pour modéliser des phénomènes complexes, l'ACL est maintenant utilisée dans des disciplines variées, telles que la psychologie, la médecine, et le marketing (Goodman, 1974; McCutcheon, 1987).

Cette approche repose sur l'idée que les relations entre les variables observées peuvent être influencées par des facteurs latents non observables. Par exemple, en psychologie, l'ACL

permet d'étudier des troubles mentaux ou des traits de personnalité en identifiant des profils latents parmi les individus (Muthén, 2001). Des études récentes, comme celle de Chiu et al. (2022), ont également utilisé cette méthode pour analyser les profils d'utilisateurs fréquents des services d'urgence en fonction de leurs comorbidités et de leurs comportements de recours aux soins.

L'ACL est particulièrement adaptée pour analyser des données catégorielles ou ordinales, c'est-à-dire des variables dont les modalités sont soit discrètes (par exemple, réponses « oui/non »), soit ordonnées (par exemple, échelles de satisfaction) (Collins et Lanza, 2009; Lanza et al., 2007; Lanza et Cooper, 2016). Si les variables comportent trop de catégories, il est recommandé de les recoder en un nombre de catégories aussi réduit que possible (Ryoo et al., 2018). Cette méthode est largement utilisée dans divers domaines, notamment :

- **Données cliniques** : par exemple, Armoon et al. (2022) ont utilisé l'analyse de classes latentes pour classer les patients atteints de troubles mentaux selon leur fréquence d'utilisation des urgences sur plusieurs années. L'étude a révélé trois profils distincts : les utilisateurs occasionnels, récurrents et très fréquents (Armoon et al., 2022). Chaque groupe présentait des caractéristiques spécifiques en matière de comorbidités, de comportements suicidaires et de précarité sociale, avec un risque de mortalité différencié selon le profil (Armoon et al., 2022).
- **Données d'enquête** : elle peut aussi être appliquée pour analyser des réponses à des enquêtes psychométriques ou à des questionnaires d'attitudes et de comportements. Par exemple, elle est souvent utilisée pour segmenter des individus selon leurs traits de personnalité ou troubles mentaux (Collins et Lanza, 2009).

La flexibilité de l'ACL permet d'inclure des covariables supplémentaires, telles que des facteurs environnementaux ou des variables modératrices, pour explorer les influences externes sur l'appartenance aux classes latentes (Collins et Lanza, 2009; Lanza et Rhoades, 2013).

L'ACL suit un processus structuré en plusieurs étapes. L'analyse débute par la définition du problème et la spécification du nombre de classes latentes à identifier. Ce choix repose sur des hypothèses théoriques, sur les spécificités du domaine étudié et sur des tests

statistiques permettant d'assurer la pertinence des sous-groupes identifiés (Collins et Lanza, 2009; Nguena Nguetack et al., 2020). Vient ensuite la spécification du modèle, qui implique la définition des variables observées et latentes ainsi que des hypothèses sous-jacentes. L'ajout éventuel de covariables permet d'affiner l'analyse et d'améliorer l'interprétation des résultats (Collins et Lanza, 2009; Lanza et Rhoades, 2013). L'estimation des paramètres repose sur des méthodes d'optimisation telles que l'algorithme espérance-maximisation (EM) ou le maximum de vraisemblance (ML). Ces techniques permettent d'obtenir les probabilités d'appartenance des individus aux différentes classes latentes, facilitant ainsi leur classification (Nguena Nguetack et al., 2020). Enfin, la sélection du modèle optimal s'appuie sur des critères statistiques comme l'*Akaike Information Criterion* (AIC) et le *Bayesian Information Criterion* (BIC). Une fois le modèle retenu, l'interprétation des classes se base sur l'analyse des caractéristiques des individus et des probabilités d'appartenance, garantissant ainsi la robustesse des résultats (Nguena Nguetack et al., 2020; Nylund et al., 2007).

L'ACL présente plusieurs avantages majeurs. Elle permet de segmenter des populations hétérogènes en groupes homogènes non observés initialement, facilitant ainsi l'identification de sous-groupes distincts (Collins et Lanza, 2009; Nguena Nguetack et al., 2020; Schreiber, 2017), comme l'a montré l'étude de Chiu et al. (2022) sur les profils des patients fréquentant les services d'urgence. De plus, l'ACL se distingue par sa flexibilité et sa robustesse : elle s'applique aussi bien à des données qualitatives que quantitatives, et permet de gérer efficacement les données manquantes sans introduire de biais significatifs (Collins et Lanza, 2009). Un autre atout important est la possibilité d'explorer les covariables, qui permettent d'identifier des facteurs influençant l'appartenance aux classes latentes (Collins et Lanza, 2009; L. McCutcheon, 1987; Nguena Nguetack et al., 2020).

Cependant, l'ACL présente certaines limites. Sa complexité augmente avec le nombre de classes ou de covariables, ce qui complique la sélection du nombre optimal de classes et induit un risque de surajustement ou de difficulté d'interprétation des résultats (Collins et Lanza, 2009; Nylund et al., 2007). En outre, les classes latentes doivent être interprétées avec prudence, car elles ne correspondent pas toujours à des entités tangibles ou clairement définies dans le monde réel (Hagenaars et McCutcheon, 2002). Pour atténuer ces limites et maximiser l'interprétabilité clinique des trajectoires identifiées, il est recommandé de

mobiliser un groupe d'experts pour attribuer des étiquettes aux profils émergents. Cette approche est soutenue par Bellemare-Lepage et al. (2023), qui soulignent l'importance de l'interprétation substantielle des classes latentes, notamment en psychologie, pour assurer la validité et la pertinence des résultats obtenus (Bellemare-Lepage et al., 2023).

- **Application dans le cadre de notre étude**

L'ACL a été utilisée pour modéliser les trajectoires des individus en fonction du maintien ou du changement de leur statut de GU d'une année à l'autre (variables dichotomiques, 10 points de suivi représentant des années). L'objectif était d'identifier des profils de trajectoires en se basant sur le maintien, la perte ou l'obtention du statut au cours du suivi. La variable d'intérêt correspond au statut annuel de GU ou non-GU attribué à chaque individu au cours de la période étudiée. Chaque année, un individu peut conserver son statut de GU, le perdre ou l'obtenir, permettant ainsi de modéliser les différentes trajectoires possibles au fil du temps.

La modélisation s'est déroulée en plusieurs étapes. Dans un premier temps, nous avons construit la base de données en attribuant à chaque individu une séquence de statuts annuels. Ces statuts ont ensuite été traités comme des états distincts dans l'ACL afin d'analyser les dynamiques de transition entre GU et non-GU. Aucun facteur explicatif n'a été intégré dans cette première analyse, afin de se concentrer exclusivement sur la structure des trajectoires. Le choix du nombre optimal de classes a été guidé par des critères statistiques tels que l'AIC et le BIC (plus faible valeur absolue = plus souhaitable), ainsi que par une interprétation empirique des profils obtenus. Un autre critère méthodologique important, à savoir la présence d'au moins 5 % de l'échantillon dans chaque classe, a également été pris en compte. Ce critère a été respecté pour l'ensemble des classes, à l'exception de la Classe 6, qui représente 4 % de l'échantillon. Cette classe a néanmoins été conservée en raison de sa forte cohérence interne et de son intérêt clinique particulier, puisqu'elle regroupe exclusivement des individus ayant été GU de manière très récurrente entre 5 et 10 années sur la période d'étude. Ce seuil vise à garantir la stabilité et la fiabilité des estimations pour chaque classe identifiée. Toutefois, il est reconnu que ce critère n'est pas absolu et peut être adapté en fonction du contexte spécifique de l'étude (Bellemare-Lepage et al., 2023). Par exemple, pour des raisons théoriques ou pratiques, il peut être justifié de conserver une classe représentant un sous-groupe cliniquement pertinent, même

si sa taille est inférieure au seuil conventionnel. Cette flexibilité est soulignée dans la littérature méthodologique, qui encourage les chercheurs à considérer la pertinence substantielle des classes en plus des critères purement statistiques (Bellemare-Lepage et al., 2023).

Cette approche a permis d'identifier des classes distinctes de trajectoires, reflétant différentes évolutions du statut de GU. Certaines classes regroupent des individus conservant leur statut de GU sur plusieurs années, tandis que d'autres montrent des profils marqués par une sortie progressive ou rapide de ce statut. Cette classification permet ainsi de mieux comprendre les dynamiques d'évolution du statut de GU au fil du temps.

b. Régression logistique multinomiale

Nous avons estimé un modèle de régression logistique multinomiale pour identifier les facteurs influençant l'appartenance aux différentes classes définies par l'ACL. L'ACL ayant permis de déterminer six classes distinctes, notre variable d'intérêt constitue la variable catégorielle *Classe*, qui comporte six modalités. Ce type de modèle est particulièrement adapté lorsque la variable dépendante est catégorielle avec plus de deux modalités (Fagerland et Hosmer, 2012). Contrairement à la régression logistique binaire, la régression logistique multinomiale permet d'estimer simultanément les probabilités d'appartenance à plusieurs catégories, en utilisant une catégorie de référence pour comparer les autres modalités (Greene, 2012).

- **Spécification du modèle**

Soit Y la variable d'intérêt représentant la classe d'appartenance des individus, avec K modalités ($k = 1, \dots, K$). La probabilité conditionnelle qu'un individu appartienne à la catégorie k , étant donné un vecteur de covariables X , est exprimée comme suit :

$$P(Y = k | X) = \frac{\exp(\beta_{k0} + X\beta_k)}{1 + \sum_{j=1}^{K-1} \exp(\beta_{j0} + X\beta_j)}$$

Où β_k représente le vecteur des coefficients associés aux variables explicatives pour la catégorie k , et β_{k0} est l'ordonnée à l'origine. L'une des catégories est choisie comme catégorie de référence et les coefficients des autres catégories sont estimés par rapport à celle-ci (Agresti, 2013).

- **Méthode d'estimation**

Les paramètres du modèle ont été estimés par la méthode du maximum de vraisemblance, qui consiste à trouver les valeurs de coefficients qui maximisent la probabilité d'observer les données disponibles (McCullagh, 2019). L'évaluation de la qualité du modèle a été réalisée à l'aide du pseudo R^2 de McFadden, qui permet d'estimer la proportion de variance expliquée, ainsi que par le test du rapport de vraisemblance, qui permet de comparer la qualité d'ajustement des modèles (Menard, 2001).

- **Validation du modèle**

Afin d'évaluer la robustesse du modèle, nous avons procédé à une analyse des résidus et vérifié l'absence de colinéarité entre les variables explicatives en calculant les facteurs d'inflation de la variance (VIF) (Kutner et al., 2005). De plus, la stabilité des estimations a été testée à l'aide de la technique de validation croisée, permettant de s'assurer que les résultats obtenus ne sont pas sensibles aux variations de l'échantillon d'entraînement et de test (James et al., 2023).

Nous avons utilisé le logiciel Stata[®] pour le traitement des données, le logiciel R a servi pour les autres analyses. Le progiciel PoLCA a servi pour l'ACL (Linzer et Lewis, 2011). Les variables explicatives ont été sélectionnées a priori sur la base du modèle comportemental d'utilisation des soins de santé d'Andersen (1995), lequel distingue les facteurs prédisposants, facilitants et de besoin. Cette approche théorique a guidé la construction initiale du jeu de variables, assurant une cohérence conceptuelle dans l'identification des déterminants potentiels de l'appartenance aux classes latentes.

Dans un second temps, pour le modèle de régression multinomiale, les variables finales ont été retenues selon une procédure de sélection ascendante basée sur le critère d'information d'Akaike (AIC par degrés). Cette méthode permet d'optimiser l'ajustement du modèle tout en limitant le risque de surajustement. Elle offre une alternative méthodologiquement plus robuste que les approches fondées uniquement sur les seuils de significativité en analyses bivariées, tout en préservant la validité théorique du modèle.

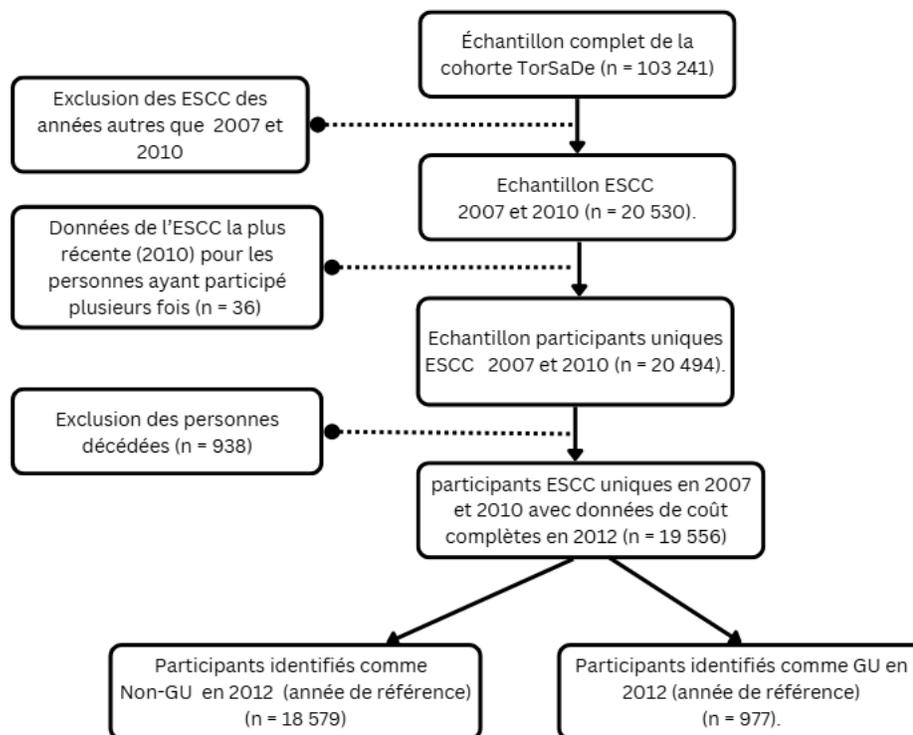
3. Résultats

3.1 Objectif 1

3.1.1 Description de l'échantillon

La figure 3 ci-dessous représente le diagramme de flux des participants. L'échantillon total de la cohorte TorSaDe compte un peu plus de 103 000 entrées d'ESCC, dont 1 093 ont été fournies par des individus ayant répondu à l'ESCC plusieurs fois. En ne conservant que les entrées de l'ESCC de 2007-2008 et de 2009-2010, nous avons 20 530 individus, dont 36 ont répondu à l'ESCC à deux reprises. Pour ces derniers, nous avons utilisé les données de l'ESCC la plus récente (2009-2010). Nous avons exclu les personnes décédées durant notre période d'étude, soit 938 individus, ce qui a permis d'obtenir 19 556 entrées uniques d'ESCC au total. Nous avons ensuite constitué deux groupes en appliquant notre critère d'identification des GU : un avec les GU 2012 ($n = 977$) et un autre avec les non-GU 2012 ($n = 18 579$).

Figure 3 : Diagramme de flux des participants

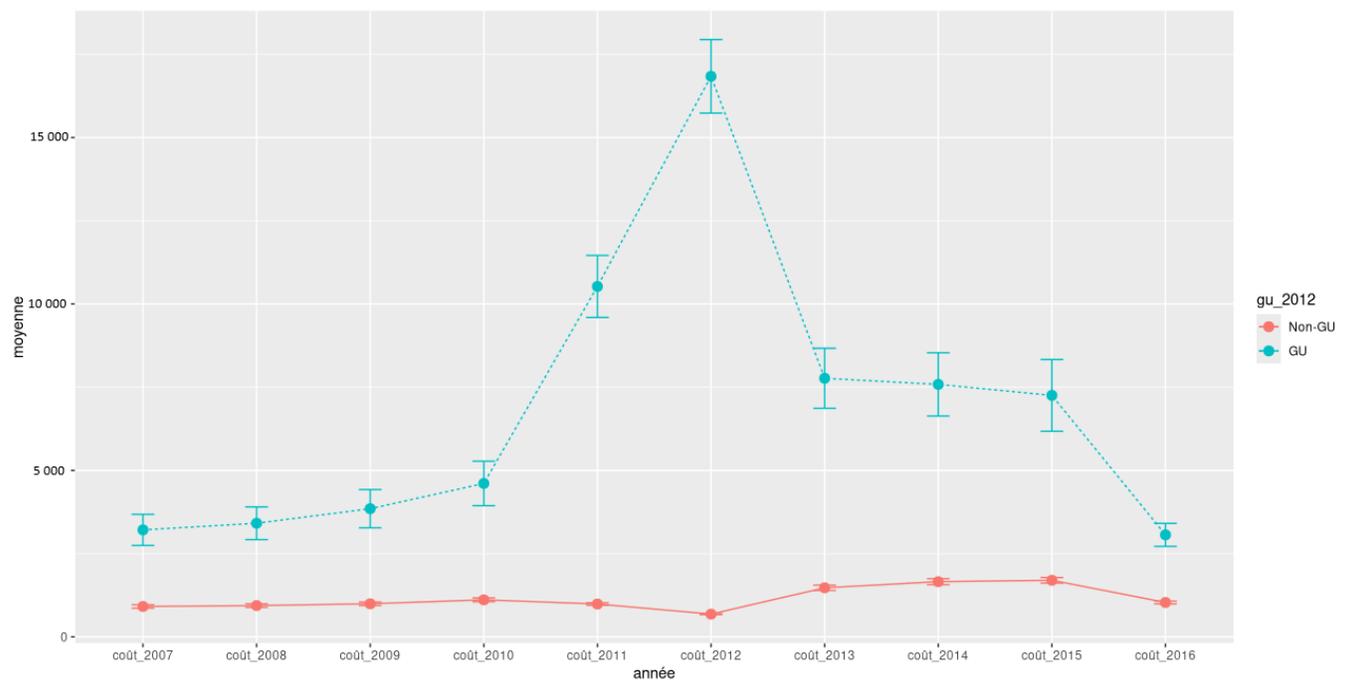


Source : auteurs

3.1.2 Distribution des coûts

L'analyse des coûts des GU et non-GU identifiés en 2012, entre 2007 et 2016, met en évidence trois grandes tendances : une différence de coûts déjà marquée avant 2012, un pic exceptionnel en 2012 (inhérent à la méthode) et une baisse progressive des coûts après 2012.

Figure 4 : Évolution des coûts moyens des GU 2012 sur la période de 2007 à 2016



Source : auteurs, analyses des données TorSaDe

Avant 2012, les GU 2012 présentait déjà des coûts moyens nettement supérieurs à ceux des non-GU 2012. En 2007, le coût moyen des GU 2012 était de 3 213,76 \$ ± 467,96 \$ contre 912,04 \$ ± 52,64 \$ pour les non-GU (annexe 1). Cette tendance s'amplifie progressivement jusqu'en 2011, où les coûts des GU 2012 atteignent 10 526,43 \$ ± 932,19 \$, alors que les non-GU 2012 restent à 985,59 \$ ± 37,50 \$. Cette différence se reflète également dans les intervalles de confiance (IC à 95 %) du graphique des coûts moyens (figure 4), où les barres d'erreur sont plus longues pour les GU 2012,

indiquant une plus grande variabilité des coûts. En revanche, les IC à 95 % des non-GU 2012 sont très courts, indiquant une faible dispersion des valeurs. Cela s'explique en partie par le nombre beaucoup plus élevé de non-GU.

L'année 2012 marque une rupture, avec une hausse des coûts des GU 2012, qui atteignent 16 840,07 \$ ± 1 103,93 \$, soit une augmentation de 60 % par rapport à 2011 (annexe 1). Ce pic est accompagné d'une forte augmentation de la variabilité des coûts, visible à travers les IC à 95 % nettement plus larges (figure 4) en 2012, ce qui indique une plus grande dispersion des valeurs. Cette dispersion est confirmée par les boîtes à moustache (annexe 2), qui montrent une boîte allongée et un grand nombre de valeurs aberrantes, ainsi que par le graphique en courbes individuelles, où plusieurs GU 2012 enregistrent des coûts extrêmement élevés, dépassant 300 000 \$ pour certains. À l'inverse, on remarque une légère baisse pour les non-GU 2012 en 2012.

Après 2012, les coûts des GU 2012 diminuent progressivement (annexe 2), passant à 7 768,22 \$ ± 901,17 \$ en 2013 et atteignant 3 064,04 \$ ± 346,51 \$ en 2016. Cependant, les IC à 95 % restent plus larges pour les GU 2012 que pour les non-GU 2012 (figure 5), ce qui montre que la variabilité des coûts persiste, bien que réduite par rapport à 2012. Les boîtes à moustache (annexe 2) et les courbes individuelles (annexe 3) illustrent cette tendance.

Les valeurs de p (annexe 1) associées à la comparaison des coûts entre GU 2012 et non-GU 2012 sont toutes inférieures à 0,05 et souvent proches de 0, indiquant que les différences observées sont statistiquement significatives chaque année.

3.1.3 Description des données de l'échantillon des grands utilisateurs et échantillon de la cohorte TorSaDe

Au total, 19 556 participants ont été inclus dans notre échantillon, dont 977 (4,99 %) sont considérés comme des GU en 2012 (année de référence), tandis que 18 579 (95,01 %) ne le sont pas (non-GU).

Les résultats du tableau 1 montrent des différences significatives entre les GU et les non-GU en termes de caractéristiques sociodémographiques, économiques et cliniques. Les GU sont en moyenne plus âgés que les non-GU (64,2 ans contre 49,2 ans) et présentent

une proportion légèrement plus élevée de femmes (55,99 % contre 53,90 %). L'âge moyen⁶ est également plus élevé chez les GU (4 476,2 contre 2 789,9 chez les non-GU). Les GU présentent un taux plus élevé de personnes n'ayant pas terminé leurs études secondaires (42,75 % contre 26,49 % pour les non-GU) et sont sous-représentés dans les niveaux d'éducation supérieure (12,95 % contre 20,08 %). Les GU sont plus nombreux à se situer dans les tranches de revenus annuels les plus bas, avec 46,00 % des GU ayant un revenu inférieur à 20 000 \$, contre 35,45 % pour les non-GU.

Les GU évaluent leur santé générale et mentale de manière moins favorable. Une mauvaise perception de la santé générale est rapportée par 28,82 % des GU contre 9,51 % des non-GU, et une mauvaise santé mentale par 7,47 % des GU contre 3,68 % des non-GU. Les GU présentent également une charge médicale plus élevée, comme en témoigne un indice de comorbidité moyen supérieur (0,8 contre 0,2), un nombre moyen de médecins différents consultés plus élevé (2,0 contre 1,0), ainsi qu'un nombre moyen de médicaments distincts plus élevé sur une période de 12 mois (10,0 contre 5,2). En se basant sur les seuils définis dans la littérature, les GU atteignent en moyenne le seuil de la polymédication excessive (≥ 10 médicaments), tandis que les non-GU franchissent celui de la polymédication standard (≥ 5 médicaments) (Hovstadius et Petersson, 2012; Sirois et al., 2019). Par ailleurs, les taux d'obésité et de tabagisme quotidien sont plus élevés chez les GU (20,87 % contre 16,44 % et 20,37 % contre 18,18 %, respectivement). Les GU sont également plus nombreux à rapporter une absence de consommation d'alcool (24,72 % contre 16,99 %). La proportion de personnes ayant un médecin de famille est aussi plus élevée chez les GU (88,13 %) que chez les non-GU (77,44 %). Les tests de robustesse ont montré que les groupes ne sont pas semblables selon le seuil de classification des GU (1 %, 5 %, 10 %), avec des différences significatives dans certaines caractéristiques sociodémographiques et cliniques (voir résultats aux annexes 5 et 6).

Tableau 1 : Description des données des GU, non-GU et échantillon de la cohorte TorSaDe

⁶ L'effet de l'âge sur certaines variables d'intérêt peut être non linéaire et l'ajout de l'âge² permet de mieux rendre compte de ces dynamiques dans les données. Dans un modèle analysant la probabilité d'être un grand utilisateur de soins de santé, l'impact de l'âge pourrait être plus important à des âges jeunes (par exemple, à cause de certaines maladies chroniques précoces) et diminuer à mesure que l'individu devient plus âgé, puis réaugmenter (à cause de la diminution des activités physiques ou de la sédentarité après un certain âge), ce qui justifie l'ajout de l'âge au carré.

Variable	Non-GU N = 18 579 ¹	GU N = 977 ¹	Échantillon TorSaDe N = 19 556 ¹	Valeur de p ²
Sexe				0,3
Femme	53,90 %	55,99 %	54,00 %	
Homme	46,10 %	44,01 %	46,00 %	
Âge en 2012, années	(49,2)	(64,2)	(49,9)	< 0,001
Perception de la santé générale				< 0,001
Mauvaise / Passable	9,51 %	28,82 %	10,47 %	
Bonne	30,96 %	35,38 %	31,18 %	
Très bonne / Excellente	59,53 %	35,79 %	58,35 %	
Perception de la santé mentale				< 0,001
Mauvaise / Passable	3,68 %	7,47 %	3,86 %	
Bonne	19,90 %	23,05 %	20,05 %	
Très bonne / Excellente	76,43 %	69,47 %	76,08 %	
Type de fumeurs				0,024
Tous les jours	18,18 %	20,37 %	18,29 %	
Fumeur occasionnel	5,01 %	3,38 %	4,93 %	
Non-fumeur	76,81 %	76,25 %	76,78 %	
Consommation d'alcool				< 0,001
Aucune consommation	16,99 %	24,72 %	17,38 %	
Occasionnelle (moins d'une fois par mois / une fois par mois / 2-3 fois par mois / une fois par semaine)	48,84 %	49,23 %	48,86 %	
Toutes les semaines (2-3 fois par semaine / 4-6 fois par semaine / tous les jours)	34,17 %	26,05 %	33,76 %	
Indice de masse corporelle, kg/m²				< 0,001
Poids insuffisant (IMC < 18)	2,61 %	3,35 %	2,65 %	
Poids normal (18 ≤ IMC < 25)	46,90 %	38,50 %	46,47 %	
Surpoids (25 ≤ IMC < 30)	34,04 %	37,28 %	34,21 %	
Obésité (30 ≥ IMC)	16,44 %	20,87 %	16,67 %	
Niveau d'éducation				< 0,001
École secondaire non terminée	26,49 %	42,75 %	27,30 %	
Diplôme d'études secondaires terminées	11,96 %	10,07 %	11,87 %	
CÉGEP/ certificat avant bac	41,48 %	34,22 %	41,11 %	
Bac/ certificat après bac / études supérieures (post-bac)	20,08 %	12,95 %	19,72 %	
Statut matrimonial				< 0,001
Marié(e)	30,89 %	35,45 %	31,12 %	
Union libre	18,90 %	11,68 %	18,54 %	
Séparé(e) / veuf(-ve) / divorcé(e)	17,91 %	35,45 %	18,79 %	
Célibataire / jamais marié(e)	32,30 %	17,42 %	31,56 %	
Revenu annuel par répondant				< 0,001
Moins de 20 000 \$	35,45 %	46,00 %	35,99 %	
20 000 à 40 000 \$	31,42 %	31,46 %	31,42 %	
40 000 à 60 000 \$	19,02 %	14,19 %	18,77 %	

60 000 à 80 000 \$	8,34 %	4,69 %	8,15 %	
80 000 \$ et +	5,78 %	3,66 %	5,67 %	
Pays de naissance				0,008
Canada	92,40 %	94,68 %	92,52 %	
Extérieur du Canada	7,60 %	5,32 %	7,48 %	
Lieu de résidence				< 0,001
Rural	56,81 %	63,05 %	57,12 %	
Urbain	43,19 %	36,95 %	42,88 %	
Indice de l'état de santé (HUI3)	(0,9)	(0,8)	(0,9)	< 0,001
Satisfaction de la vie en générale				0,026
Très satisfait(e)	41,39 %	37,75 %	41,20 %	
Satisfait(e)	51,64 %	52,37 %	51,68 %	
Insatisfait(e) / très insatisfait(e) / ni satisfait(e) ni insatisfait(e)	6,97 %	9,88 %	7,12 %	
Perception du stress – Vie quotidienne				< 0,001
Pas du tout stressé(e)	16,07 %	25,98 %	16,57 %	
Un peu stressé(e) / assez stressé(e)	60,30 %	56,11 %	60,09 %	
Très stressé(e) / extrêmement stressé(e)	23,63 %	17,91 %	23,34 %	
A un médecin de famille				< 0,001
Oui	77,44 %	88,13 %	77,97 %	
Non	22,56 %	11,87 %	22,03 %	
Indice de comorbidité	(0,2)	(0,08)	(0,2)	< 0,001
Nombre de médicaments distincts sur 12 mois	(5,2)	(10,0)	(5,6)	< 0,001
Nombre de médecins différents consultés sur 12 mois	(1,00)	(2,00)	(1,00)	< 0,001

¹ %; Moyenne ()

² Test exact de Fisher; test de Wilcoxon; test χ^2 de Pearson

GU : grand utilisateur; **non-GU** : non-grand utilisateur; **IMC** : indice de masse corporelle

Urbain : Montréal, Laval, Montérégie, Capitale-Nationale, Outaouais

Rural : Saguenay-Lac-Saint-Jean, Chaudière-Appalaches, Abitibi-Témiscamingue, Bas-Saint-Laurent, Côte-Nord, Gaspésie/Îles-de-la-Madeleine, Nord-du-Québec, Mauricie/Centre-du-Québec, Estrie, Laurentides, Lanaudière.

Extérieur du Canada : Autres pays de l'Amérique du Nord, Amérique centrale Antilles et Bermudes, Europe, Afrique.

Source : auteurs, analyses des données TorSaDe avec R

3.1.4 Caractéristiques des grands utilisateurs de soins

L'analyse par forêt aléatoire (figure 5) a permis d'identifier les principales variables influençant la classification des GU et de mettre en évidence des facteurs démographiques, cliniques, socio-économiques et comportementaux.

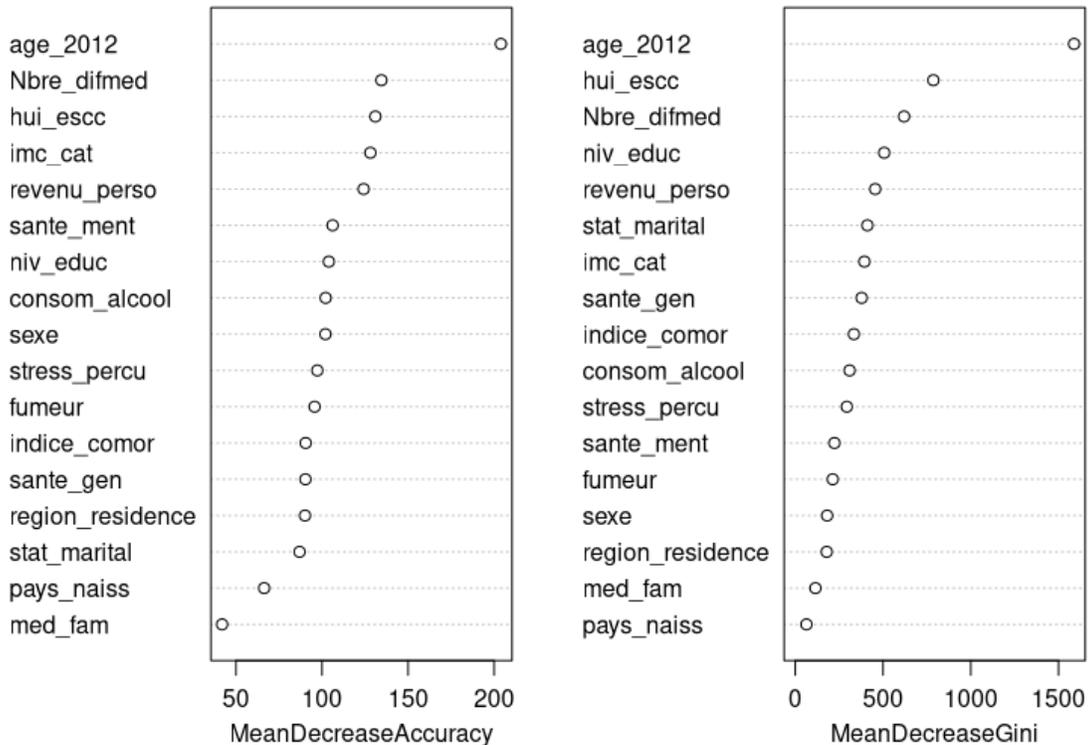
La figure 5 présente deux graphes qui permettent d'évaluer l'importance des variables utilisées dans le modèle. Le premier graphe, à gauche, repose sur la mesure de la réduction

de la précision moyenne (*mean decrease accuracy*) du modèle lorsque chaque variable est exclue. Une valeur élevée indique que la variable a une forte influence sur la précision de la classification. Le second graphe, à droite, repose sur la réduction moyenne de Gini (*mean decrease Gini*), un indicateur issu des arbres de décision qui mesure la capacité d'une variable à améliorer la séparation des classes. Plus cette valeur est élevée, plus la variable contribue à la discrimination entre les groupes.

Parmi les facteurs démographiques, l'âge constitue la variable la plus déterminante. Les facteurs cliniques jouent également un rôle majeur dans la classification des GU, ainsi que le nombre de médecins différents consultés au cours des 12 derniers mois. De plus, l'indice de l'état de santé (HUI3), l'indice de masse corporelle (IMC) et l'état de santé mentale sont des variables significativement associées à la classification des individus comme GU. Les facteurs socio-économiques sont associés à la probabilité d'appartenance au groupe des GU, notamment le niveau d'éducation et le revenu personnel. Parmi les facteurs comportementaux, la consommation d'alcool contribue à la classification des individus dans la catégorie des GU.

Figure 5 : Résultats de la prédiction des facteurs associés aux GU

Forêt aléatoire - GU



Source : auteurs, analyses des données TorSaDe

Les graphiques de l'annexe 4 présentent l'évolution des erreurs du modèle de forêt aléatoire en fonction du nombre d'arbres, aussi bien sur l'ensemble d'entraînement que sur l'ensemble de test. Trois types d'erreurs ont été analysés : l'erreur totale, le taux d'erreur pour les non-GU et le taux d'erreur pour les GU.

Dans les résultats des données d'entraînement, on observe une diminution rapide des erreurs lorsque le nombre d'arbres augmente, atteignant une stabilisation autour de 100 arbres. Après ce seuil, l'erreur totale ainsi que les taux d'erreur pour les non-GU et les GU deviennent très faibles. L'erreur totale se stabilise en dessous de 0,02; le taux d'erreur pour les non-GU autour de 0,02 et le taux d'erreur pour les GU en dessous de 0,005 indiquant que le modèle s'ajuste bien aux données d'entraînement.

Par ailleurs, les résultats sur l'ensemble de test confirment une bonne généralisation du modèle, avec des erreurs rapidement stabilisées après 100 arbres (annexe 4). L'erreur globale reste inférieure à 0,02 ce qui indique l'absence de surapprentissage.

Ainsi, ces résultats montrent que le modèle atteint une bonne performance avec environ 100 arbres, garantissant un bon équilibre entre précision et complexité, avec un seuil d'erreur total inférieur à 0,02. Les résultats de la forêt aléatoire sont robustes aux variations du seuil de classification, les principaux déterminants identifiés demeurant similaires, tant sur le plan qualitatif que quantitatif (voir résultats aux annexes 7 et 8).

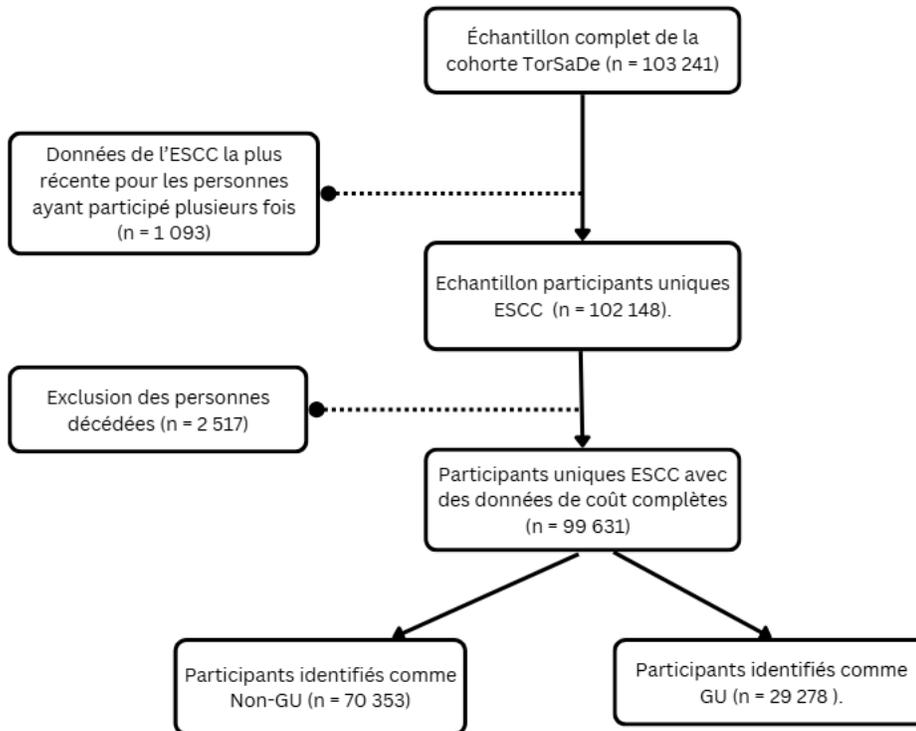
3.2 Objectifs 2 et 3

3.2.1 Description de l'échantillon utilisé

La figure 6 présente le diagramme de flux des participants. L'échantillon initial de la cohorte TorSaDe comprend 103 241 entrées issues de l'ESCC, correspondant à 102 148 participants uniques, certains individus ayant répondu à l'enquête à plusieurs occasions ($n = 1\,093$). Pour ces participants, seule l'entrée la plus récente a été conservée, conformément à une logique de non-redondance des observations.

Par la suite, 2 517 individus décédés durant la période d'étude ont été exclus, formant un échantillon final de 99 631 participants vivants au moment de l'analyse. En appliquant le critère d'identification des GU, deux sous-groupes ont été constitués : 29 278 participants ont été classés comme GU, tandis que 70 353 ont été classés comme non-GU.

Figure 6 : Flux des participants pour classifier les trajectoires



Source : auteurs

3.2.2 Résultat analyse de classes latentes

- **Critères d'information statistique pour différentes classes**

L'évaluation du nombre optimal de classes a été réalisée en analysant plusieurs critères statistiques, notamment l'AIC et le BIC, ainsi que les statistiques de vraisemblance G^2 et X^2 . Ces critères sont couramment utilisés pour sélectionner le meilleur modèle en tenant compte du compromis entre qualité d'ajustement et complexité du modèle (Burnham et Anderson, 2002).

Les résultats montrent une diminution progressive des valeurs de l'AIC et du BIC à mesure que le nombre de classes augmente, indiquant une amélioration de l'ajustement du modèle. L'analyse des différences successives (ΔAIC et ΔBIC) révèle que cette amélioration est significative jusqu'à six classes, après quoi les gains deviennent marginaux. En effet, entre cinq et six classes, la diminution de l'AIC est de 1 506,1 et celle du BIC de 1 414,9 ce qui

traduit un ajustement optimal avant que la réduction ne ralentisse brusquement au-delà de six classes.

Tableau 2 : Critères statistiques post-estimation

	AIC	BIC	Δ AIC	Δ BIC	G ²	Δ G ²	X ²	Δ X ²
2 classes	263 926,5	264 100,5	—	—	21 706,11	—	18 838,91	—
3 classes	263 469,4	263 734,5	457,1	366,0	21 227,00	479,11	470 864,99	451 026,08
4 classes	260 430,8	260 787,0	3 038,6	2 947,5	18 166,33	3 060,67	14 907,87	455 957,12
5 classes	259 014,9	259 462,2	1 415,9	1 324,8	16 728,45	1 437,88	13 649,48	1 258,39
6 classes	257 508,8	258 047,3	1 506,1	1 414,9	15 200,35	1 528,10	12 573,39	1 076,09
7 classes	257 024,5	257 654,1	484,3	393,2	14 694,03	506,32	11 951,27	622,12
8 classes	255 210,2	255 931,0	1 814,3	1 723,1	12 857,82	1 836,21	10 594,93	1 356,34

Source : auteurs, analyses des données TorSaDe

Les statistiques de vraisemblance G² et X² confirment cette tendance. Jusqu'à six classes, les valeurs de G² et X² diminuent de manière significative, indiquant une réduction progressive de l'erreur d'ajustement du modèle. Après six classes, cette diminution devient nettement plus faible, suggérant que l'ajout de classes supplémentaires n'apporte qu'une amélioration marginale (McLachlan et Peel, 2004).

Sur la base de ces résultats, le modèle avec six classes est retenu, offrant un équilibre optimal entre qualité d'ajustement et complexité du modèle. Ce choix est cohérent avec les recommandations méthodologiques en classification latente, qui préconisent de privilégier le modèle ayant le plus faible AIC et BIC, tout en évitant une augmentation excessive de la complexité (Nylund et al., 2007).

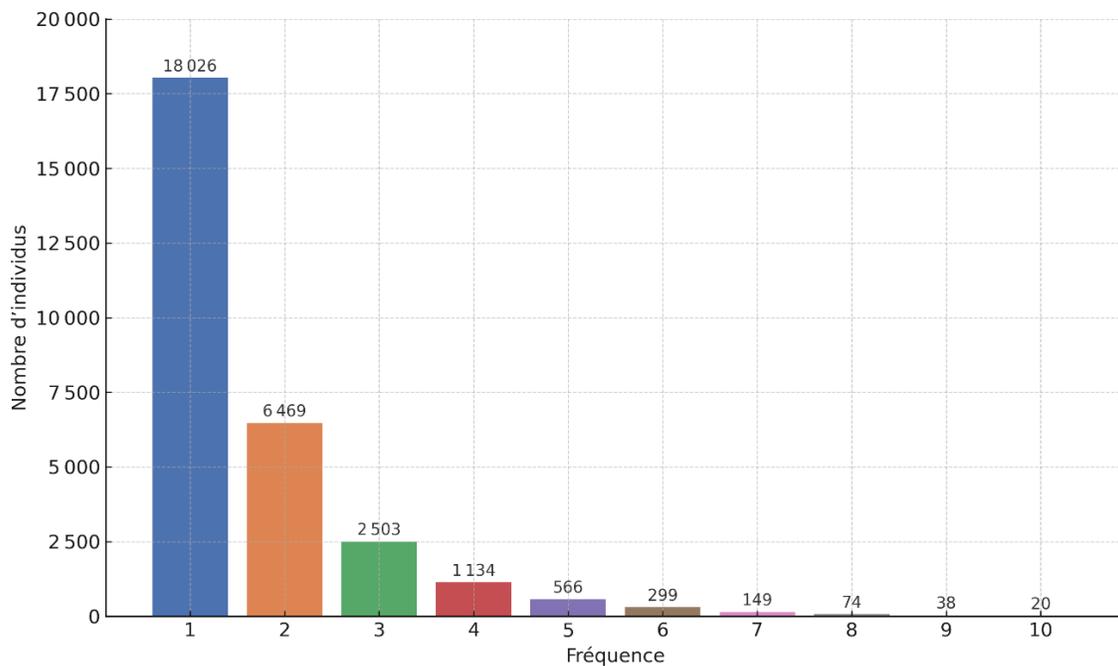
- **Caractérisation des différentes classes**

Sur les 102 241 individus de notre échantillon TorSaDe utilisés pour modéliser la trajectoire des GU, 29 278 individus (28,63 %) étaient des GU au cours d'une année au moins.

Les résultats de la figure 7 montrent que parmi les individus identifiés comme GU sur la période de suivi, la majorité d'entre eux ont été GU une seule fois ($N = 18\ 026$). Cette

proportion décroît progressivement avec l'augmentation de la fréquence d'appartenance au groupe des GU. Ainsi, 6 469 individus ont été GU deux fois, tandis que 2 503 individus l'ont été trois fois. La fréquence continue de diminuer avec 1 134 individus ayant été GU quatre fois et 566 individus cinq fois. Au-delà de six occurrences, le nombre d'individus devient marginal, avec 299 individus ayant été GU six fois, 149 sept fois, 74 huit fois, 38 neuf fois et seulement 20 individus ayant été GU dix fois.

Figure 7 : Nombre d'années pour lesquelles les individus sont des GU



Source : auteurs, analyses des données TorSaDe

Dans le tableau 3, la Classe 1 (N = 7 569, soit 26 % de l'échantillon) regroupe principalement des individus ayant été GU une seule fois (81 %), avec une faible proportion de GU récurrents. La Classe 3 (N = 2 940, soit 10 %) présente une distribution similaire avec 71,4 % des individus ayant été GU une seule fois et seulement 2,4 % ayant été GU entre 3 et 4 fois. En revanche, la Classe 2 (N = 6 228, soit 21 %) se distingue par une proportion plus importante de GU deux fois (33,5 %) et de GU entre 3 et 4 fois (22,4 %). La Classe 4 (N = 8 729, soit 30 %) suit un schéma similaire, avec 58 % des individus ayant été GU une seule fois et 25,1 % deux fois, mais également 16,3 % ayant

été GU entre 3 et 4 fois. Les Classes 5 et 6 se caractérisent par une forte récurrence des GU. La Classe 5 (N = 2 615, soit 9 %) contient 76,8 % de GU une seule fois et 24,2 % de GU deux fois, mais ne comprend aucun individu ayant été GU plus de deux fois. La Classe 6 (N = 1 197, soit 4 %), quant à elle, est spécifiquement composée d'individus ayant été GU à plusieurs reprises, avec 65,1 % ayant été GU entre 5 et 6 fois, 21,8 % entre 7 et 9 fois, et 1,7 % ayant été GU 10 fois.

Tableau 3 : Caractérisation des classes parmi les GU

	Classe 1 N = 7 569 ¹	Classe 2 N = 6 228 ¹	Classe 3 N = 2 940 ¹	Classe 4 N = 8 729 ¹	Classe 5 N = 2 615 ¹	Classe 6 N = 1 197 ¹	Total N = 29 278 ¹
GU 1 fois	6 131 (81 %)	2 747 (44,1 %)	2 099 (71,4 %)	5 066 (58 %)	1 983 (76,8 %)	0 (0 %)	18 026 (62 %)
GU 2 fois	790 (10,4 %)	2 084 (33,5 %)	771 (26,2 %)	2 192 (25,1 %)	632 (24,2 %)	0 (0 %)	6 469 (22 %)
GU 3 à 4 fois	621 (8,2 %)	1 397 (22,4 %)	70 (2,4 %)	1 412 (16,3 %)	0 (0 %)	137 (11,4 %)	3 637 (12 %)
GU 5 à 6 fois	27 (0,4 %)	0 (0 %)	0 (0 %)	59 (0,7 %)	0 (0 %)	779 (65,1 %)	865 (3,0 %)
GU 7 à 9 fois	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	261 (21,8 %)	261 (0,9 %)
GU 10 fois	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	20 (1,7 %)	20 (< 0,1 %)

¹ n (%)

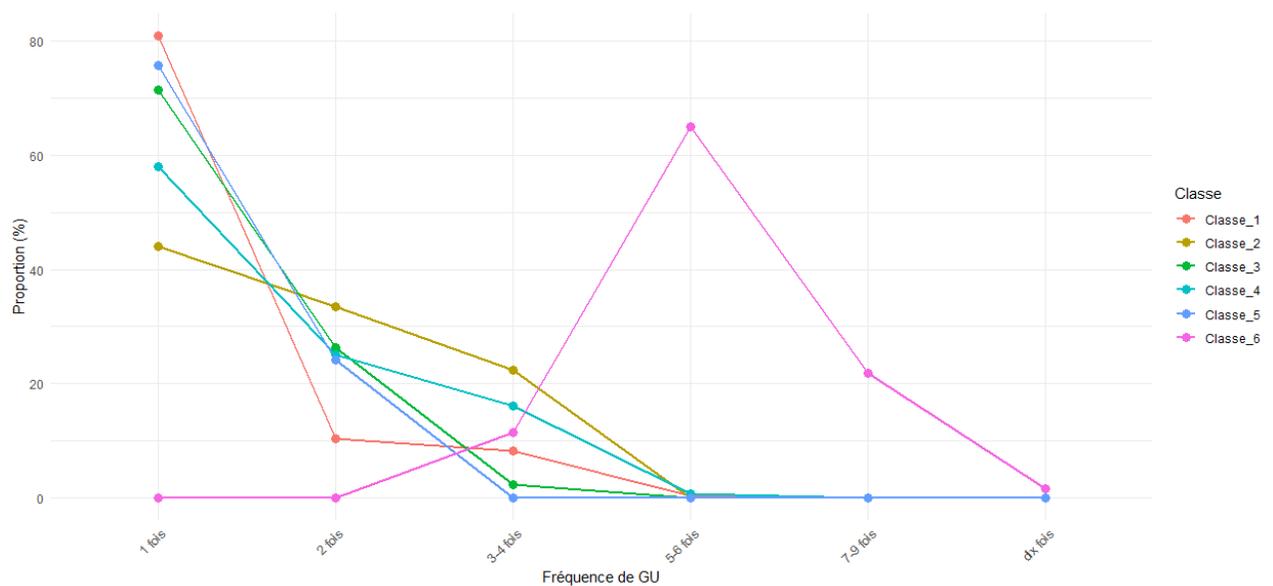
GU : grands utilisateurs

Source : auteurs, analyses des données TorSaDe

La figure 8 permet de visualiser les trajectoires des classes de GU.

On observe que la Classe 6 se distingue des autres par une forte concentration d'individus ayant été GU au moins cinq fois, avec un pic autour de 5-6 et 7-9 fois GU. Elle comprend également l'ensemble des individus ayant été GU 10 fois, soit sur la totalité de la période étudiée.

Figure 8 : Représentation graphiques des classes de GU



Source : auteurs, analyses des données TorSaDe

3.2.3 Résultat du modèle logistique multinomial

L'analyse des facteurs liés à l'appartenance aux différentes classes de GU a été réalisée à l'aide d'un modèle logistique multinomial, en prenant la Classe 1 comme référence. Les résultats montrent que l'appartenance aux classes de GU est associée à des facteurs démographiques, cliniques, socio-économiques et comportementaux (tableau 4).

- **Facteurs démographiques**

Le sexe et l'âge sont significativement associés à l'appartenance aux différentes classes de GU. Comparés aux femmes, les hommes ont une probabilité plus élevée d'appartenir aux Classes 2, 4, 5 et 6, avec un effet particulièrement marqué dans la Classe 6.

L'âge est également un facteur clé : une augmentation de l'âge est associée à une probabilité élevée d'appartenance aux Classes 2, 4, 5 et 6.

- **Facteurs cliniques**

L'état de santé perçu est fortement associé à l'appartenance aux classes de GU. Les individus déclarant une santé générale très bonne ont une probabilité significativement plus faible d'appartenir aux Classes 2, 5 et 6. L'indice de comorbidité est aussi un déterminant majeur. Il est positivement associé à l'appartenance à la Classe 6. À l'inverse, on observe un indice de comorbidité plus faible dans les Classes 4 et 5.

Tableau 4 : Facteurs liés à l'appartenance aux différentes classes de GU

Variables explicatives	Régression logistique multinomiale									
	Modèle classe 2		Modèle classe 3		Modèle classe 4		Modèle classe 5		Modèle classe 6	
	MC	MR	MC	MR	MC	MR	MC	MR	MC	MR
	RC ¹	RC ¹	RC ¹	RC ¹	RC ¹	RC ¹	RC ¹	RC ¹	RC ¹	RC ¹
Sexe										
Femme	—	—	—	—	—	—	—	—	—	—
Homme	1,19***	1,19***	0,97	0,96	1,18***	1,19***	1,17**	1,17**	1,51***	1,51***
Âge										
	1,02***	1,02***	1,00	1,00	1,03***	1,03***	1,03***	1,03***	1,03***	1,03***
Perception de la santé générale										
Mauvaise /	—	—	—	—	—	—	—	—	—	—
Passable	—	—	—	—	—	—	—	—	—	—
Bonne	0,98	0,98	1,14	1,14	0,98	0,99	1,17*	1,15	0,44***	0,43***
Très bonne /	—	—	—	—	—	—	—	—	—	—
Excellente	0,86**	0,87*	1,16	1,14	0,98	0,99	1,17*	1,12	0,27***	0,25***
Perception de la santé mentale										
Mauvaise /	—	—	—	—	—	—	—	—	—	—
Passable	—	—	—	—	—	—	—	—	—	—
Bonne	0,92	—	1,07	—	1,04	—	1,13	—	0,83	—
Très bonne /	—	—	—	—	—	—	—	—	—	—
Excellente	0,98	—	1,03	—	1,08	—	1,03	—	0,91	—
Type de fumeurs										
Non-fumeur	—	—	—	—	—	—	—	—	—	—
Fumeur occasionnel	1,20	1,21*	1,22	1,26*	1,07	1,07	1,30*	1,32*	1,15	1,19
Tous les jours	1,16**	1,16**	1,10	1,09	1,31***	1,31***	1,26**	1,25**	1,36**	1,40**
Consommation d'alcool										
Aucune consommation	—	—	—	—	—	—	—	—	—	—
Occasionnelle (moins d'une fois par mois / une fois par mois / 2-3 fois par mois / une fois par semaine)	0,96	0,96	1,05	1,05	0,98	0,98	1,13	1,10	0,81**	0,80**
Toutes les semaines (2-3 fois par semaine / 4-6 fois par semaine / tous les jours)	0,94	0,94	1,02	1,02	1,00	0,99	1,27**	1,24**	0,69**	0,67***
Indice de masse corporelle, kg/m²										
Poids normal (18 ≤ IMC < 25)	—	—	—	—	—	—	—	—	—	—
Poids insuffisant (IMC < 18)	1,13	1,15	0,90	0,92	1,07	1,07	1,19	1,19	1,29	1,29
Surpoids (25 ≤ IMC < 30)	1,04	1,04	0,91	0,91	1,03	1,03	0,93	0,93	1,05	1,05
Obésité (30 ≥ IMC)	1,09	1,10*	1,00	0,99	1,18	1,18***	1,22*	1,22*	1,59***	1,59***
Niveau d'éducation										

École secondaire non terminée	—	—	—	—	—	—	—	—	—	—
Diplôme d'études secondaires terminées	1,04		1,10		1,00		0,98		0,92	
CÉGEP/ certificat avant bac	1,00		1,03		0,96		0,87*		0,95	
Bac/ certificat après bac / études supérieures (post-bac)	0,90		1,03		0,89*		0,87		0,81	
Statut matrimonial										
Célibataire, jamais marié(e)	—	—	—	—	—	—	—	—	—	—
Séparé(e) / veuf(-ve) / divorcé(e)	0,87*	0,87*	1,13	1,13	0,82**	0,82**	0,81**	0,81**	1,13	1,14
Union libre	0,73***	0,73***	1,12	1,12	0,66***	0,66***	0,67***	0,67***	0,68**	0,69**
Marié(e)	0,83**	0,83**	1,18*	1,18*	0,73***	0,74***	0,73***	0,73***	0,72**	0,72**
Revenu annuel par répondant										
Moins de 20 000 \$	—	—	—	—	—	—	—	—	—	—
20 000 à 40 000 \$	0,90**	0,90*	1,10*	1,12*	0,92*	0,92*	1,0	0,99	0,65***	0,64***
40 000 à 60 000 \$	0,95	0,95	1,22**	1,23**	0,97	0,95	1,0	0,97	0,61***	0,58***
60 000 à 80 000 \$	1,09	1,07	1,15	1,15	0,93	0,89	1,13	1,09	0,55**	0,51**
80 000 \$ et +	0,87	0,87	0,88	0,91	0,97	0,95	1,07	1,06	0,42**	0,39**
Pays de naissance										
Extérieur du Canada	—	—	—	—	—	—	—	—	—	—
Canada	1,04		0,96		1,0		0,88		1,23	
Lieu de résidence										
Rural	—	—	—	—	—	—	—	—	—	—
Urbain	1,00		1,01		1,07*		1,09		0,96	
Perception du stress – Vie quotidienne										
Pas du tout stressé(e)	—	—	—	—	—	—	—	—	—	—
Stressé(e)	1,00		1,04		0,99		1,03		1,01	
Très stressé(e)	0,96		1,05		0,98		1,07		1,21	
A un médecin de famille										
Non	—	—	—	—	—	—	—	—	—	—
Oui	0,96	0,97	1,04	0,96	0,95	0,94	0,74***	0,74***	1,62**	1,62**
Indice de comorbidité	1,04	1,04**	0,99	0,99	0,95***	0,95***	0,84***	0,84***	1,22***	1,22***
Nombre de médecins différents consultés sur 12 mois	1,08***	1,08***	0,99	0,99	1,07***	1,07***	0,96**	0,96**	1,24***	1,24***

¹ RC : rapport de cotes

MR : modèle réduit

MC : modèle complet

Référence : Classe 1

AIC MR : 61425,61; AIC MC : 61478,60

Pseudo r2 MR : 0,353; Pseudo r2 MC : 0,354

Urbain : Montréal, Laval, Montérégie, Capitale-Nationale, Outaouais

Source : auteurs, analyses des données TorSaDe

Le nombre de médecins différents consultés durant les 12 derniers mois joue également un rôle clé. Comparativement à la Classe 1, une augmentation du nombre de consultations médicales est associée à l'appartenance aux Classes 2, 4 et 6. Enfin, l'accès à un médecin de famille influence également la répartition des classes. Les individus disposant d'un médecin de famille ont moins de chances d'appartenir à la Classe 5 ($RC = 0,74$), mais une probabilité significativement plus élevée d'appartenir à la Classe 6 ($RC = 1,62$).

- **Facteurs socio-économiques**

Le revenu annuel est un facteur clé lié à l'appartenance aux classes de GU. Comparés aux individus gagnant moins de 20 000 \$ par an, ceux ayant un revenu plus élevé présentent une faible probabilité d'appartenir aux Classes 4, 5 et 6, avec une diminution plus marquée dans la Classe 6 ($RC = 0,39$ pour les revenus supérieurs à 80 000 \$).

Le statut matrimonial joue également un rôle. Comparativement aux célibataires, les individus mariés ou en union libre ont une faible probabilité d'appartenir aux Classes 2, 4, 5 et 6.

- **Facteurs comportementaux**

Le tabagisme est un facteur qui influence l'appartenance aux classes de GU. Comparativement aux non-fumeurs, les fumeurs quotidiens ont une probabilité plus élevée d'être associés aux Classes 2, 4, 5 et 6. Les fumeurs occasionnels présentent également une probabilité élevée d'être associés aux Classes 3, 4 et 5.

En ce qui concerne la consommation d'alcool, une tendance inverse est observée. Les individus consommant de l'alcool plusieurs fois par semaine ont une probabilité plus faible d'appartenir aux Classes 5 et 6 comparativement à la Classe 1.

4. Discussion

Caractérisation des grands utilisateurs de soins

Nos résultats confirment que les GU se distinguent significativement des non-GU sur le plan sociodémographique, notamment en ce qui concerne l'âge. L'âge médian observé

dans le groupe des GU est de 68 ans, comparativement à 50 ans dans le groupe des non-GU. Cette différence est statistiquement significative et corrobore les résultats de plusieurs études antérieures ayant mis en évidence le rôle de l'avancée en âge comme facteur prédictif d'une forte utilisation des services de santé, en particulier en raison de la forte prévalence de maladies chroniques, de la fragilité fonctionnelle et du risque accru de complications médicales chez les personnes âgées (Huynh et al., 2016; Palmer et al., 2014; Sowa et al., 2021; Wong et al., 2024). Cependant, la littérature demeure hétérogène quant à l'effet de l'âge. Certaines études ont rapporté une surreprésentation des jeunes adultes parmi les GU, en particulier dans les contextes d'utilisation fréquente des services d'urgence ou de soins ambulatoires non planifiés (Hudon et al., 2017; Moe et al., 2021). Cette hétérogénéité peut être attribuée à plusieurs facteurs, notamment à la diversité des contextes d'étude, à la variabilité des systèmes de santé étudiés et aux différences méthodologiques dans la définition des GU, qui repose parfois sur la fréquence d'utilisation (Armoon et al., 2022; Moe et al., 2021) et parfois sur les coûts (Nghiem et al., 2023; Wong et al., 2024; M. Zhang et al., 2023).

Dans le cadre de cette étude, les GU ont été définis selon une approche économique, soit les 5 % de patients ayant généré les coûts les plus élevés sur une année. Ce choix méthodologique est cohérent avec des travaux antérieurs ayant montré que cette minorité est responsable d'environ 50 % des dépenses hospitalières (Nghiem et al., 2023; Wong et al., 2024; M. Zhang et al., 2023) et reflète souvent des profils de patients plus âgés, atteints de comorbidités multiples et nécessitant des hospitalisations prolongées (Karunayawong et al., 2024; Sowa et al., 2021). À l'inverse, les études qui identifient davantage de jeunes GU se fondent souvent sur des indicateurs de fréquence, tels que le nombre de visites aux services d'urgence ou de consultations ambulatoires (Hudon et al., 2017; Moe et al., 2021), ce qui met en lumière une autre dynamique d'utilisation, souvent liée à des facteurs psychosociaux (comme les troubles mentaux ou la précarité) et à une rupture dans la continuité des soins.

Ainsi, l'effet de l'âge dans l'identification des GU doit être interprété à la lumière du cadre conceptuel mobilisé (modèle comportemental d'Andersen), des indicateurs utilisés pour définir l'intensité d'utilisation, ainsi que du contexte sociosanitaire. Ces résultats sont cohérents avec ceux rapportés dans la littérature utilisant une approche similaire, ce qui

confirme la validité méthodologique de l'analyse. Ils soulignent par ailleurs l'importance de distinguer les profils selon les dimensions de l'utilisation (coût, fréquence, complexité) et appellent à des interventions différenciées selon les sous-groupes de GU identifiés.

En ce qui concerne le sexe, nos résultats révèlent une proportion légèrement plus élevée de femmes parmi les GU. Cette observation est cohérente avec certains travaux antérieurs ayant identifié une prédominance féminine dans les groupes à forte consommation de soins (Hudon et al., 2017; Palmer et al., 2014). Toutefois, d'autres études ont rapporté des résultats inverses, montrant une probabilité plus élevée d'être GU chez les hommes (Hwang et al., 2015; Paul et al., 2010), indiquant l'absence de consensus dans la littérature. Plusieurs hypothèses peuvent être avancées pour expliquer cette divergence. D'une part, les différences méthodologiques dans la définition des GU peuvent orienter les résultats dans un sens ou dans l'autre. Par exemple, certaines études centrées sur les visites aux services d'urgence ou sur les hospitalisations répétées peuvent capturer davantage de profils masculins, en particulier dans des contextes liés à des comportements à risque, à une mauvaise observance des traitements ou à des troubles liés à la consommation de substances (Chiu et al., 2022; Kurteva et al., 2023; Moe et al., 2021). D'autre part, les approches fondées sur les coûts totaux, comme celle utilisée dans la présente étude, pourraient favoriser l'identification de profils de femmes âgées, atteintes de comorbidités chroniques nécessitant un suivi régulier et une prise en charge médicale soutenue (Sowa et al., 2021; Wong et al., 2024). Par ailleurs, il convient de prendre en compte les différences de comportements de santé entre les sexes. Les femmes ont tendance à consulter plus fréquemment les professionnels de la santé, à participer davantage aux dépistages et à suivre plus rigoureusement les traitements médicaux (Hudon et al., 2016; Moe et al., 2021). Cette propension à recourir de manière plus soutenue aux soins pourrait expliquer leur surreprésentation parmi les GU, notamment dans un contexte de vieillissement où les femmes constituent une part importante de la population âgée.

Enfin, les spécificités du système de santé québécois, caractérisé par un fort engagement en faveur de l'accès universel et des soins préventifs, peuvent également moduler les différences observées selon le sexe. Il est possible que ces caractéristiques atténuent certaines barrières structurelles à l'accès aux soins pour les femmes, en particulier les plus âgées, contribuant ainsi à une plus grande intensité d'utilisation des services. Par ailleurs,

certain auteurs, comme Robert Whitley, soutiennent que la féminisation du personnel de soins et l'orientation du système vers les besoins des femmes et des enfants pourraient avoir marginalisé les hommes, notamment dans le domaine de la santé mentale (Affleck et al., 2018). Ce biais systémique pourrait en partie expliquer une moindre utilisation des services par les hommes, ou une prise en charge moins adaptée à leurs besoins spécifiques. Ainsi, la relation entre le sexe et l'appartenance au groupe des GU demeure complexe, multidimensionnelle et dépendante du cadre méthodologique, des indicateurs retenus et du contexte de l'étude.

Sur le plan clinique, les résultats obtenus indiquent que les GU présentent une perception significativement plus négative de leur état de santé général et mental, un indice de comorbidité plus élevé ainsi qu'une prévalence élevée du tabagisme et de l'obésité. Ces observations sont cohérentes avec les résultats de nombreuses études qui ont établi une association étroite entre la multimorbidité, en particulier la présence simultanée de pathologies chroniques cardiovasculaires, métaboliques et musculosquelettiques et l'intensification du recours aux services de santé (Moe et al., 2021; Sowa et al., 2021; Wong et al., 2024). Les troubles psychiatriques sont également surreprésentés parmi les GU. Des travaux antérieurs ont mis en évidence une prévalence élevée de diagnostics tels que la schizophrénie, les troubles bipolaires, la dépression majeure et les troubles liés à l'usage de substances dans cette population (Anderson et al., 2018; Armoon et al., 2022; Chiu et al., 2022). Cette cooccurrence entre troubles mentaux et pathologies physiques contribue à une complexification des besoins de soins, souvent mal pris en charge de manière intégrée, et accentue le risque de ruptures dans les trajectoires de soins. Ces résultats soulignent l'importance d'une approche coordonnée et interdisciplinaire pour cette sous-population, notamment en santé mentale.

Par ailleurs, les résultats mettent en lumière l'association entre un statut socio-économique faible et la grande utilisation des services, notamment pour la Classe 6 pour laquelle on suppose l'existence d'une endogénéité. Les GU de notre échantillon affichent un niveau d'éducation inférieur et des revenus annuels plus faibles comparativement aux non-GU, ce qui s'inscrit dans les constats de plusieurs études ayant démontré que la précarité économique, l'exclusion sociale et l'instabilité résidentielle sont des facteurs facilitant une utilisation excessive et souvent inappropriée des soins (Bodenmann et al., 2017; Hudon et

al., 2016; Sun et al., 2003). Il est également probable qu'il s'agisse d'un phénomène endogène à long terme, où les désavantages socio-économiques chroniques contribuent à une dégradation progressive de l'état de santé, entraînant une spirale d'utilisation accrue des services de soins.

Cette conjonction de vulnérabilités cliniques et sociales s'inscrit pleinement dans le cadre du modèle comportemental d'Andersen, selon lequel les facteurs de besoin (état de santé perçu et objectif) et les facteurs prédisposants (niveau d'éducation, conditions socio-économiques) interagissent pour expliquer les comportements de recours. Les résultats obtenus confirment que les GU constituent un sous-groupe cumulant des désavantages multiples cliniques, psychologiques et socio-économiques. Cela renforce l'hypothèse selon laquelle une approche populationnelle, proactive et intégrée est nécessaire pour prévenir la chronicisation de leur recours intensif aux services.

L'analyse par forêt aléatoire a permis d'identifier les variables les plus influentes dans la classification des GU. Parmi celles-ci, l'âge se démarque comme le déterminant principal, suivi du nombre de médecins différents consultés, de l'indice de santé mesuré par l'utilité (HUI3) et de l'indice de masse corporelle (IMC). Ces résultats sont cohérents avec ceux de Nghiem et al. (2023), qui ont montré que les modèles d'apprentissage automatique, en intégrant simultanément des variables cliniques, sociodémographiques et comportementales, peuvent capturer la complexité des profils de GU. Cependant, bien que ces observations soient suggestives, il convient de rester prudent quant à l'interprétation de la performance du modèle utilisé. En l'absence de comparaison directe avec d'autres approches statistiques classiques, telles que la régression logistique, il n'est pas possible de conclure que le modèle par forêt aléatoire offre une capacité prédictive supérieure. L'objectif principal de cette analyse était ici d'explorer, à partir d'un algorithme robuste, les facteurs déterminants d'interactions et de non-linéarités les plus discriminants, sans chercher à optimiser la prédiction comparative. Il convient également de souligner que l'absence de données de panel restreint la possibilité d'identifier les véritables déterminants de la grande utilisation dans une perspective dynamique. Sans suivre les individus dans le temps, il est difficile de distinguer les facteurs structurels ayant un effet causal durable de ceux qui ne sont que des corrélats circonstanciels ou conjoncturels. Cette limite méthodologique restreint la portée explicative des résultats obtenus, en particulier pour les

trajectoires évolutives de grande utilisation. Dans une perspective future, il serait pertinent d'évaluer la performance comparative de différents modèles prédictifs sur des données longitudinales afin d'identifier les méthodes les plus efficaces pour repérer de manière précoce les patients à risque de devenir GU.

Les résultats suggèrent également que l'accès aux soins primaires joue un rôle ambivalent dans le profil des GU. Bien que ceux-ci soient proportionnellement plus nombreux à déclarer avoir un médecin de famille, ils rapportent également un plus grand nombre de professionnels de santé consultés au cours de l'année. À première vue, cette dispersion pourrait être interprétée comme un indicateur de fragmentation des soins. Cependant, une telle conclusion, bien que plausible, mérite d'être relativisée. Consulter plusieurs professionnels de santé peut aussi traduire une plus grande complexité des besoins cliniques. Les GU, souvent atteints de pathologies chroniques multiples et parfois de troubles de santé mentale, peuvent nécessiter l'intervention coordonnée de plusieurs spécialistes pour assurer un suivi adéquat (Hudon et al., 2016; Moe et al., 2021; Sowa et al., 2021). Dans ce contexte, le recours à plusieurs professionnels de santé ne reflète pas nécessairement une défaillance du système de soins, mais plutôt une tentative de réponse à des besoins médicaux multiples que le médecin de famille ne peut assumer seul.

La littérature illustre d'ailleurs cette dualité. Certaines études ont montré que l'inscription auprès d'un médecin de famille est fréquente parmi les GU, en raison de la complexité de leur condition (Bieler et al., 2012; Krieg et al., 2016). Toutefois, cette inscription ne protège pas toujours contre la surutilisation, notamment lorsqu'elle s'inscrit dans un contexte de besoins déjà très lourds ou mal coordonnés (Hudon et al., 2016; Sun et al., 2003). À l'inverse, d'autres travaux suggèrent que, dans certaines conditions, l'inscription auprès d'un médecin de famille peut améliorer la continuité des soins et réduire les consultations inutiles, notamment si elle intervient dès le début du parcours (Hudon et al., 2017).

Ainsi, la multiplicité des consultations observée chez les GU peut relever autant d'un besoin objectif de prise en charge multidisciplinaire que d'une organisation des soins insuffisamment intégrée. Ces résultats soulignent l'importance de renforcer la coordination interprofessionnelle autour des patients à besoins complexes, afin de garantir la continuité des soins tout en évitant les redondances ou les ruptures de suivi.

Classification des trajectoires d'utilisation des services par les grands utilisateurs

L'analyse longitudinale des trajectoires de soins des GU a mis en évidence une hétérogénéité marquée dans leur utilisation des services de santé, permettant de dégager trois profils distincts : les GU occasionnels, récurrents et persistants. Les GU occasionnels (Classes 1, 3 et 5), majoritaires dans notre échantillon (62 %), se caractérisent par une consommation ponctuelle élevée, généralement liée à un épisode de santé aigu nécessitant une prise en charge temporaire. Les GU récurrents (Classes 2 et 4) présentent une alternance entre des périodes de forte consommation et des phases de moindre recours, reflétant souvent la gestion intermittente de maladies chroniques. Enfin, les GU persistants (Classe 6), bien que minoritaires, mobilisent de manière soutenue et disproportionnée les ressources du système, souvent en raison de comorbidités complexes, de troubles psychiatriques sévères et de vulnérabilités sociales multiples.

Cette typologie rejoint les travaux de Wodchis et al. (2016), Moe et al. (2021) et Chiu et al. (2022), qui montrent que les GU ne forment pas un groupe homogène, mais plutôt une mosaïque de trajectoires, certaines stables, d'autres chaotiques ou transitoires. Wodchis et al. (2016), ont identifié, à partir d'un échantillon ontarien, six types de trajectoires, incluant des profils stables en fin de vie, des pics transitoires post-hospitalisation et des parcours chaotiques. De même, Moe et al. (2021) ont utilisé une approche par classes latentes pour différencier des profils d'utilisation des services d'urgence, certains étant transitoires et d'autres persistants, souvent associés à des troubles psychiatriques et à la précarité. Chiu et al. (2022) ont également démontré que des profils récurrents de recours hospitalier peuvent être prédits par la complexité clinique, mais aussi par des facteurs sociaux comme le manque de soutien ou la fragmentation des soins.

En combinant des données longitudinales médico-administratives à une analyse de classes, notre étude apporte une contribution originale dans le contexte québécois, rarement exploré sous cet angle. Elle permet non seulement de distinguer des niveaux d'intensité d'utilisation, mais surtout d'identifier des trajectoires évolutives, incluant les transitions d'un statut de GU occasionnel vers une surutilisation chronique. Cette capacité à repérer les trajectoires à risque constitue un atout stratégique pour la planification de soins ciblés. La reconnaissance d'un sous-groupe de GU persistants est particulièrement importante pour la planification des soins, car ces patients concentrent les besoins les plus complexes.

Leur profil, fréquemment associé à des troubles mentaux, à la précarité sociale et à une instabilité résidentielle, souligne les limites des réponses strictement biomédicales. Armoon et al. (2022) ont montré que l'absence de suivi médical régulier chez les personnes atteintes de troubles psychiatriques sévères conduit à une utilisation répétée des services d'urgence. Pour remédier à cette situation, plusieurs stratégies ont été mises en œuvre, notamment les soins primaires intégrés assurés par une équipe interdisciplinaire, qui se sont révélés efficaces pour améliorer la gestion des maladies chroniques et réduire les hospitalisations (Moe et al., 2021).

D'autres interventions ont ciblé la coordination des soins, par la mise en place de gestionnaires de cas et de travailleurs sociaux. Ces dispositifs permettent une prise en charge plus proactive et personnalisée, réduisant les admissions (Chiu et al., 2022). Toutefois, ces stratégies atteignent leurs limites lorsqu'elles ne tiennent pas compte des déterminants sociaux de la santé. Comme l'ont souligné Anderson et al. (2018), même avec un accès amélioré aux soins, les personnes en situation de précarité ou d'itinérance peuvent rester des GU. Dans ces cas, des interventions structurelles comme l'approche *Logement d'abord*, qui vise à stabiliser la condition de vie avant d'intervenir sur les dimensions médicales et psychologiques, se montrent plus efficaces (Armoon et al., 2022).

Les GU récurrents ou présentant des troubles liés à l'usage de substances peuvent aussi bénéficier de programmes spécialisés. Moe et al. (2021) rapportent que l'accès facilité aux traitements de substitution, comme la méthadone, ainsi que les équipes mobiles en santé mentale (Chiu et al., 2022), contribuent à stabiliser les trajectoires d'utilisation et à éviter les recours systématiques aux soins d'urgence.

Parallèlement, les avancées technologiques ouvrent de nouvelles perspectives pour anticiper les trajectoires à risque. L'intelligence artificielle et les algorithmes d'apprentissage automatique permettent aujourd'hui de prédire, à partir de données cliniques et comportementales, les patients susceptibles de devenir des GU (Nghiem et al., 2023). L'identification précoce de ces profils permettrait d'agir en amont, en proposant des interventions ciblées avant que la surutilisation ne s'installe. De plus, des outils comme la télésurveillance ont démontré leur efficacité dans le suivi des patients atteints de maladies chroniques, réduisant les hospitalisations (Wong et al., 2024). Enfin, l'interconnexion des systèmes d'information entre les différents prestataires de soins favorise une meilleure

coordination et une réduction des examens redondants, ce qui optimise la trajectoire des patients atteints de comorbidités (M. Zhang et al., 2023).

Dans l'ensemble, les interventions les plus efficaces sont celles qui combinent plusieurs dimensions : amélioration de l'accès aux soins primaires, prise en charge interdisciplinaire, technologies prédictives et accompagnement social. Cependant, leur succès dépend largement de leur capacité à répondre aux besoins différenciés des GU selon leur trajectoire. Une approche uniforme est vouée à l'échec si elle ne tient pas compte des dynamiques individuelles et des conditions structurelles qui influencent la grande utilisation des soins. L'avenir de la gestion des GU repose donc sur une stratégie intégrée et adaptative, articulée autour d'une lecture fine et évolutive des parcours de soins.

Facteurs influençant l'appartenance aux différentes classes

L'analyse des facteurs associés à l'appartenance aux différentes classes de GU met en lumière à la fois des caractéristiques démographiques, cliniques, socio-économiques et comportementales. Ces résultats confirment que la grande utilisation des services de santé n'est pas le produit d'un facteur isolé, mais résulte de la combinaison de plusieurs facteurs distincts, dont les effets varient selon les trajectoires.

L'âge apparaît comme un facteur discriminant majeur. Les individus plus âgés sont significativement plus susceptibles d'appartenir aux classes récurrentes et persistantes, caractérisées par une utilisation soutenue et répétée des services de santé. Ce constat rejoint les résultats de Huynh et al. (2016), Hwang et al. (2015) et Palmer et al. (2014), qui soulignent l'effet cumulatif du vieillissement sur la prévalence des maladies chroniques et la perte d'autonomie. Toutefois, nos données révèlent également la présence de jeunes adultes dans certaines classes de GU occasionnels, ce qui montre que l'effet de l'âge n'est ni uniforme ni linéaire. L'âge agit plutôt comme un facteur de vulnérabilité lorsqu'il est combiné à d'autres déterminants, tels que la multimorbidité, un faible soutien social ou des épisodes de soins aigus.

Les indicateurs de santé perçue et de comorbidité ressortent également comme des facteurs déterminants. Les individus rapportant un mauvais état de santé général ou mental, ainsi qu'un indice de comorbidité élevé, présentent une probabilité nettement plus forte d'appartenir aux classes récurrentes ou persistantes. Ces résultats confirment les

observations de Sowa et al. (2021) et Moe et al. (2021), qui mettent en évidence le poids des maladies chroniques et des troubles psychiatriques dans les trajectoires de surutilisation. Nos analyses nuancent cependant cette relation en montrant que la seule présence de troubles mentaux ne suffit pas à expliquer la forte utilisation : c'est souvent leur combinaison avec d'autres vulnérabilités sociales comme l'instabilité résidentielle, la précarité ou l'isolement qui amplifie le recours aux soins (Anderson et al., 2018; Armoon et al., 2022).

La dimension socio-économique joue également un rôle central. Les individus à faible revenu sont nettement surreprésentés dans les classes de GU persistants, un constat cohérent avec les travaux de Bieler et al. (2012), Hudon et al. (2016) et Bodenmann et al. (2017), qui ont démontré l'effet cumulatif de la pauvreté sur l'accès restreint aux soins préventifs, la fragmentation du suivi, et les hospitalisations évitables. Par ailleurs, le fait de vivre seul, qu'il s'agisse d'un statut de célibataire, divorcé(e) ou veuf(-ve), est aussi associé à une probabilité élevée d'appartenir aux classes avec une utilisation persistante. L'absence de soutien social informel peut entraver la gestion autonome des problèmes de santé et renforcer le recours au système de santé comme principal filet de sécurité.

Les comportements de santé, notamment le tabagisme, contribuent également à structurer les trajectoires. Le tabagisme quotidien est fortement associé aux classes de GU récurrents et persistants, ce qui corrobore les conclusions de Chiu et al. (2022) et Wong et al. (2024), selon lesquelles les habitudes de vie à risque sont liées à une augmentation des complications médicales et à une intensification des besoins en soins. À l'inverse, la consommation régulière d'alcool semble, dans notre échantillon, associée à une probabilité plus faible d'appartenir à une classe de GU persistants. Ce résultat inattendu pourrait refléter un biais de déclaration, des différences de recours entre groupes ou encore des mécanismes de non-recours volontaire, et mériterait des investigations complémentaires.

Enfin, l'accès aux soins primaires, mesuré par la présence d'un médecin de famille, révèle un effet ambivalent. Nos résultats indiquent que les individus inscrits auprès d'un médecin de famille sont davantage représentés parmi les GU persistants. Ce paradoxe, également observé par Bieler et al. (2012) et Krieg et al. (2016), peut s'expliquer par un phénomène de sélection inverse : les patients ayant des besoins complexes sont plus souvent dirigés vers des médecins de famille pour un suivi rapproché. Cependant, comme le suggèrent

Hudon et al. (2017), cette inscription ne garantit pas en soi une réduction de l'utilisation des services si elle ne s'accompagne pas d'une coordination réelle, d'un accès facilité et d'une continuité de soins interprofessionnelle.

Nos résultats suggèrent que l'antécédent de recours intensif constitue un prédicteur déterminant de la persistance dans le groupe des GU. Cette persistance interroge l'efficacité du système de santé à gérer et à répondre adéquatement aux besoins des patients complexes. De futurs travaux pourraient examiner ces besoins de manière plus approfondie et identifier des pistes de solution. Cet approfondissement pourrait se faire par l'analyse de bases de données administratives et cliniques couplées permettant de mieux comprendre les déterminants structurels de la grande utilisation répétée. En outre, nos résultats soulignent le caractère multidimensionnel, dynamique et souvent cumulatif de la grande utilisation des soins. Ils renforcent l'idée que les interventions doivent être adaptées aux profils identifiés, en tenant compte non seulement des besoins médicaux, mais aussi des vulnérabilités sociales et comportementales. La détection précoce des trajectoires à risque, notamment en utilisant des outils prédictifs issus de l'apprentissage machine (Nghiem et al., 2023), combinée à des stratégies de soins primaires intégrés, pourrait contribuer à prévenir l'enracinement dans des parcours de surutilisation des services de santé.

Limites

Bien que cette étude apporte de nouvelles perspectives sur la trajectoire des soins des GU et l'identification des facteurs associés à leurs trajectoires, certaines limites doivent être reconnues. Tout d'abord, la définition des GU repose sur un critère économique, à savoir la sélection des 5 % de patients ayant les coûts de soins de santé les plus élevés. Bien que cette approche soit couramment utilisée dans la littérature (Nghiem et al., 2023; Wong et al., 2024; M. Zhang et al., 2023), elle ne prend pas en compte d'autres dimensions de l'utilisation des services de santé, telles que la fréquence des consultations ou les types de services utilisés. Par conséquent, certains patients ayant des besoins de santé complexes mais des coûts modérés peuvent ne pas être identifiés comme GU. De plus, les coûts liés aux services pharmaceutiques ne sont pas inclus dans l'analyse. Une autre limite concerne la nature des données utilisées. L'étude repose sur des données administratives ainsi que

sur des informations autodéclarées provenant de l'ESCC, ce qui peut introduire des biais de mesure, notamment en ce qui concerne la perception de l'état de santé et les comportements de recours aux soins. Par ailleurs, les données administratives ne permettent pas toujours de saisir les raisons sous-jacentes de l'utilisation des services, telles que les facteurs psychosociaux ou les préférences individuelles des patients. En outre, l'analyse repose sur une approche rétrospective qui, bien qu'elle permette de suivre les trajectoires des patients sur plusieurs années, ne tient pas compte des évolutions du système de santé pouvant avoir influencé les comportements de recours. Les changements dans l'offre de services, les politiques publiques ou les innovations en matière de soins médicaux ne sont pas directement intégrés à l'analyse, ce qui peut limiter la généralisation des résultats à d'autres contextes ou périodes.

Enfin, bien que cette étude identifie plusieurs facteurs associés à différentes classes de GU, il n'est pas possible d'établir une relation de causalité stricte entre ces variables. La présence d'une association statistique n'implique pas nécessairement qu'un facteur donné soit responsable d'un recours élevé aux soins, d'autres variables non mesurées pouvant également jouer un rôle dans ces dynamiques.

Malgré ces limites, cette étude constitue une contribution importante à la compréhension des GU de soins de santé, en mettant en lumière leur hétérogénéité ainsi que les multiples déterminants influençant leurs trajectoires. Les résultats ouvrent la voie à de futures recherches visant à affiner les modèles prédictifs et à explorer des interventions ciblées permettant d'optimiser la prise en charge de ces patients.

5. Conclusion

Cette étude met en évidence l'hétérogénéité des profils des GU de soins ainsi que les facteurs déterminants de leur appartenance aux différents groupes. Elle souligne le rôle central des facteurs cliniques, démographiques, socio-économiques et comportementaux dans la caractérisation des GU. Ainsi, une détection précoce des patients à risque, associée à une prise en charge adaptée et interdisciplinaire, pourrait contribuer à optimiser l'utilisation des ressources de santé, à réduire les coûts globaux du système de santé et à améliorer la qualité de vie des patients.

Références

- Affleck, W., Carmichael, V., et Whitley, R. (2018). Men's Mental Health : Social Determinants and Implications for Services. *The Canadian Journal of Psychiatry*, 63(9), 581-589. <https://doi.org/10.1177/0706743718762388>
- Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons.
- Andersen, R. M. (1995). Revisiting the Behavioral Model and Access to Medical Care : Does it Matter? *Journal of Health and Social Behavior*, 36(1), 1-10. <https://doi.org/10.2307/2137284>
- Anderson, M., Revie, C. W., Quail, J. M., Wodchis, W., de Oliveira, C., Osman, M., Baetz, M., McClure, J., Stryhn, H., Buckeridge, D., et Neudorf, C. (2018). The effect of socio-demographic factors on mental health and addiction high-cost use : A retrospective, population-based study in Saskatchewan. *Canadian Journal of Public Health*, 109(5), 810-820. <https://doi.org/10.17269/s41997-018-0101-2>
- Andrén, K. G., et Rosenqvist, U. (1987). Heavy users of an emergency department : A two year follow-up study. *Social Science & Medicine*, 25(7), 825-831. [https://doi.org/10.1016/0277-9536\(87\)90040-2](https://doi.org/10.1016/0277-9536(87)90040-2)
- Andreyeva, E., Chi, W., Zhang, Y., Kaushal, R., et Haynes, K. (2023). *Social Determinants of Health and High-Cost Utilization Among Commercially Insured Population*. 29. <https://www.ajmc.com/view/social-determinants-of-health-and-high-cost-utilization-among-commercially-insured-population>
- Armoon, B., Cao, Z., Grenier, G., Meng, X., et Fleury, M.-J. (2022). Profiles of high emergency department users with mental disorders. *The American Journal of Emergency Medicine*, 54, 131-141. <https://doi.org/10.1016/j.ajem.2022.01.052>

- Baker, L. C., et Baker, L. S. (1994). Excess Cost of Emergency Department Visits for Nonurgent Care. *Health Affairs*, 13(5), 162-171.
<https://doi.org/10.1377/hlthaff.13.5.162>
- Bellemare-Lepage, A., Chatelois, M., et Caron, P.-O. (2023). Exemplification méthodologique d'une analyse de classes latentes avec R. *The Quantitative Methods for Psychology*, 19(2), 217-229. <https://doi.org/10.20982/tqmp.19.2.p217>
- Berk, M. L., et Monheit, A. C. (2001). The Concentration Of Health Care Expenditures, Revisited. *Health Affairs*, 20(2), 9-18. <https://doi.org/10.1377/hlthaff.20.2.9>
- Biau, G. (2012). Analysis of a Random Forests Model. *Journal of Machine Learning Research*, 13(38), 1063-1095. <http://jmlr.org/papers/v13/biau12a.html>
- Bieler, G., Paroz, S., Faouzi, M., Trueb, L., Vaucher, P., Althaus, F., Daepfen, J.-B., et Bodenmann, P. (2012). Social and Medical Vulnerability Factors of Emergency Department Frequent Users in a Universal Health Insurance System. *Academic Emergency Medicine*, 19(1), 63-68. <https://doi.org/10.1111/j.1553-2712.2011.01246.x>
- Bodenmann, P., Velonaki, V.-S., Griffin, J. L., Baggio, S., Iglesias, K., Moschetti, K., Ruggeri, O., Burnand, B., Wasserfallen, J.-B., Vu, F., Schupbach, J., Hugli, O., et Daepfen, J.-B. (2017). Case Management may Reduce Emergency Department Frequent use in a Universal Health Coverage System : A Randomized Controlled Trial. *Journal of General Internal Medicine*, 32(5), 508-515.
<https://doi.org/10.1007/s11606-016-3789-9>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
<https://doi.org/10.1023/A:1010933404324>

- Charlson, M. E., Pompei, P., Ales, K. L., et MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies : Development and validation. *Journal of Chronic Diseases*, 40(5), 373-383.
[https://doi.org/10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8)
- Chiu, Y. M., Dufour, I., Courteau, J., Vanasse, A., Chouinard, M.-C., Dubois, M.-F., Dubuc, N., Elazhary, N., et Hudon, C. (2022). Profiles of frequent emergency department users with chronic conditions : A latent class analysis. *BMJ Open*, 12(9), e055297. <https://doi.org/10.1136/bmjopen-2021-055297>
- Collins, L. M., et Lanza, S. T. (2009). *Latent Class and Latent Transition Analysis : With Applications in the Social, Behavioral, and Health Sciences*. John Wiley & Sons.
- Corbin, J. M., et Strauss, A. L. (1988). *Unending Work and Care : Managing Chronic Illness at Home*. Jossey-Bass Publishers.
- Cutler, A., Cutler, D. R., et Stevens, J. R. (2012). Random Forests. In C. Zhang et Y. Ma (Éds.), *Ensemble Machine Learning : Methods and Applications* (p. 157-175). Springer. https://doi.org/10.1007/978-1-4419-9326-7_5
- Doran, K. M., Raven, M. C., et Rosenheck, R. A. (2013). What Drives Frequent Emergency Department Use in an Integrated Health System? National Data From the Veterans Health Administration. *Annals of Emergency Medicine*, 62(2), 151-159. <https://doi.org/10.1016/j.annemergmed.2013.02.016>
- Doupe, M. B., Palatnick, W., Day, S., Chateau, D., Soodeen, R.-A., Burchill, C., et Derksen, S. (2012). Frequent Users of Emergency Departments : Developing Standard Definitions and Defining Prominent Risk Factors. *Annals of Emergency Medicine*, 60(1), 24-32. <https://doi.org/10.1016/j.annemergmed.2011.11.036>

- Efron, B. (1979). Bootstrap Methods : Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26. <https://www.jstor.org/stable/2958830>
- Elixhauser, A., Steiner, C., Harris, D. R., et Coffey, R. M. (1998). Comorbidity Measures for Use with Administrative Data. *Medical Care*, 36(1), 8. https://journals.lww.com/lww-medicalcare/abstract/1998/01000/comorbidity_measures_for_use_with_administrative.4.aspx
- Fagerland, M. W., et Hosmer, D. W. (2012). A Generalized Hosmer–Lemeshow Goodness-of-Fit Test for Multinomial Logistic Regression Models. *The Stata Journal: Promoting Communications on Statistics and Stata*, 12(3), 447-453. <https://doi.org/10.1177/1536867X1201200307>
- Fitzpatrick, T., Rosella, L. C., Calzavara, A., Petch, J., Pinto, A. D., Manson, H., Goel, V., et Wodchis, W. P. (2015). Looking Beyond Income and Education : Socioeconomic Status Gradients Among Future High-Cost Users of Health Care. *American Journal of Preventive Medicine*, 49(2), 161-171. <https://doi.org/10.1016/j.amepre.2015.02.018>
- Fuda, K. K., et Immekus, R. (2006). Frequent Users of Massachusetts Emergency Departments : A Statewide Analysis. *Annals of Emergency Medicine*, 48(1), 16.e1-16.e8. <https://doi.org/10.1016/j.annemergmed.2006.03.001>
- Genuer, R., Poggi, J.-M., et Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225-2236. <https://doi.org/10.1016/j.patrec.2010.03.014>

- Goodman, L. A. (1974). Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika*, 61(2), 215-231.
<https://doi.org/10.2307/2334349>
- Gruneir, A., Dhalla, I. A., van Walraven, C., Fischer, H. D., Camacho, X., Rochon, P. A., et Anderson, G. M. (2011). Unplanned readmissions after hospital discharge among patients identified as being at high risk for readmission using a validated predictive algorithm. *Open Medicine*, 5(2), e104-e111.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3148002/>
- Hagenaars, J. A., et McCutcheon, A. L. (Éds.). (2002). *Applied Latent Class Analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511499531>
- Hansagi, H., Olsson, M., Sjöberg, S., Tomson, Y., et Göransson, S. (2001). Frequent use of the hospital emergency department is indicative of high use of other health care services. *Annals of Emergency Medicine*, 37(6), 561-567.
<https://doi.org/10.1067/mem.2001.111762>
- Hastie, T., Tibshirani, R., et Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832-844. IEEE Transactions on Pattern Analysis and Machine Intelligence.
<https://doi.org/10.1109/34.709601>
- Hovstadius, B., et Petersson, G. (2012). Factors Leading to Excessive Polypharmacy. *Clinics in Geriatric Medicine*, 28(2), 159-172.
<https://doi.org/10.1016/j.cger.2012.01.001>

- Hudon, C., Courteau, J., Krieg, C., et Vanasse, A. (2017). Factors associated with chronic frequent emergency department utilization in a population with diabetes living in metropolitan areas : A population-based retrospective cohort study. *BMC Health Services Research*, 17(1), 525. <https://doi.org/10.1186/s12913-017-2453-3>
- Hudon, C., Sanche, S., et Haggerty, J. L. (2016). Personal Characteristics and Experience of Primary Care Predicting Frequent Use of Emergency Department : A Prospective Cohort Study. *PLOS ONE*, 11(6), e0157489. <https://doi.org/10.1371/journal.pone.0157489>
- Huynh, C., Ferland, F., Blanchette-Martin, N., Ménard, J.-M., et Fleury, M.-J. (2016). Factors Influencing the Frequency of Emergency Department Utilization by Individuals with Substance Use Disorders. *Psychiatric Quarterly*, 87(4), 713-728. <https://doi.org/10.1007/s11126-016-9422-6>
- Hwang, A. S., Liu, S. W., Ashburner, J. M., Auerbach, B. J., Atlas, S. J., et Hong, C. S. (2015). Outcomes of primary care patients who are frequent and persistent users of the ED. *The American Journal of Emergency Medicine*, 33(9), 1320-1322. <https://doi.org/10.1016/j.ajem.2015.05.042>
- ICIS. (2023). *Tendances des dépenses nationales de santé, 2023—Analyse éclair* | ICIS. <https://www.cihi.ca/fr/tendances-des-depenses-nationales-de-sante-2023-analyse-eclair>
- ICIS. (2024). *Tendances des dépenses nationales de santé, 2024—Analyse éclair* | ICIS. <https://www.cihi.ca/fr/tendances-des-depenses-nationales-de-sante-2024-analyse-eclair>

- James, G., Witten, D., Hastie, T., Tibshirani, R., et Taylor, J. (2023). Introduction. In G. James, D. Witten, T. Hastie, R. Tibshirani, et J. Taylor (Éds.), *An Introduction to Statistical Learning : With Applications in Python* (p. 1-13). Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0_1
- Karunayawong, P., Gaewkhiew, P., Sarajan, M. H., Boonma, C., Butchon, R., Sukmanee, J., Saeraneesopon, T., Teerawattananon, Y., et Isaranuwachai, W. (2024). High-cost users still came to hospitals during the COVID-19 pandemic during first wave data in Thailand : Secondary data analysis. *BMC Public Health*, 24(1), 2917. <https://doi.org/10.1186/s12889-024-20325-y>
- Kouakou, C. R. C., Bélan, M., Poder, T. G., et Laberge, M. (2025). Care trajectories of people with mood disorders in Quebec using latent class and latent profile analysis methods. *Journal of Mood & Anxiety Disorders*, 9. <https://doi.org/10.1016/j.xjmad.2024.100101>
- Krieg, C., Hudon, C., Chouinard, M.-C., et Dufour, I. (2016). Individual predictors of frequent emergency department use : A scoping review. *BMC Health Services Research*, 16(1), 594. <https://doi.org/10.1186/s12913-016-1852-1>
- Kurteva, S., Tamblyn, R., et Meguerditchian, A. N. (2023). Predictors of frequent emergency department visits among hospitalized cancer patients : A comparative cohort study using integrated clinical and administrative data to improve care delivery. *BMC Health Services Research*, 23(1), 887. <https://doi.org/10.1186/s12913-023-09854-1>
- Kutner, M. H., Nachtsheim, C. J., Neter, J., et Li, W. (2005). *Applied linear statistical models*. McGraw-Hill. <https://thuvienso.hoasen.edu.vn/handle/123456789/9564>

- LaCalle, E. J., Rabin, E. J., et Genes, N. G. (2013). High-Frequency Users of Emergency Department Care. *Journal of Emergency Medicine*, 44(6), 1167-1173.
<https://doi.org/10.1016/j.jemermed.2012.11.042>
- Lanza, S. T., Collins, L. M., Lemmon, D. R., et Schafer, J. L. (2007). PROC LCA : A SAS Procedure for Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 671-694.
<https://doi.org/10.1080/10705510701575602>
- Lanza, S. T., et Cooper, B. R. (2016). Latent Class Analysis for Developmental Research. *Child Development Perspectives*, 10(1), 59-64.
<https://doi.org/10.1111/cdep.12163>
- Lanza, S. T., et Rhoades, B. L. (2013). Latent Class Analysis : An Alternative Perspective on Subgroup Analysis in Prevention and Treatment. *Prevention Science*, 14(2), 157-168. <https://doi.org/10.1007/s11121-011-0201-1>
- Lasser, K. E., Himmelstein, D. U., et Woolhandler, S. (2006). Access to Care, Health Status, and Health Disparities in the United States and Canada : Results of a Cross-National Population-Based Survey. *American Journal of Public Health*, 96(7), 1300-1307. <https://doi.org/10.2105/AJPH.2004.059402>
- Linzer, D. A., et Lewis, J. B. (2011). poLCA : An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42, 1-29.
<https://doi.org/10.18637/jss.v042.i10>
- L.McCutcheon, A. (1987). *Latent Class Analysis*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781412984713>

- Malone, R. E. (1995). Heavy users of emergency services : Social construction of a policy problem. *Social Science & Medicine*, 40(4), 469-477.
[https://doi.org/10.1016/0277-9536\(94\)E0116-A](https://doi.org/10.1016/0277-9536(94)E0116-A)
- McCullagh, P. (2019). *Generalized Linear Models* (2^e éd.). Routledge.
<https://doi.org/10.1201/9780203753736>
- McLachlan, G. J., et Peel, D. (2004). *Finite Mixture Models*. John Wiley & Sons.
- Menard, S. (2001). *Applied Logistic Regression Analysis*. SAGE Publications.
- Moe, J., Bailey, A. L., Oland, R., Levesque, L., et Murray, H. (2013). Defining, quantifying, and characterizing adult frequent users of a suburban Canadian emergency department. *Canadian Journal of Emergency Medicine*, 15(4), 214-226. <https://doi.org/10.2310/8000.2013.130936>
- Moe, J., O'Sullivan, F., McGregor, M. J., Schull, M. J., Dong, K., Holroyd, B. R., Grafstein, E., Hohl, C. M., Trimble, J., et McGrail, K. M. (2021). Identifying subgroups and risk among frequent emergency department users in British Columbia. *JACEP Open*, 2(1). <https://doi.org/10.1002/emp2.12346>
- More, A. S., et Rana, D. P. (2020). An Experimental Assessment of Random Forest Classification Performance Improvisation with Sampling and Stage Wise Success Rate Calculation. *Procedia Computer Science*, 167, 1711-1721.
<https://doi.org/10.1016/j.procs.2020.03.381>
- Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables : New opportunities for latent class-latent growth modeling. In *New methods for the analysis of change* (p. 291-322). American Psychological Association. <https://doi.org/10.1037/10409-010>

- Nghiem, N., Atkinson, J., Nguyen, B. P., Tran-Duy, A., et Wilson, N. (2023). Predicting high health-cost users among people with cardiovascular disease using machine learning and nationwide linked social administrative datasets. *Health Economics Review*, 13(1), 9. <https://doi.org/10.1186/s13561-023-00422-1>
- Nguena Nguefack, H. L., Pagé, M. G., Katz, J., Choinière, M., Vanasse, A., Dorais, M., Samb, O. M., et Lacasse, A. (2020). Trajectory Modelling Techniques Useful to Epidemiological Research : A Comparative Narrative Review of Approaches. *Clinical Epidemiology*, 12, 1205-1222. <https://doi.org/10.2147/CLEP.S265287>
- Nylund, K. L., Asparouhov, T., et Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling : A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- OECD. (2019, novembre 7). *Health at a Glance 2019*. OECD. https://www.oecd.org/en/publications/health-at-a-glance-2019_4dd50c09-en.html
- O'Toole, T. P., Pollini, R., Gray, P., Jones, T., Bigelow, G., et Ford, D. E. (2007). Factors identifying high-frequency and low-frequency health service utilization among substance-using adults. *Journal of Substance Abuse Treatment*, 33(1), 51-59. <https://doi.org/10.1016/j.jsat.2006.12.002>
- Palmer, E., Leblanc-Duchin, D., Murray, J., et Atkinson, P. (2014). Emergency department use. *Canadian Family Physician*, 60(4), e223-e229. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4046533/>
- Paul, P., Heng, B. H., Seow, E., Molina, J., et Tay, S. Y. (2010). Predictors of frequent attenders of emergency department at an acute general hospital in Singapore.

- Emergency Medicine Journal*, 27(11), 843-848.
<https://doi.org/10.1136/emj.2009.079160>
- Qureshi, D., Isenberg, S., Tanuseputro, P., Moineddin, R., Quinn, K., Meaney, C., McGrail, K., Seow, H., Webber, C., Fowler, R., et Hsu, A. (2020). Describing the characteristics and healthcare use of high-cost acute care users at the end of life : A pan-Canadian population-based study. *BMC Health Services Research*, 20(1), 997. <https://doi.org/10.1186/s12913-020-05837-8>
- Rask, K. J., Williams, M. V., McNagny, S. E., Parker, R. M., et Baker, D. W. (1998). Ambulatory health care use by patients in a public hospital emergency department. *Journal of General Internal Medicine*, 13(9), 614-620.
<https://doi.org/10.1046/j.1525-1497.1998.00184.x>
- Ronksley, P. E., Liu, E. Y., McKay, J. A., Kobewka, D. M., Rothwell, D. M., Mulpuru, S., et Forster, A. J. (2016). Variations in Resource Intensity and Cost Among High Users of the Emergency Department. *Academic Emergency Medicine*, 23(6), 722-730. <https://doi.org/10.1111/acem.12939>
- Rosella, L. C., Fitzpatrick, T., Wodchis, W. P., Calzavara, A., Manson, H., et Goel, V. (2014). High-cost health care users in Ontario, Canada : Demographic, socio-economic, and health status characteristics. *BMC Health Services Research*, 14(1), 532. <https://doi.org/10.1186/s12913-014-0532-2>
- Ryoo, J. H., Wang, C., Swearer, S. M., Hull, M., et Shi, D. (2018). Longitudinal Model Building Using Latent Transition Analysis : An Example Using School Bullying Data. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00675>

- Samadoulougou, S., Letarte, L., et Lebel, A. (2022). Association between Neighbourhood Deprivation Trajectories and Self-Perceived Health : Analysis of a Linked Survey and Health Administrative Data. *International Journal of Environmental Research and Public Health*, 20(1), 486. <https://doi.org/10.3390/ijerph20010486>
- Schreiber, J. B. (2017). Latent Class Analysis : An example for reporting results. *Research in Social and Administrative Pharmacy*, 13(6), 1196-1201. <https://doi.org/10.1016/j.sapharm.2016.11.011>
- Simard, M., Sirois, C., et Candas, B. (2018). Validation of the Combined Comorbidity Index of Charlson and Elixhauser to Predict 30-Day Mortality Across ICD-9 and ICD-10. *Medical Care*, 56(5), 441. <https://doi.org/10.1097/MLR.0000000000000905>
- Sirois, C., Domingues, N. S., Laroche, M.-L., Zongo, A., Lunghi, C., Guénette, L., Kröger, E., et Émond, V. (2019). Polypharmacy Definitions for Multimorbid Older Adults Need Stronger Foundations to Guide Research, Clinical Practice and Public Health. *Pharmacy*, 7(3), Article 3. <https://doi.org/10.3390/pharmacy7030126>
- Sowa, P. M., Venuthurupalli, S. K., Hoy, W. E., Zhang, J., Cameron, A., Healy, H. G., et Connelly, L. B. (2021). Identification of factors associated with high-cost use of inpatient care in chronic kidney disease : A registry study. *BMJ Open*, 11(8), e049755. <https://doi.org/10.1136/bmjopen-2021-049755>
- Sun, B. C., Burstin, H. R., et Brennan, T. A. (2003). Predictors and Outcomes of Frequent Emergency Department Users. *Academic Emergency Medicine*, 10(4), 320-328. <https://doi.org/10.1111/j.1553-2712.2003.tb01344.x>

- Tille, F., Gibis, B., Balke, K., Kuhlmeier, A., et Schnitzer, S. (2017). Sociodemographic and health-related determinants of health care utilisation and access to primary and specialist care : Results of a nationwide population survey in Germany (2006–2016). *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, *126*, 52-65. <https://doi.org/10.1016/j.zefq.2017.07.012>
- Ullman, R., Block, J. A., et Stratmann, W. C. (1975). An Emergency Room's Patients : Their Characteristics and Utilization of Hospital Services. *Medical Care*, *13*(12), 1011. https://journals.lww.com/lww-medicalcare/abstract/1975/12000/An_Emergency_Room_s_Patients_Their.3.aspx
- Vanasse, A., Chiu, Y. M., Courteau, J., Dorais, M., Bartlett, G., Zawaly, K., et Benigeri, M. (2021). Cohort Profile : The Care Trajectories-Enriched Data (TorSaDE) cohort. *International Journal of Epidemiology*, *50*(4), 1066-1066h. <https://doi.org/10.1093/ije/dyaa167>
- Wiens, K., Rosella, L. C., Kurdyak, P., Chen, S., Aubry, T., Stergiopoulos, V., et Hwang, S. W. (2021). Factors associated with higher healthcare costs in a cohort of homeless adults with a mental illness and a general cohort of adults with a history of homelessness. *BMC Health Services Research*, *21*(1), 555. <https://doi.org/10.1186/s12913-021-06562-6>
- Wodchis, W. P., Austin, P. C., et Henry, D. A. (2016). A 3-year study of high-cost users of health care. *CMAJ*, *188*(3), 182-188. <https://doi.org/10.1503/cmaj.150064>
- Wong, J. J., Côté, P., Tricco, A. C., Watson, T., et Rosella, L. C. (2024). Characterizing high-cost healthcare users among adults with back pain in Ontario, Canada : A

population-based cohort study. *PAIN*, 165(9), 1944.

<https://doi.org/10.1097/j.pain.0000000000003200>

Zhang, M., Ma, J., Xie, F., et Thabane, L. (2023). Identifying factors associated with high use of acute care in Canada : A population-based retrospective study. *The European Journal of Health Economics*, 24(9), 1505-1515.

<https://doi.org/10.1007/s10198-022-01558-3>

Annexes

Annexe 1 : Coût moyen par année et par groupe (dollars canadiens de 2019)

Année	Coût moyen ± IC à 95 %		Valeur de p
	GU	Non-GU	
2007	3 213,76 ± 467,96	912,04 ± 52,64	6,65 ^e -21
2008	3 413,02 ± 490,62	938,84 ± 55,58	7,60 ^e -22
2009	3 850,12 ± 574,68	992,93 ± 52,44	2,23 ^e -21
2010	4 608,08 ± 667,67	1 111,35 ± 58,50	1,88 ^e -23
2011	10 526,43 ± 932,19	985,59 ± 37,50	2,62 ^e -75
2012	16 840,07 ± 1103,93	682,89 ± 16,36	5,64 ^e -132
2013	7 768,22 ± 901,17	1 473,58 ± 79,65	5,09 ^e -39
2014	7 584,30 ± 951,26	1 656,90 ± 89,41	7,02 ^e -32
2015	7 253,71 ± 1076,92	1 698,59 ± 84,03	7,35 ^e -23
2016	3 064,04 ± 346,51	1 032,22 ± 40,85	1,62 ^e -28

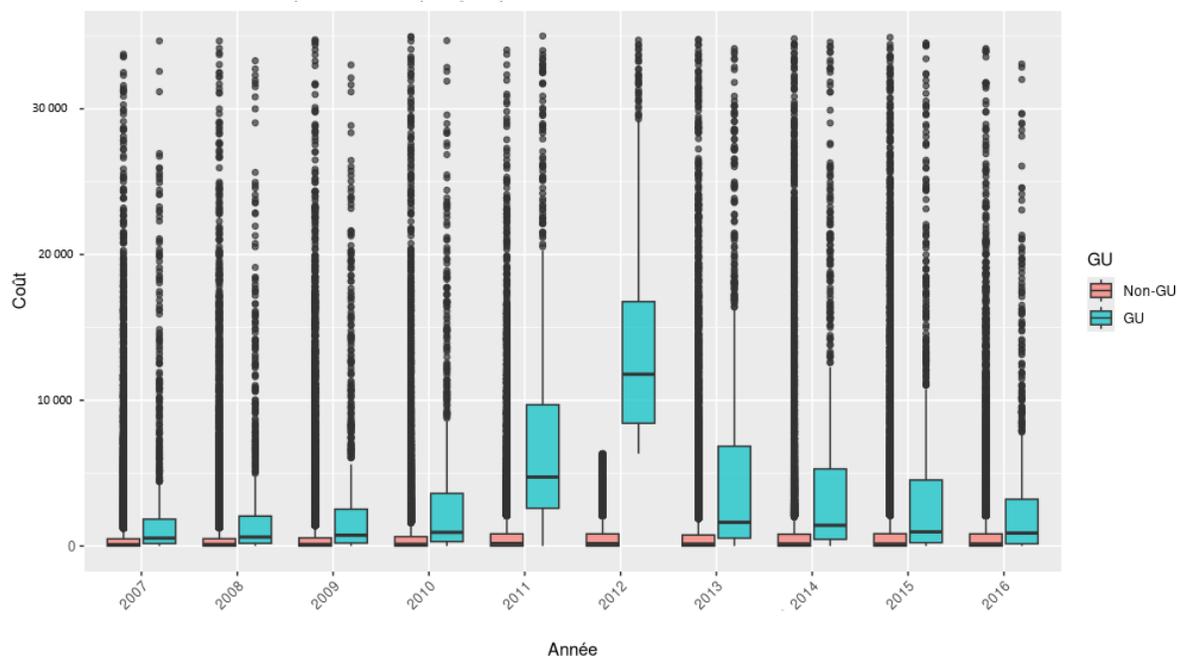
GU : grands utilisateurs; Non-GU : non-grands utilisateurs,

Valeur de p : statistique du test t de Student,

IC à 95 % : intervalle de confiance

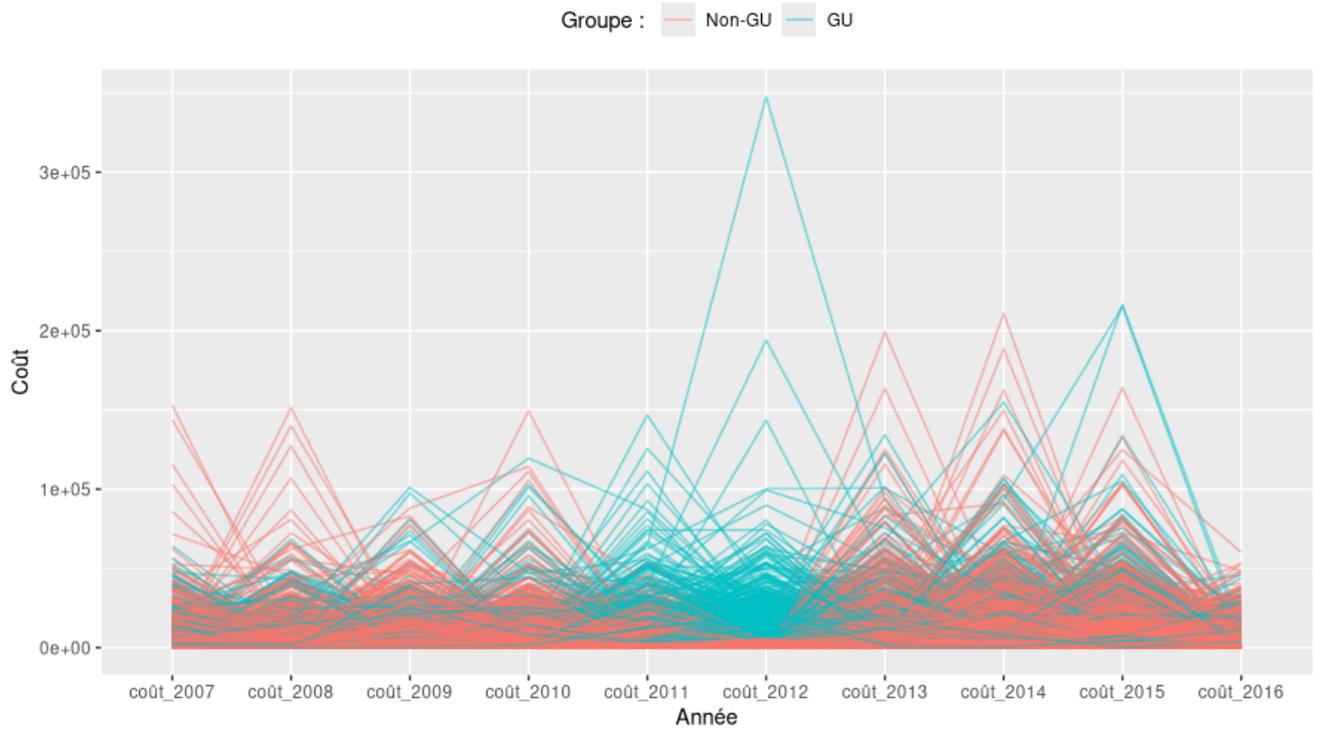
Source : auteurs, analyses des données TorSaDe avec R

Annexe 2 : Boîtes à moustaches des coûts par groupe et par année



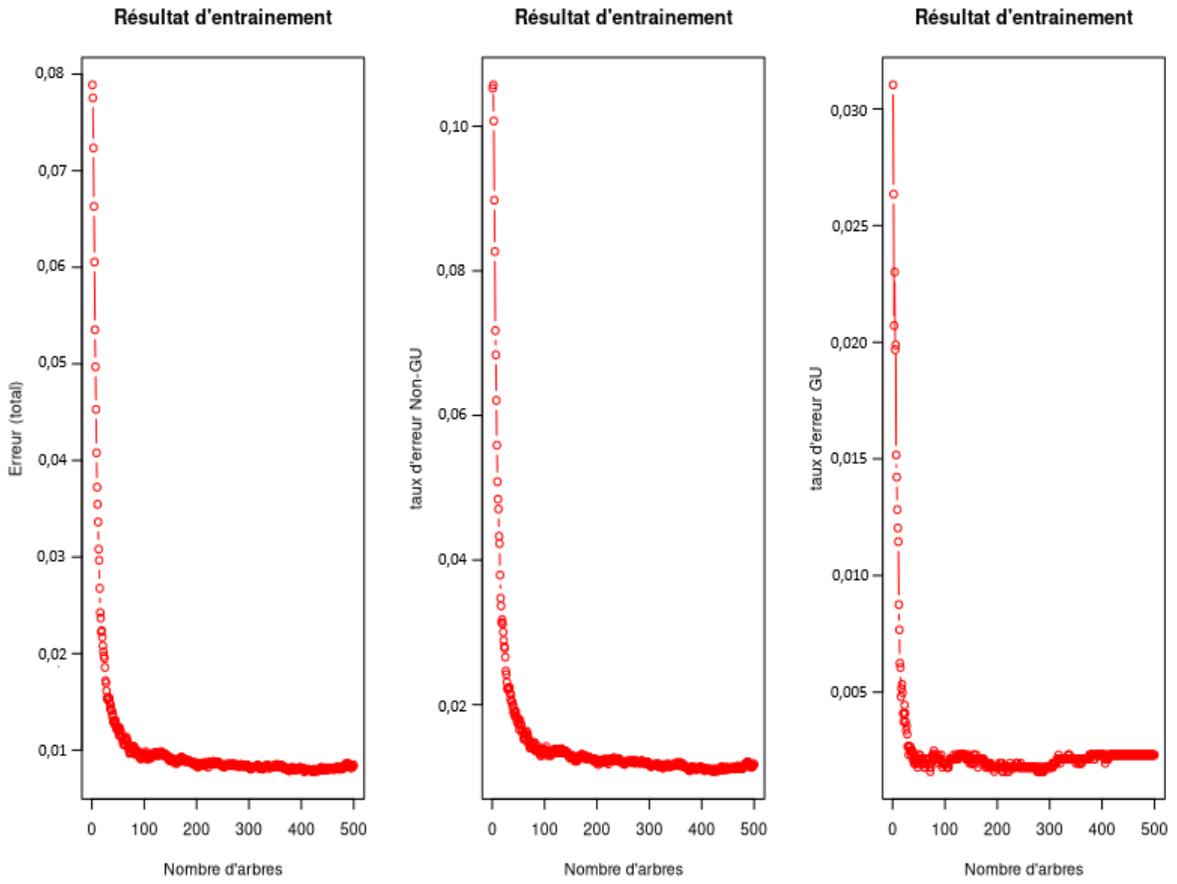
Source : auteurs, analyses des données TorSaDe

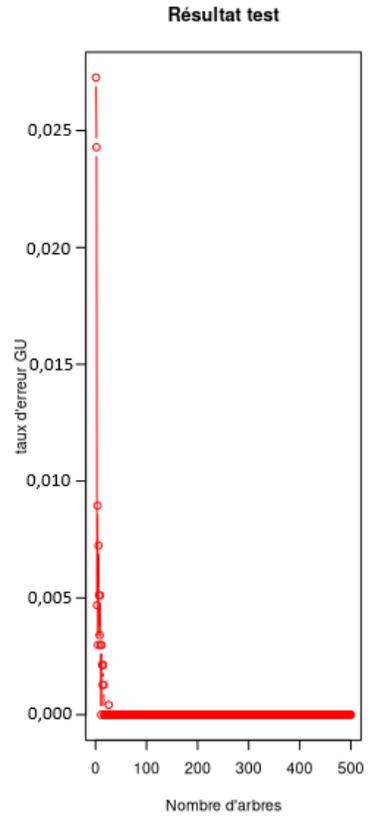
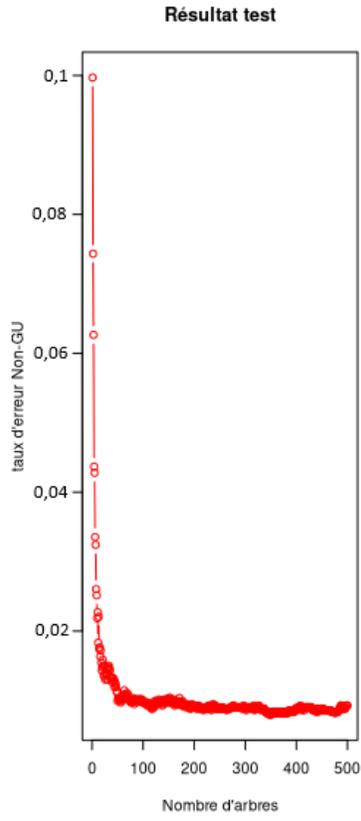
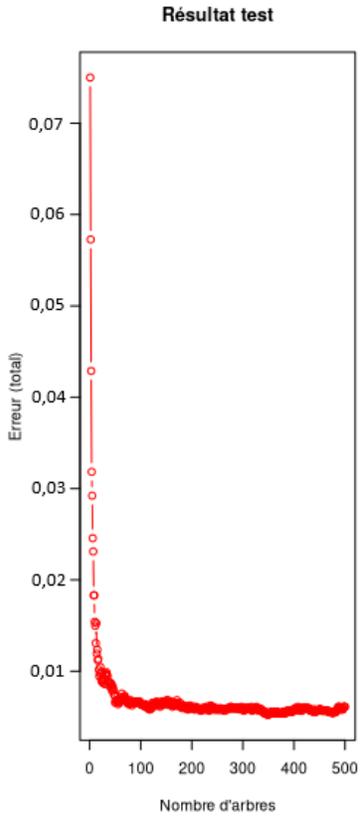
Annexe 3 : Trajectoire individuelle des coûts



Source : auteurs, analyses des données TorSaDe

Annexe 4 : Test post estimation forêt aléatoire





Annexe 5 : Description des données des GU, non-GU et échantillon de la cohorte TorSaDe, seuil de 10 %

Variable	Non-GU N = 17 601 ^l	GU N = 1 955 ^l	Échantillon TorSaDe N = 19 556 ^l	Valeur de p ²
Sexe				< 0,001
Femme	53,25 %	60,77 %	54,00 %	
Homme	46,75 %	39,23 %	46,00 %	
Âge en 2012, années	48,9	58,8	49,9	< 0,001
Perception de la santé générale				< 0,001
Mauvaise / Passable	8,94 %	24,27 %	10,47 %	
Bonne	30,72 %	35,33 %	31,18 %	
Très bonne / Excellente	60,34 %	40,40 %	58,35 %	
Perception de la santé mentale				< 0,001
Mauvaise / Passable	3,51 %	7,06 %	3,86 %	
Bonne	19,80 %	22,37 %	20,05 %	
Très bonne / Excellente	76,69 %	70,57 %	76,08 %	
Type de fumeurs				0,005
Tous les jours	18,10 %	20,05 %	18,29 %	
Fumeur occasionnel	5,07 %	3,68 %	4,93 %	
Non-fumeur	76,84 %	76,27 %	76,78 %	
Consommation d'alcool				< 0,001
Aucune consommation	16,69 %	23,57 %	17,38 %	
Occasionnelle (moins d'une fois par mois / une fois par mois / 2-3 fois par mois / une fois par semaine)	48,71 %	50,18 %	48,86 %	
Toutes les semaines (2-3 fois par semaine / 4-6 fois par semaine / tous les jours)	34,60 %	26,25 %	33,76 %	
Indice de masse corporelle, kg/m²				< 0,001
Poids insuffisant (IMC < 18)	2,61 %	3,01 %	2,65 %	
Poids normal (18 ≤ IMC < 25)	46,96 %	42,20 %	46,47 %	
Surpoids (25 ≤ IMC < 30)	34,16 %	34,67 %	34,21 %	
Obésité (30 ≥ IMC)	16,27 %	20,12 %	16,67 %	
Niveau d'éducation				< 0,001
École secondaire non terminée	26,22 %	37,05 %	27,30 %	
Diplôme d'études secondaires terminées	12,04 %	10,28 %	11,87 %	
CÉGEP / certificat avant bac	41,51 %	37,56 %	41,11 %	
Bac/ certificat après bac / études supérieures (post-baccalauréat)	20,23 %	15,11 %	19,72 %	
Statut matrimonial				< 0,001
Marié(e)	30,85 %	33,49 %	31,12 %	
Union libre	18,84 %	15,82 %	18,54 %	
Séparé(e) / veuf(-ve) / divorcé(e)	17,53 %	30,11 %	18,79 %	
Célibataire / jamais marié(e)	32,78 %	20,58 %	31,56 %	
Revenu annuel par répondant				< 0,001
Moins de 20 000 \$	34,99 %	44,74 %	35,99 %	
20 000 à 40 000 \$	31,34 %	32,06 %	31,42 %	
40 000 à 60 000 \$	19,20 %	15,01 %	18,77 %	

Variable	Non-GU N = 17 601 ¹	GU N = 1 955 ¹	Échantillon TorSaDe N = 19 556 ¹	Valeur de p ²
60 000 à 80 000 \$	8,50 %	5,12 %	8,15 %	
80 000 \$ et +	5,97 %	3,07 %	5,67 %	
Pays de naissance				0,3
Canada	92,45 %	93,15 %	92,52 %	
Extérieur du Canada	7,55 %	6,85 %	7,48 %	
Lieu de résidence				< 0,001
Rural	56,59 %	61,89 %	57,12 %	
Urbain	43,41 %	38,11 %	42,88 %	
Indice de l'état de santé (HUI3)	0,9	0,8	0,9	< 0,001
Satisfaction de la vie en générale				0,004
Très satisfait(e)	41,70 %	36,85 %	41,20 %	
Satisfait(e)	51,37 %	54,38 %	51,68 %	
Insatisfait(e) / très insatisfait(e) / ni satisfait(e) ni insatisfait(e)	6,93 %	8,76 %	7,12 %	
Perception du stress – Vie quotidienne				< 0,001
Pas du tout stressé(e)	15,92 %	22,35 %	16,57 %	
Un peu stressé(e) / assez stressé(e)	60,38 %	57,52 %	60,09 %	
Très stressé(e) / extrêmement stressé(e)	23,70 %	20,12 %	23,34 %	
A un médecin de famille				< 0,001
Oui	77,11 %	85,73 %	77,97 %	
Non	22,89 %	14,27 %	22,03 %	
Indice de comorbidité	0,2	0,6	0,2	< 0,001
Nombre de médicaments distincts sur 12 mois	4,9	9,1	5,6	< 0,001
Nombre de médecins différents consultés sur 12 mois	0,9	1,8	1,0	< 0,001

¹ %

² Test exact de Fisher; test de Wilcoxon; test χ^2 de Pearson

GU : grand utilisateur; **non-GU** : non-grand utilisateur; **IMC** : indice de masse corporelle

Urbain : Montréal, Laval, Montérégie, Capitale-Nationale, Outaouais

Rural : Saguenay-Lac-Saint-Jean, Chaudière-Appalaches, Abitibi-Témiscamingue, Bas-Saint-Laurent, Côte-Nord, Gaspésie/Îles-de-la-Madeleine, Nord-du-Québec, Mauricie/Centre-du-Québec, Estrie, Laurentides, Lanaudière.

Extérieur du Canada : Autres pays de l'Amérique du Nord, Amérique centrale Antilles et Bermudes, Europe, Afrique.

Annexe 6 : Description des données des GU, non-GU et échantillon de la cohorte TorSaDe, seuil de 1 %

Variables	Non-GU N = 19 361 ¹	GU N = 195 ¹	Échantillon TorSaDe N = 19 556 ¹	Valeur de p ²
Sexe				0,10
Femme	54,07 %	47,69 %	54,00 %	
Homme	45,93 %	52,31 %	46,00 %	
Âge en 2012, années	49,7	69,0	49,9	< 0,001
Perception de la santé générale				< 0,001
Mauvaise / Passable	10,19 %	38,66 %	10,47 %	
Bonne	31,17 %	32,47 %	31,18 %	
Très bonne / Excellente	58,64 %	28,87 %	58,35 %	
Perception de la santé mentale				0,020
Mauvaise / Passable	3,83 %	6,91 %	3,86 %	
Bonne	20,01 %	24,47 %	20,05 %	
Très bonne / Excellente	76,16 %	68,62 %	76,08 %	
Type de fumeurs				0,6
Tous les jours	18,28 %	19,49 %	18,29 %	
Fumeur occasionnel	4,94 %	3,59 %	4,93 %	
Non-fumeur	76,78 %	76,92 %	76,78 %	
Consommation d'alcool				0,003
Aucune consommation	17,29 %	25,64 %	17,38 %	
Occasionnelle (moins d'une fois par mois / une fois par mois / 2-3 fois par mois / une fois par semaine)	48,86 %	48,72 %	48,86 %	
Toutes les semaines (2-3 fois par semaine / 4-6 fois par semaine / tous les jours)	33,85 %	25,64 %	33,76 %	
Indice de masse corporelle, kg/m²				0,11
Poids insuffisant (IMC < 18)	2,64 %	3,28 %	2,65 %	
Poids normal (18 ≤ IMC < 25)	46,55 %	38,25 %	46,47 %	
Surpoids (25 ≤ IMC < 30)	34,17 %	37,70 %	34,21 %	
Obésité (30 ≥ IMC)	16,63 %	20,77 %	16,67 %	
Niveau d'éducation				< 0,001
École secondaire non terminée	27,12 %	45,64 %	27,30 %	
Diplôme d'études secondaires terminées	11,87 %	11,28 %	11,87 %	
CÉGEP / certificat avant bac	41,20 %	32,31 %	41,11 %	
Bac / certificat après bac / études supérieures (post- bac)	19,81 %	10,77 %	19,72 %	
Statut matrimonial				< 0,001
Marié	31,07 %	35,38 %	31,12 %	
Union libre	18,64 %	8,21 %	18,54 %	
Séparé / veuf / divorcé	18,57 %	40,51 %	18,79 %	
Célibataire, jamais marié	31,72 %	15,90 %	31,56 %	
Revenu annuel par répondant				0,001

Variables	Non-GU N = 19 361 ¹	GU N = 195 ¹	Échantillon TorSaDe N = 19 556 ¹	Valeur de p ²
Moins de 20 000 \$	35,84 %	50,58 %	35,99 %	
20 000 à 40 000 \$	31,47 %	26,74 %	31,42 %	
40 000 à 60 000 \$	18,81 %	15,12 %	18,77 %	
60 000 à 80 000 \$	8,19 %	4,65 %	8,15 %	
80 000 \$ et +	5,70 %	2,91 %	5,67 %	
Pays de naissance				0,2
Canada	92,49 %	94,87 %	92,52 %	
Extérieur du Canada	7,51 %	5,13 %	7,48 %	
Lieu de résidence				0,067
Rural	57,06 %	63,59 %	57,12 %	
Urbain	42,94 %	36,41 %	42,88 %	
Indice de l'état de santé (HUI3)	0,9	0,8	0,9	< 0,001
Satisfaction de la vie en générale				0,4
Très satisfait(e)	41,25 %	36,08 %	41,20 %	
Satisfait(e)	51,66 %	53,61 %	51,68 %	
Insatisfait(e) / Très insatisfait(e) / ni satisfait(e) ni insatisfait(e)	7,09 %	10,31 %	7,12 %	
Perception du stress – Vie quotidienne				< 0,001
Pas du tout stressé(e)	16,43 %	30,26 %	16,57 %	
Un peu stressé(e) / assez stressé(e)	60,15 %	54,36 %	60,09 %	
Très stressé(e) / extrêmement stressé(e)	23,42 %	15,38 %	23,34 %	
A un médecin de famille				< 0,001
Oui	77,86 %	89,74 %	77,97 %	
Non	22,14 %	10,26 %	22,03 %	
Indice de comorbidité	0,2	0,9	0,2	< 0,001
Nombre de médicaments distincts sur 12 mois	5,4	10,7	5,6	< 0,001
Nombre de médecins différents consultés sur 12 mois	1,0	2,3	1,0	< 0,001

¹ %

² Test exact de Fisher; test de Wilcoxon; test χ^2 de Pearson

GU : grand utilisateur; **non-GU** : non-grand utilisateur; **IMC** : indice de masse corporelle

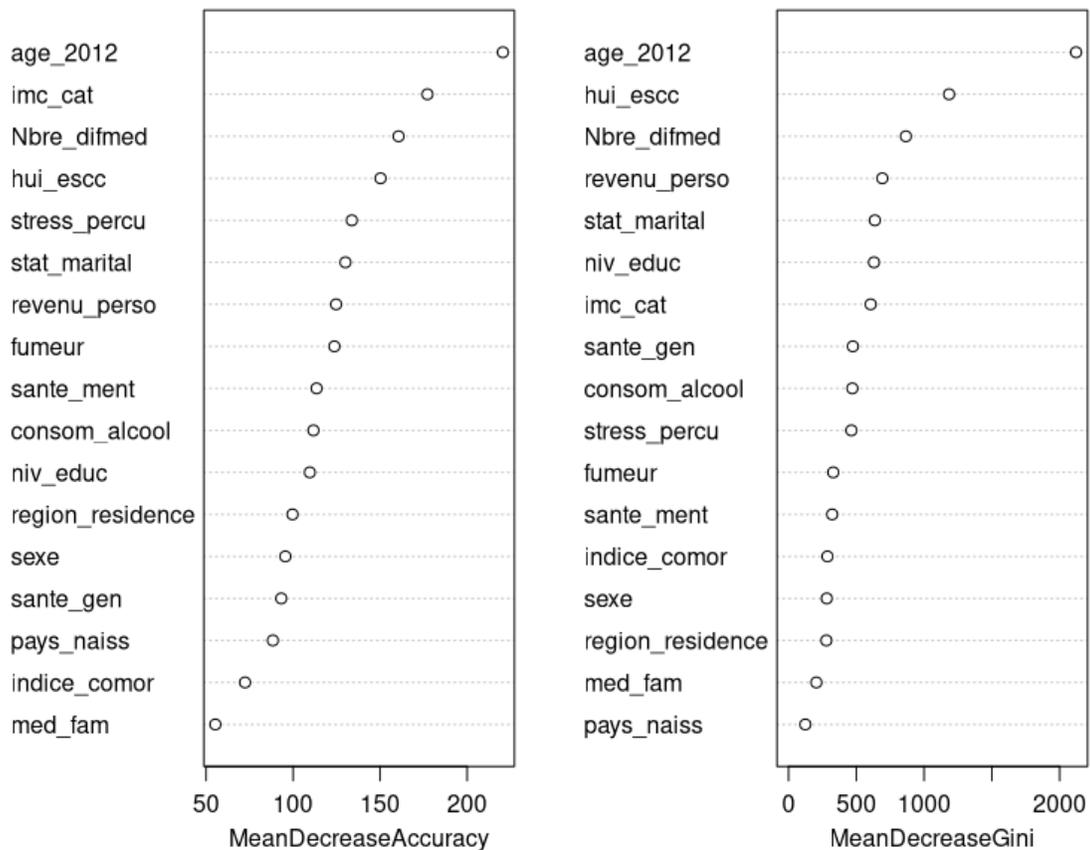
Urbain : Montréal, Laval, Montérégie, Capitale-Nationale, Outaouais

Rural : Saguenay-Lac-Saint-Jean, Chaudière-Appalaches, Abitibi-Témiscamingue, Bas-Saint-Laurent, Côte-Nord, Gaspésie/Îles-de-la-Madeleine, Nord-du-Québec, Mauricie/Centre-du-Québec, Estrie, Laurentides, Lanaudière.

Extérieur du Canada : Autres pays de l'Amérique du Nord, Amérique centrale Antilles et Bermudes, Europe, Afrique.

Annexe 7 : Test de robustesse forêt aléatoire, seuil de 10 %

Forêt aléatoire - GU seuil 10 %



Annexe 8 : Test de robustesse forêt aléatoire, seuil de 1 %

Forêt aléatoire - GU- 1%

