Mondes appariés synthétiques pour évaluer la déshérence médicale dans le diabète de type 2 : un banc d'essai de simulation mécaniste des schémas d'évaluation fondés sur les données de remboursement

Un banc d'essai où la vérité est connue par construction montre comment les études sur données de remboursement faussent l'effet de la déshérence médicale.

6 novembre 2026 (version révisée, R1, v019i), Jérôme Vétillard, VP R&D Qualees*

Résumé

Contexte. Dans les données de remboursement, un patient qui sort du suivi d'une maladie chronique n'est pas observé, et son contrefactuel individuel est inobservable : les évaluations des programmes de vigilance portent donc un biais de taille et de signe inconnus.

Objectif. L'objectif principal est méthodologique : construire un banc d'essai de simulation dans lequel la vérité est connue par construction, et mesurer dans quelle mesure des schémas d'évaluation spécifiés, fondés sur les données de remboursement, retrouvent des contrastes causaux spécifiés de la déshérence médicale, avec le diabète de type 2 comme premier cas d'usage. La détection par dispositifs connectés et une analyse de marge (*headroom*) du programme de vigilance PREDICARE sont des exemples secondaires, traités sous hypothèses explicites.

Méthodes. predicare-synth est une microsimulation causale mécaniste du syndrome cardio-réno-métabolique, non entraînée sur des données de patients, dont les paramètres sont posés à partir de statistiques publiées ; la calibration est en attente. Un tronc quotidien alimente des données de remboursement françaises simulées (SNDS), des dispositifs connectés (IoMT) et une application patient. Un tronc mensuel suit 60 000 personnes atteintes de diabète de type 2 pendant 20 ans dans des mondes appariés partageant des nombres aléatoires communs : suivi continu, statu quo, réengagement seul, une variante IoMT et la séquence PREDICARE (ciblage sur les données de remboursement, bilan en unité mobile, télésurveillance pendant la déshérence). Les estimands sont définis explicitement. Les résultats déterministes sont répliqués sur 10 graines avec leurs erreurs-types de Monte Carlo (MCSE), et les schémas fondés sur les données de remboursement ont été rejoués sous 14 mécanismes générateurs de données alternatifs. Le gain de performance du registre IoMT est imposé sous forme de courbe prescrite : l'entraînement du modèle, le choix des étiquettes, la validation temporelle et le transfert entre territoires ne sont pas simulés. Les coûts sont exprimés en euros 2024 ; l'analyse coût-utilité (perspectives payeur et collective, actualisation à 2,5 %, deux analyses probabilistes de 300 tirages) rapporte une marge, c'est-à-dire le coût fixe de fonctionnement maximal compatible avec un seuil de lecture.

Résultats. Parmi les personnes satisfaisant le critère de silence dans les données de remboursement, le contraste simulé apparié, sous le modèle supposé, montre que la déshérence augmente la mortalité à 10 ans de 5,8 points (MCSE 0,1) tout en réduisant les coûts payeur non actualisés de 1 981 € (MCSE 76). Restreindre la cohorte à la réapparition hospitalière biaise le coût à la hausse de 3 087 € (MCSE 323 ; positif pour chaque graine) ; l'analyse de survie en intention de suivre dilue l'effet sur la mortalité (2,1

contre 5,3 points dans l'émulation de cohorte à graine unique ; 3,3 contre 5,8 points sur 10 graines) et l'analyse selon l'exposition effective (*as-treated*) manque une mortalité différée. Sur 14 mécanismes posés et non calibrés, le biais à la hausse des cohortes fondées sur la réapparition persiste (139 des 140 couples mécanisme–graine), mais le biais de mortalité en intention de suivre s'inverse lorsque la déshérence est un marqueur de gravité latente, et le signe de la marge économique du réengagement dépend du mécanisme. À titre d'illustration secondaire et conditionnelle, la marge de la séquence PREDICARE (le coût fixe de fonctionnement maximal compatible avec 50 k€/QALY sous des paramètres posés et non calibrés) dépend principalement de l'atténuation, non mesurée, de l'excès de risque sous télésurveillance et de la participation ; ce n'est pas une estimation de ce que PREDICARE pourrait coûter. Sous la sélection posée sur la charge latente, la comparaison des participants aux non-participants surestime l'effet sur la mortalité d'un facteur six environ ; l'estimand défendable est l'effet en intention de traiter de la politique d'offre parmi toutes les personnes éligibles, qui, sous le budget d'invitation simulé, se trouve dilué à −0,31 point de mortalité à 10 ans. Dans une expérience de prédiction interne, les variables IoMT simulées portent l'AUC-ROC pour la décompensation cardiaque à 30 jours de 0,57 à 0,70, avec une valeur prédictive positive de 5 % pour un quota d'alertes de 5 %.

Conclusions. La contribution principale est un banc d'essai de simulation dans lequel la vérité est connue par construction, montrant que, dans les mécanismes testés, les schémas d'évaluation fondés sur les données de remboursement faussent l'effet de la déshérence médicale : la sélection sur la réapparition hospitalière en surestime le coût sous tous les mécanismes, tandis que le sens du biais de mortalité dépend du mécanisme causal posé. Il montre aussi pourquoi l'effet de la politique d'offre parmi toutes les personnes éligibles, et non l'effet parmi les participants, est l'estimand qu'une évaluation randomisée peut identifier. Les analyses de registre, de prédiction et de coût-utilité sont des illustrations conditionnelles ; le banc d'essai ne montre ni que PREDICARE fonctionne ni ce qu'il pourrait coûter, et ses résultats non calibrés ne disent rien de l'incidence, de la mortalité ou des coûts en France.

Mots-clés. déshérence médicale ; perdus de vue ; microsimulation ; données synthétiques ; données de remboursement ; étude de simulation ; estimands ; analyse coût-utilité ; télésurveillance ; diabète de type 2 ; schéma d'évaluation

1. Introduction

Les maladies chroniques dominent désormais la dépense de santé en France : environ 13,7 millions de personnes (une sur cinq) relevaient du dispositif des affections de longue durée (ALD) en 2021, représentant environ deux tiers des dépenses de l'Assurance maladie [1]. Ces patients suivent des trajectoires continues, mais ils sont observés de manière intermittente : une personne atteinte d'un diabète de type 2 bien équilibré ne voit un médecin que quelques fois par an. Lorsque le suivi s'interrompt, rien dans le système ne lève d'alerte : le patient cesse simplement de produire des données.

Nous appelons ce phénomène *déshérence médicale* (en anglais, *medical disengagement*). Notre thèse de travail est qu'il ne s'agit pas principalement d'un défaut d'attention des professionnels, mais du résultat attendu d'un régime d'observation discret appliqué à un risque continu. C'est donc une propriété architecturale, qu'un programme de vigilance ne peut modifier qu'en changeant ce qui est observé, à quelle fréquence, et qui agit sur cette observation.

PREDICARE est l'un de ces programmes. Son premier objectif est de constituer un registre longitudinal de personnes en déshérence, par une observation IoMT renforcée, afin d'identifier rétrospectivement les motifs qui précèdent les décompensations ; la valeur prédictive et clinique de ces motifs reste à établir. Les données de remboursement localiseraient les ruptures de suivi et retrouveraient les

événements, un bilan initial en unité mobile caractériserait l'état clinique de départ, et les dispositifs connectés fourniraient les mesures répétées entre ce bilan et un éventuel événement, précisément ce qui manque aux données de soins courants pour des patients qui ne consultent plus. La séquence prévue consiste à constituer le registre, relier les séries à des événements cliniquement confirmés, apprendre rétrospectivement, puis valider sur une période ultérieure et un autre territoire, les alertes étant d'abord calculées en mode silencieux ; ce n'est qu'ensuite que les alertes serviraient à orienter les patients vers les soins aigus. À plus long terme, le programme vise à préserver la continuité de la surveillance des personnes en ALD malgré la baisse de la densité médicale. Le présent article n'apporte de preuve pour aucune de ces étapes. Il traite une question préalable : les schémas fondés sur les données de remboursement qui serviraient à évaluer un tel programme peuvent-ils seulement retrouver un effet connu ? Et son registre est une courbe de performance prescrite, non un processus d'apprentissage simulé.

Trois obstacles empêchent d'évaluer un tel programme sur données réelles. Premièrement, le patient en déshérence est invisible précisément pendant qu'il l'est ; les études l'identifient généralement après coup, lorsqu'il réapparaît, souvent par l'hôpital. Deuxièmement, le contrefactuel individuel n'est pas observé : le même patient, resté suivi, ne peut être observé. Troisièmement, le bénéfice de la prévention est un événement évité, inobservable au niveau individuel (le « paradoxe de la prévention »). Dans ces conditions, une estimation observationnelle ne peut être confrontée à la vérité.

Les données synthétiques sont habituellement proposées pour protéger la vie privée ou pour augmenter des jeux d'entraînement. Nous les utilisons à une autre fin : construire des *mondes dans lesquels la vérité est connue*, afin que les méthodes de détection, les règles de ciblage et les schémas d'évaluation puissent être notés avant d'être imposés aux patients. La logique est l'inverse de l'émulation d'essai cible [2] : au lieu d'émuler un essai dans des données réelles, nous générons des données autour d'un mécanisme causal connu et vérifions si les schémas appliqués à ces données le retrouvent, comme dans les études de simulation qui évaluent des méthodes statistiques [3].

La contribution principale de cet article est un banc d'essai de simulation dans lequel la vérité est connue par construction, montrant que, dans les mécanismes testés, les schémas d'évaluation fondés sur les données de remboursement faussent l'effet de la déshérence médicale : la sélection sur la réapparition hospitalière en surestime le coût sous tous les mécanismes, tandis que le sens du biais de mortalité dépend du mécanisme causal posé. Son objectif principal est donc méthodologique : mesurer dans quelle mesure des schémas spécifiés, fondés sur les données de remboursement, retrouvent des contrastes causaux spécifiés, sous le mécanisme posé et sous des mécanismes alternatifs dans lesquels la déshérence est en partie ou en totalité un marqueur de gravité. L'expérience de prédiction avec dispositifs connectés et les analyses coût-utilité sont secondaires ; elles montrent à quoi le banc d'essai peut servir, et ce qu'il ne peut pas établir.

Cet article apporte cinq contributions :

1. un générateur mécaniste de trajectoires cardio-réno-métaboliques doté de têtes d'observation de type SNDS, IoMT et application, dans lequel chaque paramètre porte un domaine de validité et un statut épistémique explicites ;

2. un schéma de mondes appariés fondé sur des nombres aléatoires communs, qui produit des issues contrefactuelles simulées connues par construction, de sorte que les effets moyens de la déshérence et du réengagement sont mesurés sans biais de comparaison, conditionnellement au modèle posé ;

3. un script de ciblage fondé uniquement sur les données de remboursement, transposable au SNDS réel, dont la précision, le rappel et le biais d'estimation sont mesurés par rapport à la vérité terrain du simulateur (et non par rapport à la déshérence réelle) ;

4. la démonstration que, dans les configurations testées, les schémas observationnels testés sur données de remboursement (cohortes définies par la réapparition, survie en intention de suivre et attribution selon l'exposition effective) manquent ou surestiment chacun l'effet de la déshérence ; que, dans une classe de 14 mécanismes générateurs de données posés, le biais à la hausse des cohortes fondées sur la réapparition persiste alors que le sens du biais de mortalité ne persiste pas ; et un cadrage coût-utilité qui remplace la question « combien cela fera-t-il économiser ? » par « combien cela peut-il coûter ? », rapporté sous forme de marge selon trois perspectives et sous analyse de sensibilité probabiliste ;

5. un modèle de la séquence PREDICARE (monde PE) (ciblage sur les données de remboursement, bilan en unité mobile à T0 et télésurveillance IoMT démarrant pendant la déshérence) montrant que, sous une participation sélective posée, le contraste naïf participants contre non-participants surestime l'effet d'un facteur six environ, et que l'évaluation doit randomiser l'offre.

2. Méthodes

2.1 Schéma d'étude

Le banc d'essai sépare trois choses que les données réelles confondent : le mécanisme (un modèle causal), l'observation de ce mécanisme (données de remboursement, dispositifs, application) et les analyses. Les analyses ne lisent que les observations ; la vérité terrain est stockée à part et n'est lue que par l'étape d'évaluation (Figure 1).

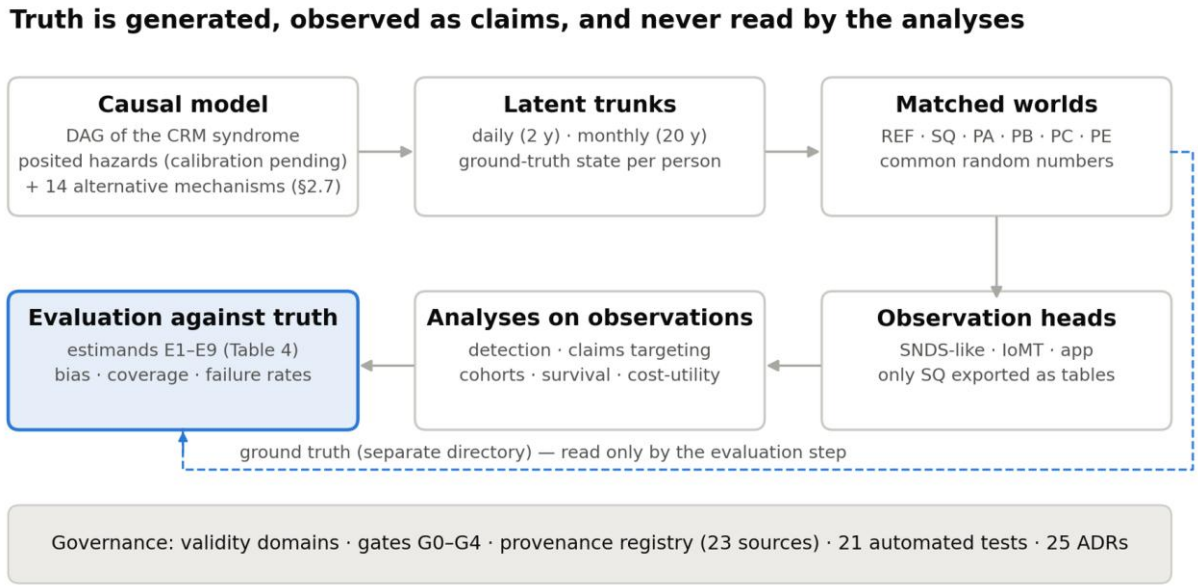

Figure 1 - Chaîne du banc d'essai : six étapes et un chemin de vérité. Les risques instantanés sont posés, non calibrés ; les contrefactuels ne sont connus que sous le modèle supposé.

Chaque expérience compare des bras qui ne diffèrent que par un seul facteur (un ensemble de variables, un schéma, un ordre de file d'attente ou un monde), à vérité latente, population et partition des données identiques. Les résultats sont donc relatifs par construction. L'étude est organisée selon la structure

ADEMP des études de simulation (objectifs, mécanismes générateurs de données, estimands, méthodes, mesures de performance) [3] : les estimands sont définis au §2.6 (Tableau 4), les mécanismes générateurs alternatifs au §2.7, et chaque résultat déterministe de long terme est répliqué sur 10 graines (§2.4).

2.2 Le générateur : une microsimulation causale mécaniste, non un modèle entraîné

Aucune donnée de patient n'a servi à entraîner le générateur. predicare-synth n'est pas un modèle génératif profond (GAN, VAE, modèle de diffusion ou modèle de langage) ajusté sur des dossiers réels. C'est une microsimulation stochastique dont la structure est un modèle causal écrit à la main et dont les paramètres sont posés à partir d'agrégats publiés (la calibration formelle est en attente, voir §2.5 et Tableau supplémentaire S2). Ce choix est délibéré. Un générateur entraîné sur des données réelles reproduit le régime d'observation de ces données (y compris leurs angles morts) et ne peut produire précisément ce qui manque : des trajectoires ambulatoires longitudinales de patients qui ont cessé d'être observés.

Structure causale. Un graphe orienté acyclique (DAG) relie les affections du syndrome cardio-réno-métabolique, selon le cadre CKM de l'American Heart Association [4] : obésité → diabète de type 2, hypertension, dyslipidémie → maladie rénale chronique (MRC) stadifiée par le DFGe (3a, 3b, 4, 5) [5] → insuffisance cardiaque, avec l'infarctus du myocarde et l'AVC comme événements. La BPCO est conservée comme affection témoin hors graphe. Les arêtes portent des rapports de risques instantanés multiplicatifs de l'ordre de ceux rapportés dans la littérature (par exemple, le diabète multiplie le risque instantané de MRC par ~3,5). Le graphe est acyclique par construction ; la boucle cardio-rénale est rompue dans le sens majoritaire métabolique → rénal → cardiaque, et cette simplification est documentée.

Deux troncs latents, une seule source de paramètres (Tableau 1).

Tableau 1 - Les deux troncs latents : pas de temps, horizon, population et finalité.

Tronc	Pas temporel	Horizon	Population	Finalité
Quotidien	1 jour	2 anss	40,000–100,000 adultes, CRM + COPD	Détection à horizon court (≤ 30 jours), Signaux IoMT
Mensuel	1 mois	20 ans (1 + 10 + 10)	60,000 personnes atteintes de diabete type 2	Problèmes de long terme, coûts, QALYs, émulation de cohorte

Les deux troncs lisent les mêmes paramètres de risque instantané dans les mêmes fichiers de configuration, de sorte qu'une modification du modèle causal se propage aux deux.

État latent. Chaque personne simulée porte un état de vérité terrain qu'aucune règle ni aucun modèle ne lit jamais : affections avec date de survenue, stade de MRC, physiologie quotidienne dans le tronc quotidien (poids, pression artérielle, indice glycémique et temps dans la cible dérivé de la MCG, fréquence cardiaque de repos, activité, sommeil, SpO_2, DFGe, kaliémie), une charge microvasculaire à trois domaines dans le tronc mensuel (rétinopathie 0–2, albuminurie A1–A3, pied diabétique grade 0–2), l'état de traitement et les épisodes de déshérence.

Déshérence. La rupture est tirée comme un risque instantané tant que le patient est suivi (~4,7 % par an) et elle est récurrente : retour spontané, reconnexion après un séjour hospitalier (60 %), puis rupture ultérieure. Quatre mois après le début d'une rupture, les risques instantanés de décompensation sont multipliés par 3,5, les risques cardiovasculaires par 1,5, et la progression de la MRC et des atteintes microvasculaires est accélérée (×1,8 à ×2,5). Une forme partielle maintient la délivrance en pharmacie pendant 2 à 6 mois.

Ce que « déshérence » signifie dans le modèle. Une rupture combine trois mécanismes qu'une étude sur données de remboursement ne peut séparer : la perte de contact observé (pas de consultation de médecin généraliste, pas de biologie), l'interruption du traitement (pas de délivrance, sauf pendant une rupture partielle avec pharmacie résiduelle) et un excès de risque clinique qui débute quatre mois après le début de la rupture. *Déshérence médicale* désigne la rupture latente ; *silence observable* désigne sa trace dans les données de remboursement (estimand E8, §2.6). Dans les analyses principales, la totalité de l'excès de risque est causée par la rupture ; le §2.7 relâche cette hypothèse.

Hiérarchie microvasculaire. La charge définit trois strates (faible, intermédiaire, élevée). Un gradient posé (cardiovasculaire ×1,0/1,4/2,0 ; décompensation diabétique ×1,0/1,2/1,5 ; mortalité ×1,0/1,2/1,5) suit le sens rapporté par les cohortes populationnelles dans lesquelles le risque cardiovasculaire et le risque d'amputation augmentent avec le nombre de complications microvasculaires [6,7].

Escalade thérapeutique (tronc mensuel). La metformine constitue le socle ; les inhibiteurs de SGLT2, les agonistes du récepteur du GLP-1 et l'insuline basale sont initiés pendant le suivi, interrompus pendant la rupture et repris au retour, en respectant la disponibilité française (AR GLP-1 à partir de 2009, iSGLT2 remboursés à partir de 2020). Leurs effets sont des rapports de risques instantanés poolés issus de méta-analyses d'essais de sécurité cardiovasculaire (CVOT) [8,9] (Tableau 2).

Tableau 2 - Effets des traitements appliqués tant que le patient est traité.

Critère	GLP-1 RA (8 CVOTs)	SGLT2i (6 CVOTs)
MACE (appliqué à l'IDM et à l'AVC)	0.86 (MI 0.91, stroke 0.84)	0.90
Hospitalisation pour insuffisance cardiaque	0.90	0.68
Critère rénal composite (incidence et progression de la MRC)	0.83	0.62
Macroalbuminurie nouvelle	0.74	—
Mortalité de fond (résiduelle, posée)	0.95	0.95

Mortalité et dialyse. La mortalité de fond suit une loi de Gompertz multipliée par l'insuffisance cardiaque, la MRC avancée, la dialyse, la strate microvasculaire et la rupture ; les événements aigus portent leur propre létalité. La MRC de stade 5 conduit à la dialyse, non programmée lorsqu'elle débute pendant une rupture.

Sources des paramètres. Toute donnée externe entre par un registre de provenance (23 sources en v0.17 : identifiant, producteur, URL, licence, usage prévu, statut, empreinte SHA-256 une fois téléchargée). Les sources comprennent le jeu de données synthétique SNDS « top diabète » du Health Data Hub [10] (utilisé pour l'alignement de schéma uniquement, non pour l'épidémiologie), le jeu de données Cnam des dépenses par pathologie (millésime 2024, téléchargé et empreinté) [11], l'indice des prix à la consommation de l'INSEE [12], le registre REIN [13], NHANES, KDIGO [5], la déclaration CKM de l'AHA [4], des études de télésurveillance de l'insuffisance cardiaque [14,15], UKPDS 62 [16] et les méta-analyses de CVOT [8,9]. Une exécution de génération refuse de démarrer si le registre ou le cadre de gouvernance n'est pas conforme.

2.3 Modèles d'observation, interventions et mondes appariés

Têtes d'observation. L'état latent est projeté sur trois canaux dotés de leur propre échantillonnage et de leurs propres données manquantes, jamais sur la vérité elle-même (Tableau 3).

Tableau 3 - Têtes d'observation : ce que chaque canal enregistre, sa cadence et ses données manquantes.

Canal	Ce qu'il enregistre	Cadence	Données manquantes
Données de type SNDS	actes datés sans résultats : délivrances (ATC), consultations de médecin généraliste et de spécialiste, actes de biologie, séjours hospitaliers (motif PMSI, durée, entrée par les urgences), passages aux urgences, dates d'ALD, décès	pilotée par les événements ; un diabétique suivi ≈ 4 consultations de MG/an	structurelles : silence après rupture ; décès signalés à 98 %
IoMT kit	poids, pression artérielle, MCG, FC de repos, activité, sommeil, SpO_2, température plantaire	Quotidien	les dispositifs actifs s'arrêtent ~15 jours avant la rupture administrative ; les passifs ~30 jours après ; couplage MNAR optionnel avant décompensation
Application patient	symptômes déclarés, point mensuel structuré | spontanée ou sollicitée	spontanée ou sollicitée	la non-réponse est un état enregistré, distinct de l'absence de symptôme

La tête SNDS a été alignée sur le schéma de la version synthétique du SNDS publiée par le Health Data Hub (tables ER_PRS_F, ER_PHA_F, IR_IMB_R, PMSI). Les nomenclatures sont des libellés symboliques, avec des codes candidats signalés pour confirmation.

Modèle d'intervention. Chaque mois, une file de vigilance liste les patients dont le silence *observable* dépasse la règle de ciblage (\> 90 jours sans délivrance et \> 180 jours sans consultation de MG). Sous un budget de capacité (0,2 % de la population vivante par mois dans le tronc de long terme, 2–5 % dans les expériences à court horizon), la file est ordonnée soit par silence le plus long (politique A), soit par charge microvasculaire observable (politique B). Un contact réussit avec une probabilité de 0,60 et conduit au retour avec une probabilité de 0,50 ; le retour éteint l'excès de risque après un mois et rétablit les traitements. Chaque alerte coûte 40 €. La probabilité de retour est le paramètre le moins documenté du banc d'essai et elle est traitée comme une variable de politique.

Mondes appariés à nombres aléatoires communs. Selon le principe des nombres aléatoires communs en simulation de maladies [17], chaque mois consomme un tableau fixe de tirages uniformes dont les colonnes ont une signification fixe (rupture, retour, chaque type d'événement, décès, progression, chaque observation, contact). Tous les mondes lisent le même tableau, et un événement survient lorsque le tirage tombe sous le risque instantané du monde. Deux mondes ne diffèrent donc que là où leurs risques instantanés diffèrent, et ils sont identiques pour un patient jusqu'à sa première rupture, propriété garantie par des tests automatisés.

Quatre mondes de base sont exécutés conjointement (la v0.15 ajoute les mondes IoMT PC/PD et la séquence PREDICARE, monde PE, décrits plus bas) :

- **REF** - suivi continu (risque de rupture fixé à zéro) ;

- **SQ** - statu quo (ruptures, retours spontanés, reconnexion après hospitalisation) ;

- **PA -** réengagement seul, file ordonnée par silence le plus long (politique A) ;

- **PB** - réengagement seul, file ordonnée par charge microvasculaire observable (politique B).

En moyenne sur les patients, SQ − REF mesure l'effet de la déshérence et PA − SQ l'effet du réengagement (contrastes populationnels simulés appariés, §2.6), sans biais de comparaison et avec une variance de Monte Carlo réduite. Ces issues contrefactuelles simulées sont connues par construction, non observées : elles ne valent que conditionnellement au mécanisme posé, et les différences pour un patient individuel dépendent en outre du couplage choisi des tirages aléatoires, qui est un dispositif de réduction de variance et non une identification des contrefactuels individuels. Nous ne rapportons donc que des moyennes de population et de strate. Seul le monde SQ est exporté vers des tables de type SNDS ; les autres restent dans un répertoire de vérité terrain séparé.

Variante IoMT après réengagement (mondes PC/PD, v0.15). Dans PA et PB, un patient réengagé revient au suivi standard. PC et PD testent une variante dans laquelle le patient réengagé est en outre équipé d'un kit IoMT et devient mieux suivi qu'un patient diabétique standard, les mesures quotidiennes alimentant un registre longitudinal dont la performance de détection est supposée s'améliorer avec les épisodes qu'il accumule. Les mondes PC (file A) et PD (file B ; implémenté et testé, non rapporté ici puisque le choix de file est examiné au §3.3) mettent cela en œuvre sur un flux aléatoire dédié, de sorte que SQ et PA restent inchangés. Lors d'un retour obtenu par contact, le patient accepte le kit avec une probabilité de 0,80 et l'abandonne à raison de 2 % par mois ; le signal s'arrête pendant la rupture, et sa perte fait du patient un candidat de la file le mois suivant, avant les silences détectés dans les données de remboursement. Une décompensation cardiaque, une insuffisance rénale aiguë ou une décompensation diabétique en cours chez un patient équipé et suivi est détectée avec une sensibilité s(t), puis évitée avec une probabilité de 0,40. Le gain de performance du registre est prescrit, non simulé (aucun algorithme d'apprentissage n'est exécuté) : s(t) = s_min + (s_max − s_min) × R/(R + R½), où R est le nombre cumulé de décompensations observées chez des patients équipés, s_min = 0,05 correspond au hasard pour un budget d'alertes de 5 %, et s_max (0,24 pour l'insuffisance cardiaque, 0,30 pour l'atteinte rénale aiguë, 0,20 pour la décompensation diabétique) ainsi que l'évitabilité sont reportés depuis le tronc quotidien, par un couplage paramétrique des deux troncs. Le plafond pour l'insuffisance cardiaque est le rappel opérationnel mesuré dans l'expérience de prédiction interne (0,24 pour un quota de 5 % par date, §3.1) ; la valeur antérieure (0,39, issue d'un budget poolé du top 5 % dans le tronc quotidien v0.10) est conservée comme cas de sensibilité optimiste. Ce pont est une hypothèse de couplage, non une validation empirique, tout comme l'évitabilité d'un événement détecté (0,40), reportée depuis le tronc quotidien. Ses différences sont traitées comme suit. (i) Unité du quota : dans le tronc quotidien, les dates index sont espacées de 30 jours et le quota signale 5 % des patients équipés à chaque date ; dans le tronc mensuel, le quota est de 5 % des personnes équipées par mois, de sorte que le taux nominal d'alertes est le même, et seul le coût des alertes y est lié. (ii) Alertes répétées et dédoublonnage : dans le tronc quotidien, un patient peut être signalé à des dates successives et chaque signalement compte comme une alerte ; le tronc mensuel ne simule pas les alertes une à une mais tire, pour chaque décompensation à venir d'une personne équipée, si elle est détectée avec une probabilité s(t), de sorte que les alertes répétées, leur dédoublonnage et le nombre de fenêtres précédant un événement ne sont pas représentés. (iii) Définition de l'épisode : dans le tronc quotidien, un positif est une hospitalisation pour insuffisance cardiaque aiguë dans les 30 jours, comptée à une date repère au plus ; dans le tronc mensuel, c'est une décompensation survenant dans le mois. (iv) Taux d'événements :

les personnes équipées dans PE sont surveillées pendant la déshérence et décompensent plus souvent que les insuffisants cardiaques équipés du tronc quotidien, ce qui modifie le rappel à quota fixe dans un sens que nous ne pouvons signer. Aucune de ces différences n'est corrigée ; nous en bornons la conséquence par des plafonds de 0,12 et 0,06 (§3.6). Le registre n'apprend pas dans cet article : aucun algorithme d'apprentissage n'est exécuté dans les mondes de long terme, et son gain de performance est une courbe prescrite. L'équipement coûte 25 € par patient équipé et par mois dans la valeur par défaut du code (110 € dans les résultats rapportés du monde PE ; §2.4), plus 20 € par alerte. Dans les mondes de long terme, le budget d'alertes est un quota mensuel (5 % des personnes équipées par mois), non un seuil de score. Un registre réel devrait définir les épisodes positifs (une décompensation hospitalière dans les 30 jours suivant une date index), les négatifs, un calendrier de réentraînement, une validation hors période et la transférabilité entre patients et territoires ; rien de cela n'est simulé, et la courbe est une hypothèse plausible de gain de performance, non le résultat d'une expérience d'apprentissage. Des variantes restreignent l'équipement aux patients avec insuffisance cardiaque ou MRC observable, pré-alimentent le registre, ou orientent les pertes de signal vers une file de télésurveillance dédiée, hors du budget de réengagement.

Séquence PREDICARE : bilan en unité mobile et télésurveillance pendant la déshérence (monde PE). PC ajoute l'IoMT au réengagement ; le programme prévu commence plus tôt. Le monde PE enchaîne cinq étapes : (1) ciblage des patients silencieux sur les données de remboursement, sous un budget mensuel d'invitations ; (2) invitation, joignabilité (0,60) et venue dans une unité mobile de santé (« médicobus », 0,50 si joint), modulée par la charge microvasculaire latente (×1,15, ×1,0, ×0,75), sélection posée ; (3) bilan clinique et biologique initial à T0 (180 €), qui rend connues la MRC et l'insuffisance cardiaque ; (4) équipement (0,80) et télésurveillance se déroulant pendant la déshérence, avec sa propre attrition (4 % par mois tant que dure la déshérence), que le suivi conventionnel reprenne ou non (0,35) ; (5) alerte et réponse : détection avec la sensibilité prescrite du registre, évitement (0,40), et reconnexion aux soins après une vraie alerte (0,50). La Figure supplémentaire S1 présente le parcours avec ses points d'entrée de coûts (C1–C4) et d'effets sur la santé (H1–H5). Deux variables de politique ne sont pas documentées et sont balayées : la part de l'excès de risque de déshérence effacée par la télésurveillance (de 0 à 1) et le renouvellement des traitements par l'équipe de télésurveillance. L'évaluation conserve le dénominateur de tous les patients invités et sépare l'effet de l'offre (intention de traiter, vérité appariée), l'effet parmi les participants (une strate principale, estimée par l'effet causal moyen chez les observants, CACE = ITT / participation, valide sous la restriction d'exclusion et la non-observance unilatérale [18]) et le contraste naïf participants contre non-participants.

Décisions à T0 et capacité clinique (v0.16). Deux mécanismes ont été ajoutés à PE, avec des valeurs par défaut destinées à reproduire la v0.15 (le test automatisé vérifie seulement qu'une capacité d'alerte non contraignante laisse le monde PE inchangé). (i) Le bilan à T0 mesure la MRC, l'insuffisance cardiaque et une strate microvasculaire élevée avec des sensibilités et un taux de faux positifs posés ; sous une politique stratifiée, seuls les participants classés à haut risque sont équipés. (ii) Les alertes sont traitées par une équipe dont la capacité est plafonnée (alertes examinées pour 100 patients équipés par mois, face à un budget d'alertes de 5), éventuellement priorisées selon le risque à T0 ; seules les alertes traitées sont valorisées, et l'évitabilité est pondérée par $g(d)/g(2)$ avec $g(d) = \max(0, 1 - d/13)$, d étant le délai de réponse en jours et 13 jours le délai d'anticipation médian du tronc quotidien.

2.4 Analyses

Détection à court horizon (apprentissage automatique ; secondaire). La prédiction d'une décompensation hospitalière dans les 30 jours a utilisé un schéma à dates repères (*landmark*) : dates index mensuelles, variables strictement observables à la date index (cadences et historique des remboursements ; pentes et niveaux IoMT sur 7–30 jours ; déclarations dans l'application ; valeurs

biologiques lorsqu'elles existent), critère dans (t, t + 30 jours]. L'algorithme d'apprentissage était le même dans chaque bras, HistGradientBoostingClassifier de scikit-learn (300 itérations, gestion native des valeurs manquantes) [19], avec une partition apprentissage/test par patient figée avant tout filtrage. Seul l'ensemble de variables changeait entre les bras. Les campagnes historiques (v0.2–v0.11, 3 à 5 graines ; Tableau supplémentaire S9) ont fait varier la durée du prodrome, le bruit et le taux d'équipement, comparé par ablation le nombre de ports aux valeurs physiologiques, et couplé l'équipement à la déshérence (gradient social). Pour l'insuffisance cardiaque, l'expérience a été rejouée comme expérience de simulation interne, explicitement non comme une validation : 10 graines ; patients disjoints entre apprentissage (50 %), calibration (20 %) et test (30 %) ; apprentissage sur les dates index jusqu'au jour 420 et test au-delà ; budget d'alertes appliqué sous forme de quota par date ; intervalles par bootstrap en grappes sur les patients au sein des graines (400 réplicats). Les métriques étaient l'AUC-ROC, l'AUC-PR, le rappel et la valeur prédictive positive au quota, le score de Brier et la pente de calibration.

Schémas d'évaluation. Sur les mondes appariés, nous avons comparé les vrais événements évités (ITT, par différence entre mondes) aux estimateurs qu'un programme réel pourrait calculer : alertés revenus contre alertés non revenus, avant/après chez les revenus, un ECR individuel émulé, un essai en escalier (*stepped-wedge*) et un ajustement par pondération sur l'inverse de la probabilité de traitement (IPTW, score de propension logistique) du contraste naïf. Chaque estimateur a été évalué sur 10 graines (tronc quotidien, 40 000 personnes par graine) par son rapport à la vérité, la couverture de son intervalle bootstrap à 95 % (B = 300) et ses taux d'échec (signe erroné ; biais relatif supérieur à 50 %) (Tableau supplémentaire S8).

Émulation de cohorte à partir des données de remboursement. Un script de ciblage ne lisant que les tables de type SNDS a défini :

- **population** : ALD 8 (diabète) ouverte avant 2007 ;
- **silence** : \> 90 jours sans délivrance et \> 180 jours sans consultation de MG, sans décès connu ;
- **index** : dernier contact avant le premier silence observé en 2007–2016, garantissant ≥ 10 ans de suivi ;
- **sorties concurrentes** : réapparition hospitalière, urgences seules, retour en soins primaires, décès, silence persistant, à 5 et 10 ans ;
- **comparateur** : appariement exact 1:1 avec remise au mois index parmi les patients vivants, en contact dans les 6 mois et jamais silencieux auparavant. Les clés de niveau 1 étaient la tranche d'âge de 5 ans, le sexe, la classe d'ancienneté du diabète (\< 5, 5–9, ≥ 10 ans depuis l'ouverture de l'ALD), l'hypertension et la dyslipidémie traitées (12 mois), l'intensité thérapeutique (metformine seule / un traitement ajouté / insuline ou ≥ 2 classes), le tertile de coût remboursé sur les 12 mois précédents, la strate microvasculaire observable, et les ALD de MRC et d'insuffisance cardiaque. Les niveaux 2 et 3 relâchaient les clés de manière hiérarchique. L'équilibre a été évalué par les différences moyennes standardisées (DMS) [20].

Des variables d'événement identiques ont été construites pour les deux cohortes : date et délai de la première hospitalisation, première décompensation (insuffisance cardiaque, insuffisance rénale aiguë sur MRC, décompensation diabétique), premier passage aux urgences et décès, avec censure au décès ou au 31 décembre 2026 ; années de vie et coût par année de vie dans chaque fenêtre ; coût ventilé en pharmacie, soins ambulatoires, dialyse, hôpital et urgences ; épisodes de silence récurrents.

Survie. Courbes de Kaplan–Meier [21] (intention de suivre) ; incidence cumulée par l'estimateur d'Aalen–Johansen [22] avec le décès comme risque concurrent ; courbes selon l'exposition effective censurant chaque patient lorsqu'il change d'état (retour aux soins pour les personnes en déshérence, premier silence ultérieur pour les suivis) ; et taux d'incidence par personne-année selon l'état d'exposition variant dans le temps.

Évaluation économique. Perspective de base : l'assureur obligatoire français (la perspective collective du cas de référence de la HAS est décrite plus bas) ; tous les coûts sont exprimés en euros 2024 (Tableau

supplémentaire S1-C). Les postes tirés du jeu de données Cnam des dépenses par pathologie [11] sont lus dans son millésime 2024 (soins de ville et séjours hospitaliers ; prestations en espèces exclues) : diabète 2 296 € par an, dont une fraction ambulatoire posée de 60 % est appliquée aux mois suivis, et dialyse chronique 67 538 € par an. Les événements aigus (insuffisance cardiaque 5 824 €, syndrome coronarien aigu 8 915 €, AVC 14 209 €) sont valorisés par la composante hospitalière de l'année de l'épisode, et non par le coût annuel de cette année, qui compterait deux fois les soins de ville déjà valorisés mensuellement ; le total annuel est conservé comme analyse de sensibilité en borne haute. Les postes posés sont déclarés aux prix 2024 et signalés. Les coûts variables du programme (contact ou invitation 40 €, bilan à T0 180 €, équipement et plateforme 110 € par mois équipé, traitement d'une alerte 20 €) sont modélisés explicitement (le coût d'équipement inclut environ 86 € par mois pour un capteur de glucose en continu à la base de remboursement française ; la valeur par défaut du code de 25 € est corrigée exactement, le coût n'ayant pas de rétroaction sur les trajectoires ; Tableau supplémentaire S14) ; le coût fixe de fonctionnement ne l'est pas. Utilités : valeur de base 0,785 avec des décréments additifs issus d'UKPDS 62 [16] (IDM −0,055, insuffisance cardiaque −0,108, AVC −0,164, amputation −0,280) et, pour l'insuffisance rénale terminale (−0,263) et la dialyse (−0,305), de valeurs rapportées par McEwan et al. et Brändle et al. telles que tabulées dans une revue systématique d'évaluations économiques du diabète [23] (une revue de modèles d'insuline dans le diabète de type 1 : ce sont des valeurs de seconde main, et les sources primaires de type 2 doivent être vérifiées et citées directement, ou remplacées, au jalon G2) ; la valeur de base se situe dans la fourchette examinée par Beaudet et al. [24]. Coûts et résultats ont été actualisés à 2,5 % par an, comme le recommande la HAS pour les horizons inférieurs à 30 ans [25]. Parce que le coût fixe de fonctionnement F de chaque stratégie (réengagement seul ou séquence PREDICARE) est inconnu, il a été traité comme un paramètre (€ par patient couvert et par an, appliqué aux patients-années couverts actualisés, environ 11 par personne diabétique sur 20 ans) : RDCR(F) = (ΔC + F × patients-années couverts) / ΔQALY, et nous rapportons la *marge* (*headroom*), le F maximal compatible avec des seuils de lecture de 30, 50 et 100 k€/QALY (la France n'a pas de seuil officiel). La marge est une grandeur de seuil, non un budget : elle est exprimée par personne atteinte de diabète de type 2 dans le territoire couvert (estimand E4) et ne doit couvrir que des ressources qui ne varient pas avec le volume (unité mobile, effectif de base, licence de plateforme, coordination ; Tableau supplémentaire S1-C). Les résultats économiques sont donc un exemple traité d'analyse de marge, conditionnel à des effets hypothétiques déclarés, et non une estimation économique de PREDICARE. Les analyses de scénarios univariées ont couvert 13 scénarios, chacun exécuté sur 10 graines. Parce que le cas de référence de la HAS est une perspective collective incluant l'aide informelle [25] (pertes de productivité uniquement en analyse complémentaire), la v0.15 ajoute une couche de valorisation sans rétroaction : les survivants d'un AVC, d'une amputation, d'une mise en dialyse ou d'une décompensation cardiaque deviennent dépendants à domicile ou institutionnalisés avec des probabilités posées, en plus de la perte d'autonomie liée à l'âge ; la dépendance est valorisée avec les coûts unitaires DREES 2019 [26] convertis en euros 2024 par l'indice des prix à la consommation de l'INSEE (facteur 1,1496) [12] (555 € par mois à domicile ; 2 155 € d'hébergement plus 586 € de dépendance par mois en EHPAD) et des heures d'aide informelle posées (à confronter aux enquêtes CARE [27]) au coût de remplacement. Une analyse de sensibilité probabiliste a tiré conjointement 23 paramètres, selon les recommandations ISPOR–SMDM sur l'analyse d'incertitude [28], dans des distributions déclarées (lognormale, bêta, uniforme) centrées sur les valeurs de base (mécanisme de déshérence, réengagement, coûts, utilités et IoMT), et a resimulé les mondes appariés pour chacun des 300 tirages (20 000 patients par tirage) ; une seconde analyse sur le monde PE a ajouté sept distributions propres à la séquence (Tableau S3).

Erreur de Monte Carlo et valeur de l'information. Les résultats déterministes de long terme utilisent 10 graines (42–51) de 60 000 personnes (30 000 pour la couche de perspective collective) et sont rapportés en moyenne ± erreur-type de Monte Carlo (MCSE : écart-type inter-graines divisé par √10), avec [min ; max] sur les graines ; les contrastes appariés sont calculés graine par graine. Les différences inférieures à environ deux MCSE ne sont pas interprétées. Dans les analyses probabilistes, chaque tirage simule 20 000 personnes, de sorte que les distributions brutes combinent l'incertitude sur les paramètres (de second

ordre) et l'incertitude de Monte Carlo (de premier ordre). La variance de premier ordre, estimée au sein des tirages, a été soustraite de la variance inter-tirages (analyse de variance [29]) ; elle représentait 5–12 % de la variance des grandeurs rapportées, de sorte que les intervalles bruts surestiment légèrement l'incertitude sur les paramètres. Pour la séquence PREDICARE, le bénéfice net incrémental d'un tirage était 50 000 × ΔQALY − ΔC − F × personnes-années couvertes ; la valeur attendue de l'information parfaite (EVPI) et de l'information parfaite partielle (EVPPI) ont été estimées à partir des 300 tirages par régression non paramétrique de ce bénéfice net sur les paramètres [30] (polynôme cubique pour chaque paramètre, ajustement additif pour les groupes), avec des erreurs-types bootstrap sur les tirages ; les valeurs non distinguables de leur erreur-type ne sont pas interprétées.

2.5 Gouvernance et contrôle qualité

Le banc d'essai s'applique à lui-même la gouvernance qu'il défend : aucune assertion ne peut voyager au-delà du domaine de validité du composant qui l'a produite. Ces contrôles disciplinent le développement et rendent les hypothèses inspectables ; ils ne se substituent ni à une validation clinique ni à une revue indépendante du modèle causal [31].

Domaines de validité. Chaque composant (ontologie, DAG, chaque tête d'observation, chaque émulateur, valorisation, utilités, registre de provenance ; 21 composants) déclare dans un fichier lisible par machine ce qu'il permet d'affirmer, ce qu'il interdit, et son statut épistémique courant. Chaque composant est au statut *hypothèse*, sauf le registre de provenance (*convention*). Par exemple, le tronc de long terme autorise des énoncés sur la *forme* et le *signe* des différences entre mondes appariés et interdit toute incidence, mortalité ou montant en euros réels.

Jalons de décision. Un statut ne peut changer que par un jalon, jamais par la rédaction : G0 adoption du cadre au lancement ; G1 atelier clinique (hypothèse → convention) ; G2 calibration sur une source enregistrée (→ calibré) ; G3 décision go/no-go du démonstrateur ; G4 confrontation aux données réelles (→ corroboré). Chaque jalon désigne une autorité — certaines non délégables à l'équipe technique — et ce que son franchissement n'autorise *pas*. Une vérification automatisée refuse les transitions illégales ; pendant le développement, elle en a intercepté et bloqué une.

Provenance. Aucune donnée externe n'entre sans une entrée au registre ; une source téléchargée doit porter son empreinte SHA-256, et chaque exécution de génération écrit un manifeste avec les empreintes des fichiers d'ontologie, de physiologie, de gouvernance et de registre.

Déterminisme et invariance. Les flux aléatoires sont isolés par patient × composant × sous-flux via SeedSequence de NumPy, de sorte que modifier une politique d'équipement n'altère pas les trajectoires latentes (un bogue précoce de ce type a été détecté et corrigé). Une graine donnée reproduit les données octet pour octet.

Tests automatisés (21 ; Tableau supplémentaire S13). Une chaîne de recette vérifie chaque jointure de la chaîne applicative (ciblage → jumeau individuel → détection de dérive → file d'attente → cockpit), y compris un test négatif volontairement en échec : lorsqu'une personne extérieure à l'ensemble des silencieux observables est injectée dans la liste ciblée, la chaîne doit échouer. D'autres tests vérifient le déterminisme ; l'identité du suivi continu et du statu quo avant la première rupture d'une personne ; l'identité de REF, PA et PB avec SQ pour les personnes jamais en rupture ; que le monde PC est égal à PA sans équipement, et que l'ajout du monde PC ou l'activation d'un mécanisme structurel laisse les autres mondes inchangés (flux aléatoires isolés) ; qu'une capacité d'alerte non contraignante laisse le monde PE inchangé ; la séparation stricte entre les tables de remboursement et le répertoire de vérité (le script de ciblage refuse de s'exécuter sur ce dernier) ; et une précision (\> 0,85) et un rappel (\> 0,80) minimaux

du ciblage sur données de remboursement par rapport à la définition de rupture du simulateur. Que la v0.18 reproduise la v0.17 octet pour octet sous sa structure par défaut a été vérifié séparément sur les sorties d'exécution. Aucun de ces tests ne vérifie l'exactitude clinique des mécanismes.

Registres de décision. Vingt-cinq registres de décision d'architecture (ADR-001 à ADR-025) documentent chaque choix structurel, ses alternatives rejetées et ses conséquences.

Revue contradictoire. Un fichier de contradiction écrit liste les dix objections les plus fortes au programme (circularité, seuils, paternalisme, faux positifs, gradient social, évitement inobservable…), avec la position actuelle, la faiblesse et les éléments de preuve qui trancheraient chacune. Une revue critique du banc d'essai (13 biais) et une revue épidémiologique externe du classeur de cohorte (variables d'issue asymétriques, facteurs de confusion manquants) ont conduit aux corrections rapportées ici.

Validation du modèle. Selon la Task Force 7 ISPOR–SMDM [31] :

- **validité apparente** : la structure causale et les paramètres ont été examinés par l'auteur et par la revue contradictoire ci-dessus, pas encore par un panel clinique indépendant ;

- **vérification** : les tests automatisés et la chaîne de recette ci-dessus, y compris les contrôles d'invariance entre versions et entre mondes (Tableau supplémentaire S13) ;

- **validité croisée** : aucune comparaison avec un modèle établi du diabète de type 2 (par exemple l'UKPDS Outcomes Model) n'a été faite ;

- validité externe et prédictive : aucune. La calibration sur l'incidence et la mortalité françaises est prévue au jalon G2 ; le Tableau supplémentaire S2 liste les taux agrégés simulés, dont plusieurs (notamment les mises en dialyse et l'infarctus du myocarde) paraissent élevés au regard des ordres de grandeur français. Le banc d'essai est donc validé, au mieux, pour son usage déclaré (tester si des méthodes retrouvent une vérité connue) et non pour prédire des résultats en France.

2.6 Estimands

Chaque grandeur rapportée dans cet article est l'un des estimands du Tableau 4, ou un estimateur de l'un d'eux. Trois types sont distingués tout au long du texte. Les *contrastes populationnels simulés appariés* comparent la même population simulée sous deux stratégies ; les deux mondes partagent des nombres aléatoires communs, ce qui réduit la variance de Monte Carlo, et les contrastes sont des moyennes sur une population, rapportées uniquement comme telles. Les *contrastes de sous-groupes définis par le modèle* sont des contrastes dans des sous-groupes définis par une issue dans un monde, par exemple « les patients qui réapparaissent à l'hôpital dans le statu quo » ; ce sont des strates principales [32], dont les valeurs dépendent du modèle posé et du couplage choisi des tirages aléatoires, et qui ne sont pas identifiables dans des données réelles. Les *estimateurs observationnels* sont ce qu'une étude restreinte aux données de remboursement pourrait calculer ; leur biais est la différence entre l'estimateur et le contraste apparié correspondant.

Ce que « supprimer la déshérence » signifie. Le contraste SQ − REF (E1) est l'effet d'une intervention hypothétique qui empêcherait toute rupture, maintenant chaque personne dans le régime d'observation, de traitement et de risque d'un patient suivi. Il agrège les trois composantes d'une rupture (§2.2) et ne correspond à aucune politique de santé unique : dans la réalité, la déshérence mêle accès interrompu, choix du patient, maladie silencieuse, déménagement, suivi hors du système remboursé et

interruption de traitement, qui appellent des réponses différentes. L'expérience structurelle fait varier certaines de ces sources (soins hors système, ruptures partielles, causalité inverse ; §2.7), pas toutes. Les contrastes pertinents pour la décision publique sont ceux d'une intervention qui peut effectivement être proposée : le réengagement (PA − SQ) et l'offre PREDICARE (E5).

Le temps est compté en mois de simulation. Le mois 0 est janvier 2006, le suivi débute au mois 12, et la fenêtre d'entrée pour l'indexation couvre les mois 12–131, de sorte que chaque index est suivi au moins 120 mois et qu'il n'y a aucune censure avant 10 ans, sauf par décès. Le décès fait partie de chaque issue : coûts et QALY ne s'accumulent que tant que la personne est en vie. Les retours aux soins, les nouvelles ruptures, l'escalade thérapeutique et l'interruption de traitement surviennent tous dans l'horizon et ne sont pas des événements de censure dans les contrastes appariés ; dans l'analyse de survie selon l'exposition effective (Tableau 7), un changement d'état est un événement de censure par construction de cette analyse.

Tableau 4 - Estimands : population et dénominateur, temps zéro, contraste, critère et statut.

#	Estimand (where reported)	Population, denominator and time zero	Contrast, outcome and summary measure	Status
E1	Effet de la déshérence parmi les silencieux identifiés dans les données de remboursement (Tableau 5 ; ligne vérité du Tableau 6 ; Figure 3)	Personnes atteintes de DT2 dans le monde statu quo qui satisfont pour la première fois le critère de silence observable (\> 90 jours sans délivrance et \> 180 jours sans consultation de MG) au cours des mois 12–131 : toutes ces personnes (environ 18 100 sur 60 000), quelle que soit leur évolution ultérieure. Temps zéro : dernier contact observé avant le silence	| SQ contre REF pour les mêmes personnes simulées ; décès à 5 et 10 ans, coût payeur cumulé par composante (non actualisé), années de vie dans la fenêtre ; différence de moyennes	Contraste apparié, mais sur une population sélectionnée sur un événement (le silence) qui ne survient que dans SQ : défini par le modèle, conditionnel au couplage
E2	Coût et mortalité dans des sous-groupes définis par l'issue (Tableau 9 ; Tableau supplémentaire S12 ; Figure 2)	Sous-ensembles de E1 définis par la première sortie observée dans SQ et/ou par la strate microvasculaire latente au temps zéro	Comme E1	Strate principale définie par le modèle ; non identifiable dans des données réelles. Seules les lignes par strate observable dans les données de remboursement décrivent des sous-groupes qu'une étude réelle pourrait constituer
E3	Estimateurs de E1 fondés sur les données de	(a) Tous les silencieux contre des suivis appariés exactement	Silencieux observés contre suivis observés, tous deux	Estimateurs ; biais = estimateur − E1

	remboursement (Tableau 6, lignes 2–4 ; Tableau 7)	1:1 (avec remise, au mois index) ; (b) le sous-ensemble de (a) réapparu à l'hôpital contre les mêmes comparateurs. Temps zéro : mois index	dans SQ ; critères comme E1, plus survie de Kaplan–Meier, d'Aalen–Johansen et selon l'exposition effective, et rapports de taux par état variant dans le temps	
E4	Coût-utilité populationnel d'une stratégie (Tableaux 8, 10, 11 ; Figures 5 et 6 ; Tableaux supplémentaires S11–S12 ; les deux analyses probabilistes)	Toutes les personnes simulées atteintes de DT2 vivantes au mois 12 (60 000 par graine ; 20 000 par tirage) : chaque personne atteinte de DT2 dans le territoire couvert, et non les personnes ciblées, invitées ou équipées. Temps zéro : mois 12	Monde stratégie (PA, PB, PC, PE, REF) contre SQ, chacun étant une politique mensuelle dynamique ; coûts et QALY actualisés à 10 et 20 ans ; ΔC par personne, ΔQALY pour 1 000, marge F_max(λ) = (λ·ΔQALY − ΔC) / personnes-années couvertes (environ 11 personnes-années actualisées par personne sur 20 ans)	Contraste populationnel apparié. La marge est une grandeur de seuil, non un budget, et n'exclut rien de ce que le modèle compte déjà comme coût variable (Tableau supplémentaire S1-C)
E5	Effet de l'offre (intention de traiter) parmi toutes les personnes éligibles, séquence PREDICARE (§3.5 ; Figure 6, droite)	Toutes les personnes qui satisfont pour la première fois le critère de ciblage sur données de remboursement (silence observable) au cours des mois 12–131, qu'elles soient invitées ou non (file limitée par le budget ; réinvitation possible 6 mois après un contact infructueux) : la population de E1, environ 18 100 sur 60 000. Temps zéro : mois où le critère est satisfait pour la première fois. Sous randomisation en grappes, la même règle appliquée dans les deux bras avant l'allocation définit le dénominateur	PE contre SQ pour les mêmes personnes (trajectoires identiques jusqu'à la première invitation, qui requiert le silence) ; décès à 10 ans et coût payeur à 10 ans ; différence de risques, € par personne éligible	Contraste apparié. Effet de la politique simulée, budget et file compris (et non d'une invitation reçue par chaque personne éligible). Estimand principal pour le démonstrateur, identifiable en randomisant des grappes avant le début du ciblage, chacune avec sa propre capacité (§3.5). L'effet parmi les personnes effectivement invitées (temps zéro : première invitation) est secondaire : cette population est sélectionnée par la file et le budget du programme
E6	| Effet parmi les participants (Figure 6, droite)	Personnes éligibles venues au moins une fois dans l'horizon après	| PE contre SQ, comme E5 ; CACE = ITT / effet de premier	Strate principale (observants), identifiée par le CACE sous non-

		des invitations répétées (18,3 % de E5 dans le scénario de référence ; 67,1 % des invités)	niveau sur la participation	observance unilatérale (vérifiée par construction du schéma) et restriction d'exclusion (vérifiée par construction ; doit être argumentée sur le terrain). Secondaire
E7	Contraste naïf participants contre non-participants (Figure 6, droite)	Participants contre non-participants parmi les invités, tous dans PE (temps zéro : première invitation)	Différence de mortalité observée à 10 ans	Estimateur, biaisé par la sélection posée ; présenté uniquement pour quantifier ce biais
E8	Exactitude du ciblage sur données de remboursement (Tableau supplémentaire S9-D)	Personnes vivantes présentant un signal de diabète observable, évaluées chaque mois	Sortie de la règle contre état de rupture du simulateur ; précision, rappel	Concordance avec la propre définition de rupture du simulateur, par une règle construite dans le même univers ; non une estimation de performance pour le SNDS réel
E9	Discrimination à court horizon du détecteur IoMT (§3.1 ; Tableau supplémentaire S9)	Personnes équipées avec insuffisance cardiaque connue dans les données de remboursement, à des dates repères tous les 30 jours ; patients de test disjoints de l'apprentissage, dates index après le jour 420	Données de remboursement seules contre données de remboursement + IoMT (même algorithme) ; ≥ 1 hospitalisation pour insuffisance cardiaque aiguë dans (t, t + 30 jours] ; AUC-ROC, AUC-PR, calibration, rappel et VPP à un quota par date	Expérience de simulation interne sur un prodrome injecté ; non une validation de modèle

Unités et interférence. Sauf mention contraire, un montant en euros dans les Tableaux 8, 10 et 11 s'entend « par personne atteinte de diabète de type 2 dans le territoire couvert » (E4). Lorsque le texte traduit la marge en montants par personne invitée, par participant ou par patient-année équipé (§3.5), il redistribue le même total actualisé sur un dénominateur plus petit ; ces traductions sont étiquetées comme telles et ne remplacent jamais E4. Les personnes n'interagissent qu'à travers des ressources partagées à capacité limitée (budgets mensuels, capacité de traitement des alertes et courbe du registre), de sorte que E4 et E5 sont définis pour la politique dans son ensemble, sous cette interférence ; le §3.5 précise quel effet chaque schéma d'essai possible estimerait. L'expression « mêmes patients » est réservée aux contrastes appariés et signifie toujours « sous le modèle et le couplage supposés ».

2.7 Expérience structurelle : mécanismes générateurs de données alternatifs

Les analyses principales explorent des valeurs de paramètres au sein d'une structure causale unique, dans laquelle une rupture cause la totalité de l'excès de risque. Pour tester si les biais des schémas sont des propriétés de l'évaluation sur données de remboursement ou de cette structure, nous avons rejoué les schémas fondés sur les données de remboursement sous 14 mécanismes générateurs de données qui changent *qui cause quoi* (v0.18 ; Tableau supplémentaire S7). Laissé à ses valeurs par défaut, le bloc structurel de la v0.18 reproduit la v0.17 octet pour octet (vérifié sur les sorties d'exécution ; un test automatisé vérifie que le flux structurel laisse le monde REF inchangé).

- **Référence causale (M0).** Le mécanisme du §2.2.
- **Déshérence comme marqueur de gravité latente (M1–M3).** Une fragilité latente *Z* (lognormale, moyenne 1, écart-type du log 0,6), jamais observée, multiplie le risque instantané de rupture et, dans chaque monde, les risques instantanés d'événements et la mortalité de fond. Seule une fraction θ de l'excès de risque de rupture est causale : θ = 0,5 (M1) ou θ = 0 (M2). Dans M3, θ = 0 et le traitement se poursuit aussi pendant la rupture, de sorte que la déshérence ne cause rien sur le plan clinique.
- **Confusion par la précarité ou l'accès (M4).** Effet causal complet, plus une fragilité partagée agissant sur la rupture et, plus faiblement, sur les événements.
- **Causalité inverse par maladie aiguë (M5).** Le risque instantané de rupture est multiplié par 4 dans les 3 mois suivant tout séjour hospitalier.
- **Déshérence partielle et intermittente (M6).** Pharmacie résiduelle dans 80 % des ruptures, pendant 6–12 mois, et retours spontanés rapides.
- **Soins hors du système observable (M7).** 30 % des silences dans les données de remboursement ne sont pas des ruptures cliniques : soins et traitements se poursuivent, non enregistrés et non payés par l'assureur.
- **Reconnexion absente, réduite ou retardée (M8, M8b).** Pas de reconnexion, ou une reconnexion réduite, après un séjour hospitalier ; excès de risque éteint sur 6 mois après le retour.
- **Degrés d'interruption du traitement (M9, M9b).** La moitié, ou aucune partie, de l'effet du traitement est perdue pendant la rupture.
- **Ampleur et délai de l'effet (M10, M11).** Excès de risque divisé par deux et débutant dès le premier mois ; ou différé au mois 8.

Pour chaque mécanisme et chacune des 10 graines (60 000 personnes), nous avons calculé la vérité appariée (E1, 10 ans) et les estimateurs sur données de remboursement (E3) : tous les silencieux contre des suivis appariés (coût) ; personnes réapparues à l'hôpital contre suivis appariés (coût) ; différence de mortalité en intention de suivre ; et rapport des taux de décès selon l'exposition effective, silence contre suivi. Les mesures de performance étaient le biais (estimateur − vérité, moyenne ± MCSE sur les graines) ; la couverture, c'est-à-dire la part des graines dans lesquelles l'intervalle à 95 % intra-graine de l'estimateur contient la vérité de la même graine (ces intervalles utilisent des erreurs-types appariées et ignorent la réutilisation des témoins, ils sont donc anticonservateurs) ; et les taux d'échec (signe opposé à la vérité ; biais relatif supérieur à 50 %). Les mécanismes ne sont pas recalibrés pour maintenir constante l'association observée entre rupture et événements, car cela exigerait la calibration du jalon G2. Chaque mécanisme a donc sa propre vérité, et seule la relation entre estimateur et vérité est interprétée.

3. Résultats

Les résultats proviennent de versions successives du banc d'essai, et chacun est étiqueté avec la version qui l'a produit. Pour cette révision, chaque résultat de long terme (§3.2–3.6), à l'exception de l'émulation de cohorte à graine unique du §3.3 (graine 42, n = 20 000), a été rejoué sur 10 graines, sous la v0.17 (coûts en euros 2024 ; mondes REF, SQ, PA et PB) ou la v0.19 (plafond du registre pour l'insuffisance cardiaque 0,24 ; mondes PC et PE ; monde PE à un coût de télésurveillance de 110 € par mois équipé, rejoué avec la v0.19.2 ou corrigé exactement) ; l'expérience structurelle (§3.7) a utilisé la v0.18, dont la structure par défaut est identique à celle de la v0.17. Les résultats du tronc quotidien qui n'ont pas été rejoués conservent leur étiquette de version d'origine. La provenance de chaque résultat (version, réplication, population et script) est résumée dans la section Disponibilité des données et du code et détaillée dans le Tableau supplémentaire S6.

3.1 Détection, observabilité et schémas d'évaluation

La détection à court horizon est une expérience secondaire et interne. Les campagnes historiques de détection (v0.2–v0.11 ; Tableau supplémentaire S9) ont montré que les données de remboursement seules étaient proches du hasard pour la décompensation à 30 jours, et que le capteur utile dépendait de l'organe : signaux IoMT quotidiens pour l'insuffisance cardiaque, biologie délocalisée dense pour l'insuffisance rénale aiguë sur MRC. Dans l'expérience de prédiction interne rejouée pour cette révision (v0.18 ; pas une validation), dix graines ont fourni des ensembles de test poolés de 1 998 patients équipés avec insuffisance cardiaque connue dans les données de remboursement, observés à 15 730 dates repères, avec 176 hospitalisations distinctes pour insuffisance cardiaque aiguë (12–24 par graine). L'AUC-ROC était de 0,57 [IC 95 % 0,52–0,62] avec les données de remboursement seules et de 0,70 [0,65–0,74] avec les variables IoMT (étendue entre graines 0,37–0,69 et 0,61–0,79). Avec un quota d'alertes de 5 % des patients à chaque date index, le modèle IoMT a capté 24 % des épisodes avec une valeur prédictive positive de 5,3 % par alerte (patient-date), soit environ 19 alertes par vrai épisode ; les alertes répétées pour un même patient à des dates successives ont été comptées séparément. Les scores bruts étaient mal calibrés (pente 0,37), et une recalibration isotonique sur des ensembles de calibration d'environ dix événements par graine ne l'a pas corrigé (pente 0,12 ; Figure supplémentaire S2). Ce sont des propriétés d'un prodrome injecté détecté par un algorithme sur des patients simulés, et non la preuve qu'un registre réel apprendrait une signature transférable. La charge d'alertes est une contrainte de conception au moins aussi contraignante que la discrimination ; dans PE, la capacité clinique derrière les alertes est examinée au §3.6.

Résultats historiques du tronc quotidien. Les campagnes antérieures (v0.1–v0.14, non rejouées) sur l'ablation des décomptes de port des dispositifs, les données manquantes informatives, l'équipement couplé à la déshérence future et le ciblage sur données de remboursement sont rapportées dans le Tableau supplémentaire S9-D. Les résultats de ciblage qui y figurent sont une concordance avec la propre définition de rupture du simulateur, par une règle construite dans le même univers, et ne prédisent pas la performance sur le SNDS réel.

Les contrastes fondés sur les participants sont biaisés exactement dans la mesure de la sélection que l'on pose. Avec dix graines (v0.18, tronc quotidien ; Tableau supplémentaire S8), le vrai nombre de décompensations évitées après réengagement était en moyenne de 9,5 par graine (étendue 6–18). L'ECR individuel était sans biais en moyenne (rapport à la vérité 1,03 ± 0,45) mais si imprécis que son estimation avait le mauvais signe dans 3 graines sur 10. L'essai en escalier surestimait d'un facteur deux environ (2,19 ± 0,81), parce que l'exposition est confondue avec le temps calendaire, et les comparaisons avant/après chez les revenus échouaient (−0,23 ± 0,17). Le contraste naïf revenus contre non revenus et

son ajustement IPTW étaient sans biais en moyenne dans cette expérience (1,02 ± 0,15 et 1,01 ± 0,16 ; couverture 90 %), parce que le retour après contact est tiré indépendamment du risque dans le tronc quotidien ; la surestimation rapportée dans une version antérieure (×1,66–1,71 sur trois graines, v0.5–v0.6) n'est pas reproduite et est retirée. Rapproché du résultat de long terme, où le contraste naïf participants contre non-participants surestimait l'effet d'un facteur six environ parce que la participation était posée comme dépendante de la charge latente (§3.5), cela montre que le biais des comparaisons fondées sur les participants est porté par le mécanisme de sélection que l'on suppose, qu'une étude réelle ne peut pas observer.

3.2 Trajectoires de long terme, coûts et biais des schémas observationnels

Les résultats de cette section proviennent du tronc mensuel (v0.17, 10 graines × 60 000 personnes, environ 18 100 satisfaisant le critère de silence par graine) et sont rapportés en moyenne ± MCSE [min ; max] sauf mention contraire. Les contrastes sont des contrastes simulés appariés sous le modèle supposé (E1) ou des sous-groupes définis par le modèle (E2), à l'exception des estimateurs sur données de remboursement (E3).

Une cohorte définie par la réapparition voit moins de la moitié de la déshérence. Le critère de silence a capté 89,4 % des vraies premières ruptures. En 10 ans, 51 % des personnes en déshérence sont réapparues par un séjour hospitalier, 24 % par un passage aux urgences seul, 19 % sont revenues en soins primaires et 5,7 % sont décédées sans réapparaître. Au total, seules 46 % des vraies ruptures entreraient dans une cohorte définie par la réapparition hospitalière (Figure 2).

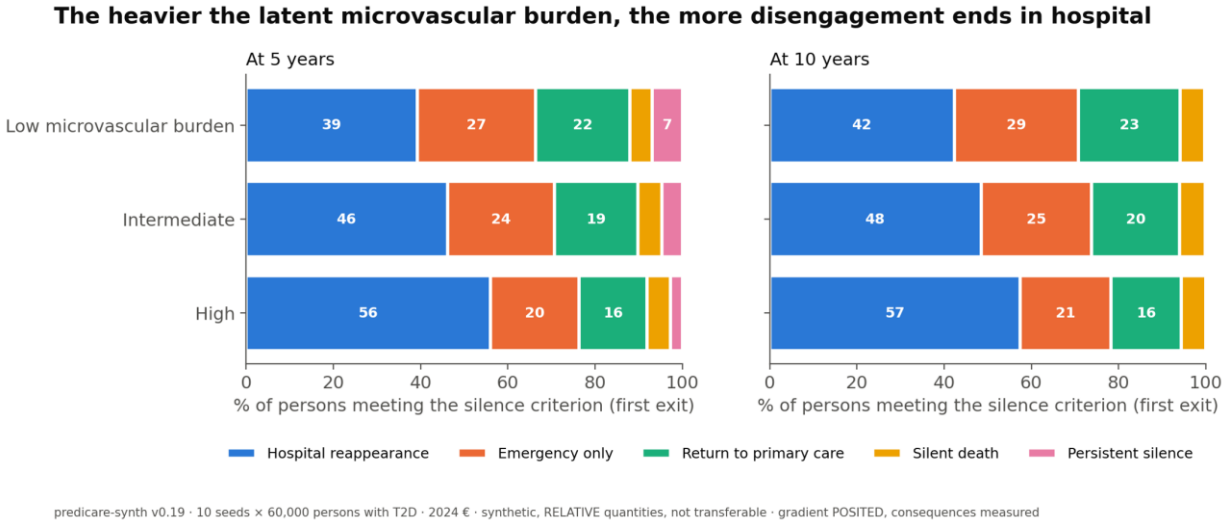

Figure 2 — Sorties concurrentes de la déshérence selon la strate microvasculaire latente. Le gradient sur les risques instantanés d'événements est posé ; ses conséquences sur les sorties et les coûts sont mesurées.

La hiérarchie microvasculaire façonne les issues, mais les données de remboursement ne la voient pas. La réapparition hospitalière à 10 ans passait de 42 % (charge faible) à 48 % (intermédiaire) et 57 % (élevée). La mortalité à 10 ans était de 33,4 % contre 27,4 % (faible), 38,9 % contre 33,0 % (intermédiaire) et 46,1 % contre 40,4 % (élevée) dans le statu quo comparé au suivi continu. Pourtant, seuls 14,0 % des patients à charge latente élevée étaient classés en charge élevée à partir des traces de remboursement, et les strates faible et intermédiaire étaient indistinguables (86–87 % classés en charge faible) : la rétinopathie non proliférante, l'albuminurie et les lésions du pied de grade 1 ne laissent aucune trace

dans la tête de remboursement simulée, par construction (dans le SNDS réel, les examens du fond d'œil et les dosages d'albuminurie sont des actes remboursés, mais sans leurs résultats).

Sous le modèle supposé, la déshérence coûte des vies mais pas d'argent (Tableau 5, Figure 3).

Tableau 5 - Différences simulées appariées, statu quo moins suivi continu, parmi les personnes satisfaisant le critère de silence (fenêtre de 10 ans, perspective payeur, non actualisé ; moyenne ± MCSE [min ; max] sur 10 graines). Statut : résultat répliqué (10 graines).

Fenêtre de 10 ans, SQ − REF (E1)	Toutes strates	Charge latente élevée
Différence de coût total	−1,981 ± 76 € [−2,282; −1,480]	−241 ± 160 € [−933; +782]
dont soins ambulatoires et pharmacie	−9,004 ± 26 €	−9,698 ± 41 €
dont séjours hospitaliers	+3,817 ± 14 €	+4,145 ± 15 €
dont dialyse	+3,141 ± 83 €	+5,256 ± 148 €
Mortalité (points)	+5.8 ± 0.1 (41.2% vs 35.3%)	+5.7 ± 0.1
Années de vie perdues dans la fenête	0.42	0.44

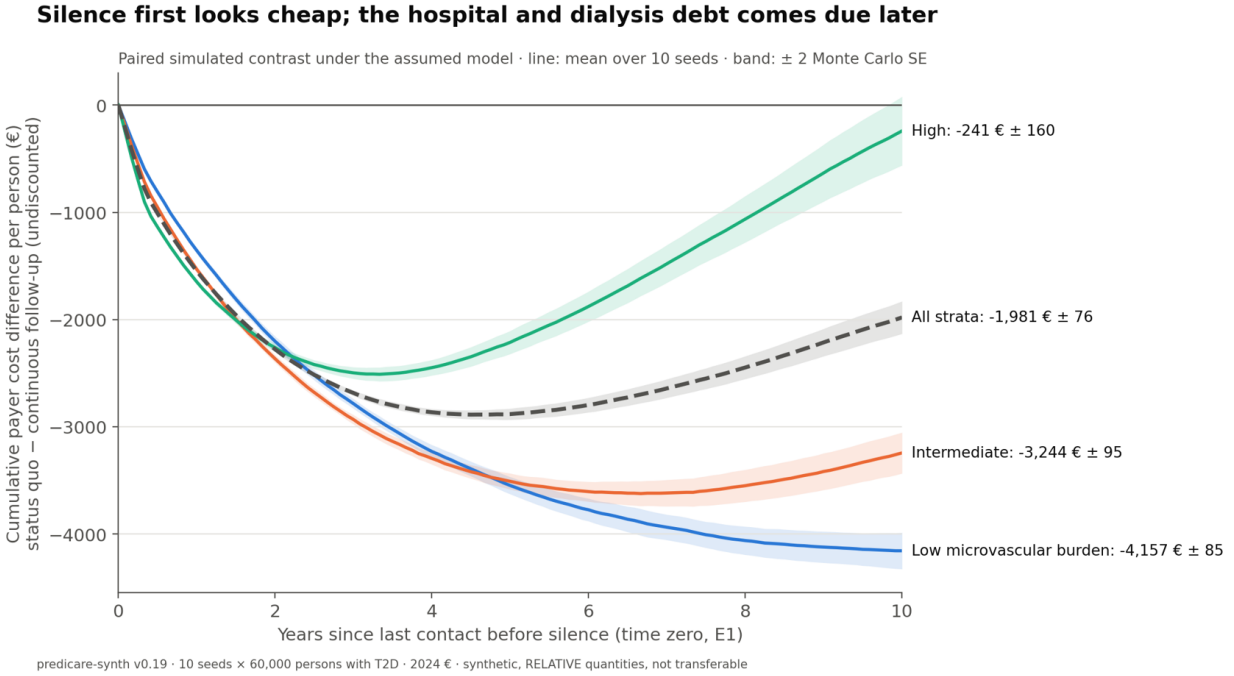

Figure 3 - Différence de coût cumulé, statu quo moins suivi continu, sous le modèle supposé. Le silence paraît d'abord bon marché (soins non consommés) ; la dette hospitalière et de dialyse vient à échéance plus tard. Dans la strate à charge élevée, la différence nette à 10 ans n'est pas distinguable de zéro (−241 ± 160 €), mais ses composantes dialyse et hôpital sont robustes (+5 256 ± 148 € et +4 145 ± 15 €). Bande : ± 2 MCSE sur 10 graines.*

Parmi les patients encore silencieux et vivants à 5 ans, la dette latente était mesurable : MRC de stade 4–5 chez 8,4 % contre 6,0 %, strate élevée 66,8 % contre 57,5 %, et les coûts des années 5–10 dépassaient ceux du suivi continu de 1 610 € après une économie de 4 600 € sur les années 0–5.

Restreindre la cohorte aux patients réapparus surestime le coût de la déshérence (Tableau 6, Figure 4).

Tableau 6 — Surcoût attribué à la déshérence : vérité appariée contre deux schémas de cohorte sur données de remboursement (non actualisé ; moyenne ± MCSE sur 10 graines). Statut : résultat répliqué (10 graines).*

Surcoût attribué à la déshérence	5 ans	10 ans
Vérité (contraste simulé apparié, E1)	−2,882 ± 26 €	−1,981 ± 76 €
Étude sur données de remboursement : tous les silencieux contre suivis appariés	−5,020 ± 85 €	−4,628 ± 212 €
Étude sur données de remboursement : patients réapparus seuls contre suivis appariés	−1,165 ± 164 €	**+1,106 ± 372 €** [−1,096; +3,442]
Biais du schéma « réapparus seuls » (étude − vérité)	+1,717 ± 154 €	+3,087 ± 323 € [+1,183; +4,922]

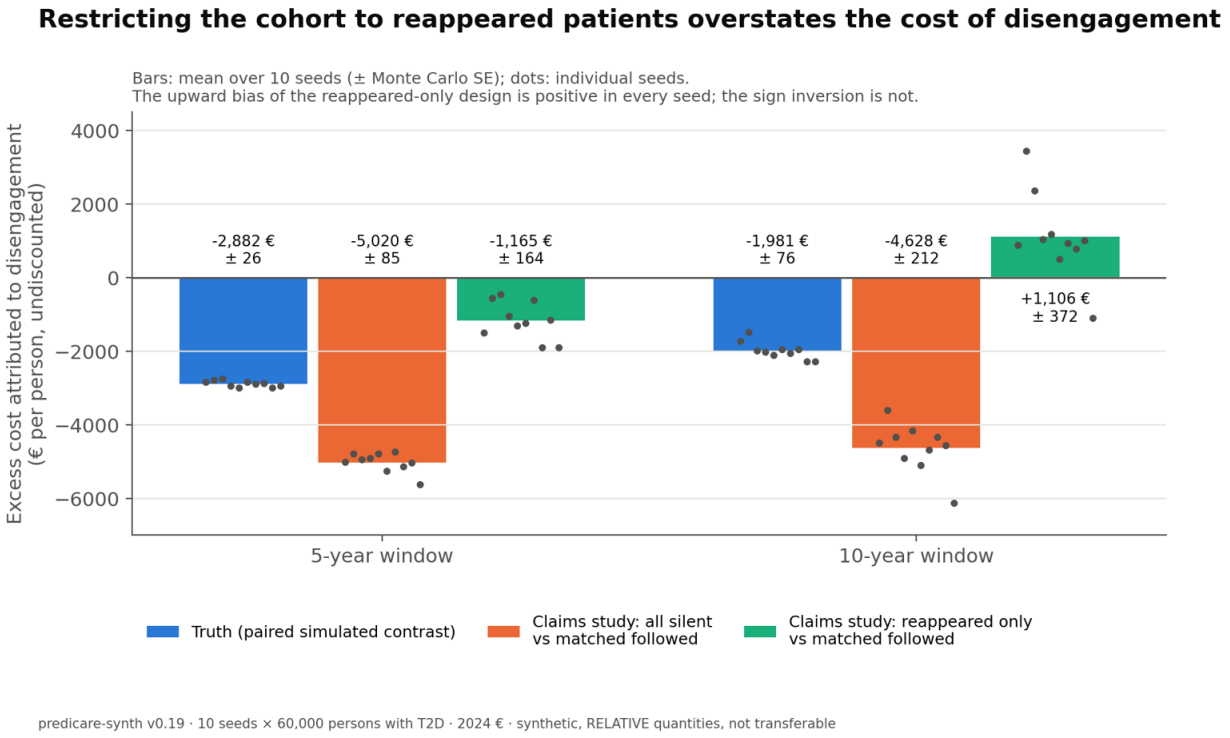

Figure 4 — Sélectionner sur la décompensation fabrique une partie du surcoût que l'on cherche à montrer. Barres : moyenne ± MCSE sur 10 graines ; points : graines individuelles.*

Le biais à la hausse du schéma « réapparus seuls » était positif pour chaque graine ; l'inversion de signe elle-même ne l'était pas (une graine sur dix donnait une estimation négative à 10 ans). Le résultat robuste est le sens et l'ampleur du biais, non l'inversion.

Sensibilité. Sur 13 scénarios (10 graines × 60 000 chacun), trois résultats ont tenu : l'excès de mortalité à 10 ans (+5,3 à +9,6 points) ; l'écart de coût entre les strates à charge élevée et faible, resté positif (+3 300 à +5 400 €) et persistant lorsque le gradient microvasculaire posé était supprimé (+5 318 €), ce qui montre qu'il est porté par le mécanisme rénal (MRC ≥ 3b, progression accélérée, dialyse) plutôt que par le gradient injecté ; et le biais à la hausse du schéma « réapparus seuls », présent dans chaque scénario (+930 € avec un excès de risque de décompensation ×5 à +6 660 € avec ×2). Le *signe* du coût

net de la déshérence n'a pas tenu : il devenait positif lorsque la fraction ambulatoire du coût de suivi était de 30 % (+402 ± 75 €), était indistinguable de zéro lorsque les coûts d'épisode étaient multipliés par 1,5 (−72 ± 76 €), et approchait zéro lorsque l'excès de risque de décompensation pendant la rupture était de ×5 (−722 ± 79 €). L'excès de mortalité n'était pas monotone en fonction du multiplicateur de décompensation (6,1 points à ×2, 5,7 à ×5) : plus de décompensations signifie plus de reconnexions hospitalières et donc des ruptures plus courtes, propriété du mécanisme de reconnexion posé.

3.3 Analyses de survie et coût-utilité

Émulation de cohorte (export v0.19, n = 20 000, graine 42). Le script fondé uniquement sur les données de remboursement a identifié 6 333 patients en déshérence et les a tous appariés au niveau 1. Toutes les DMS étaient ≤ 0,005 en valeur absolue, sauf le coût des 12 mois précédents (−0,076), dont le déséquilibre résiduel reste à l'intérieur des tertiles. La mortalité vraie à 10 ans de ces patients était de 42,2 % dans le statu quo et de 36,9 % s'ils étaient restés suivis (+5,3 points).

Chaque analyse de survie voit une partie différente de l'effet (Tableau 7).

Tableau 7 - Ce que chaque analyse de survie voit de l'effet de la déshérence sur la mortalité (graine 42, n = 20 000 ; les issues de survie sont identiques en v0.14 et v0.19, le modèle causal étant inchangé). Statut : analyse à graine unique, non répliquée ; son équivalent sur 10 graines pour l'intention de suivre est donné au §3.7.

Analyse (10 ans sauf mention contraire)	En déshérence	Suivis appariés	Ce que cela montre
Survie globale, intention de suivre (Kaplan–Meier)	60.4%	62.5%	| écart de 2,1 points contre 5,3 dans la vérité : contamination — 41 % des suivis entrent ensuite en déshérence, la plupart des personnes en déshérence reviennent
Sans hospitalisation à 5 ans (Kaplan–Meier)	30.0%	45.8%	l'écart s'ouvre au cours des années 1–3, puis les courbes deviennent parallèles
Incidence cumulée d'hospitalisation (Aalen–Johansen)	79.0%	69.4%	1 − KM surestimait l'incidence d'~8 points (86,8 % / 77,0 %) ; l'écart entre cohortes est préservé
| Survie globale, selon l'exposition effective (censure au changement d'état), 5 ans	77.7%	84.1%	l'écart de mortalité réapparaît, sous censure informative
Rapport de taux silence / suivi, décompensation	×2.17	×2.74	le silence double à triple le taux de décompensation tant qu'il dure
Rapport de taux silence / suivi, décès	×1.01	×1.15	le taux instantané de décès change à peine

Les deux dernières lignes réconcilient les autres : la mortalité due à la déshérence est **différée**. Les patients décompensent pendant le silence, réapparaissent à l'hôpital, sont reconnectés et meurent plus tard, comptés comme « suivis ». Toute analyse qui attribue le décès à l'état observé au moment du décès manquera l'essentiel de l'effet ; seul le contrefactuel apparié le mesure.

Coût-utilité (tronc mensuel, 20 ans, actualisation à 2,5 % ; Tableau 8, Figure 5).

Tableau 8 - Coût-utilité du suivi continu et du réengagement contre statu quo (tronc mensuel, 20 ans, actualisation à 2,5 %, perspective payeur ; moyenne ± MCSE sur 10 graines). Statut : résultat répliqué (10 graines).

Stratégie contre statu quo	Δ coût par personne	Δ QALY pour 1 000	RDCR sans coût fixe	Marge à 30 / 50 / 100 k€ par QALY
Suivi continu (borne théorique)	+689 ± 28 €	+175.5 ± 1.8	~3,900 €/QALY	—
Réengagement, file A (silence le plus long)	+89 ± 10 €	+21.3 ± 0.4	~4,200 €/QALY	49 / **88** / 184 €
Réengagement, file B (charge microvasculaire)	+131 ± 11 €	+19.9 ± 0.4	~6,600 €/QALY	42 / 78 / 168 €

Les marges sont exprimées par personne couverte atteinte de diabète de type 2 et par an (E4) ; la MCSE de la marge à 50 k€/QALY de la file A était de 2 €. À un horizon de 10 ans, cette marge tombait à 24 ± 1 € [19 ; 31], parce que les gains en QALY s'accumulent tardivement alors que les coûts de fonctionnement sont payés dès le premier jour. Sur les 13 scénarios de sensibilité, elle allait de 43 € (probabilité de retour 0,25, ou budget mensuel de 0,1 %) à 216 € (budget de 0,5 % par mois). Le réengagement était dominant en moyenne (moins coûteux et plus efficace ; ΔC −25 €) lorsque la fraction ambulatoire du coût de suivi était de 30 %. L'ajout des traitements (v0.13 → v0.14) a porté le RDCR sans coût fixe de ~2 800 à ~4 900 €/QALY, parce que les patients réengagés reprennent des médicaments coûteux, mais a laissé la marge presque inchangée (93 → 90 €).

Les deux files ont gagné les mêmes années de vie au niveau de la population, mais pas pour les mêmes patients : la file B donnait une mortalité à 10 ans plus faible dans la strate à charge élevée (44,9 % contre 45,3 % pour A et 46,1 % dans le statu quo ; différence −0,39 ± 0,04 point), la file A dans la strate à charge faible (32,1 % contre 32,9 % ; +0,82 ± 0,05 point). Avec des silences longs, la file B gagnait près de deux fois plus d'années de vie par patient en déshérence (0,085 contre 0,047 an en v0.13).

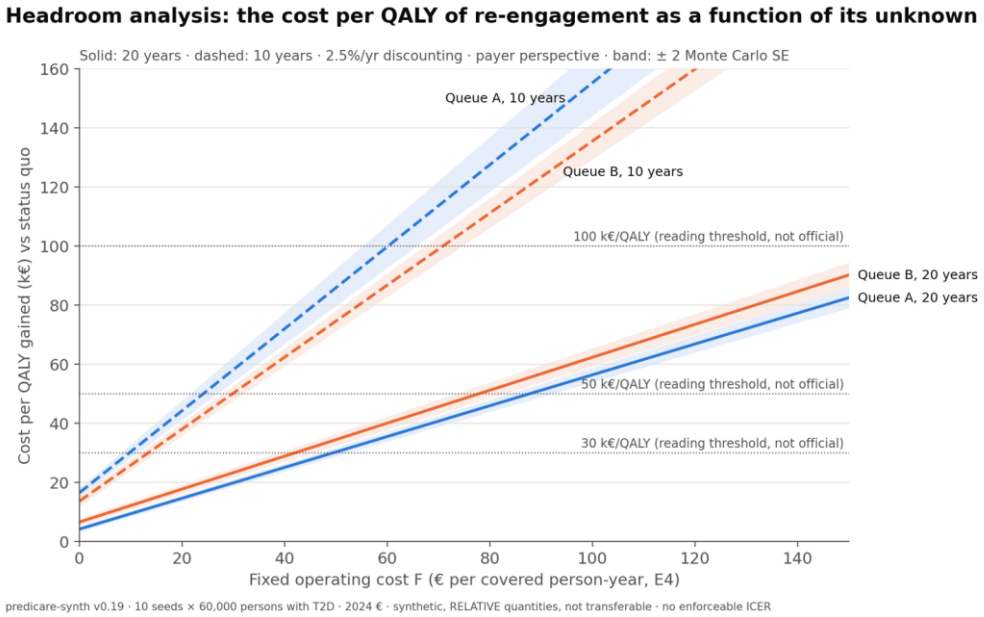

Figure 5 - Analyse de marge : RDCR du réengagement en fonction du coût fixe de fonctionnement inconnu (€ par personne-année couverte), horizons de 10 et 20 ans ; bande ± 2 MCSE sur 10 graines. Les seuils sont des aides à la lecture, non des valeurs officielles françaises.

3.4 Destins croisés, variante IoMT et sensibilité probabiliste

Variante IoMT après réengagement (PC, v0.19). Équiper les patients réengagés ajoutait +0,33 ± 0,14 QALY pour 1 000 au réengagement seul (+21,3 ± 0,4). Sa valeur venait surtout du traitement de la perte du signal d'un patient équipé hors du budget de réengagement (+1,57 ± 0,16 avec une file de télésurveillance dédiée), c'est-à-dire de la continuité plutôt que des décompensations évitées ; la courbe du registre de ce monde s'élevait à un rappel de 0,22 après vingt ans, pour un plafond de 0,24 (Tableau supplémentaire S11, Figure supplémentaire S3). Le forfait IoMT maximal de PC (9 € par personne-mois équipée, 32 € avec une file de télésurveillance dédiée) ne dépend pas du coût de kit supposé, et il se situe en dessous du coût d'un capteur de glucose en continu seul (environ 86 € par mois) : avec un kit qui en comprend un, équiper les personnes réengagées n'est efficient à 50 k€/QALY en moyenne dans aucun scénario du modèle.

Deux questions, deux populations. « Combien coûte la déshérence ? » est la question que pose le payeur lorsque le patient disparaît, et la population pertinente est celle de tous les patients silencieux. « Combien coûtent les patients qui reviennent abîmés ? » concerne le sous-groupe qui réapparaît à l'hôpital ; le bon comparateur est constitué des mêmes patients s'ils étaient restés suivis, et non d'un groupe de suivis ordinaires. Le banc d'essai calcule les deux (fenêtre de 10 ans, SQ − REF, non actualisé ; Tableau 9). Les lignes définies par une issue observée dans SQ sont des strates principales définies par le modèle (E2) ; seules les lignes définies par la strate observable dans les données de remboursement décrivent des sous-groupes qu'une étude réelle pourrait constituer.

Tableau 9 - Coût de la déshérence par population, perspective payeur (10 ans, SQ − REF, non actualisé ; moyenne ± MCSE sur 10 graines ; 60 000 personnes par graine ; perspectives collectives dans le Tableau supplémentaire S12). Statut : résultat répliqué (10 graines).

Population	Part des silencieux	Coût, payeur	Décès à 10 ans, SQ contre REF
Tous les silencieux	100%	−1,981 ± 76 €	41.2% vs 35.3%
Réapparus à l'hôpital	51%	+835 ± 134 € [350; 1,683]	40.7% vs 35.2%
Réapparus à l'hôpital, charge latente élevée	28%	+2,796 ± 235 € [2,114; 4,295]	45.3% vs 39.9%
Charge latente élevée, toutes issues	49%	−241 ± 160 € [−933; 782]	46.1% vs 40.4%

Le schéma usuel (patients réapparus contre suivis ordinaires) rapporte +1 106 € (Tableau 6) : il surestime le coût de la déshérence pour ce sous-groupe d'environ un tiers (+1 106 € contre +835 €), avec une incertitude large, mêlant le prix de la déshérence au prix d'avoir mal évolué. Les sous-groupes définis par une issue dans le monde statu quo sont des strates principales [32] : leur comparaison avec REF est exacte sous le modèle et sous le couplage choisi des tirages aléatoires, mais non identifiable dans des données réelles. En pratique, l'avenir d'un patient silencieux est inconnu au moment où le silence est détecté ; ce qui peut être observé est la strate microvasculaire lue dans les données de remboursement, selon laquelle la probabilité de réapparition hospitalière passait de 45 % à 61 % et 69 %, et les années de vie perdues de 0,39 à 0,46 et 0,50, alors qu'aucune strate observable ne présentait de surcoût (−2 280 € à −1 320 €), et que seuls 14 % des patients réellement à charge élevée étaient identifiés.

Perspective collective. L'ajout de la dépendance à domicile ou en établissement et de l'aide informelle (cas de référence de la HAS [25]) creusait l'économie apparente pour l'ensemble des silencieux,

multipliait le surcoût de ceux qui reviennent abîmés, et abaissait les marges d'environ 5–13 % ; le signe du coût de la déshérence était négatif dans les trois perspectives (Tableau supplémentaire S12).
Analyse de sensibilité probabiliste. Sur 300 tirages conjoints de 23 distributions de paramètres déclarées (mondes REF, SQ, PA et PC ; 20 000 personnes par tirage), la déshérence représentait une économie pour le payeur sur 10 ans dans 95 % des tirages (par personne diabétique, c'est-à-dire en moyenne sur toute la population simulée, dont environ un tiers connaît au moins une déshérence : médiane −613 € [−1 388 ; +112], actualisé) tout en augmentant la mortalité à 10 ans dans chaque tirage (niveau population : médiane +1,6 point [0,9 ; 2,5]). La marge du réengagement à 50 k€/QALY avait une médiane de 83 € [33 ; 193] sur 20 ans et de 23 € [2 ; 70] sur 10 ans ; le bruit de Monte Carlo représentait 7 % de sa variance. Pour un coût fixe de 30 € par personne et par an, le réengagement était en dessous de 50 k€/QALY dans 98 % des tirages ; à 50 €, 86 % ; à 100 €, 29 %. La probabilité de retour après contact dominait (ρ de Spearman = 0,65), suivie de la joignabilité (0,36), de l'excès de mortalité pendant la rupture (0,31) et de l'accélération rénale (0,26). L'ajout IoMT (schéma de base, sans file dédiée) n'était positif que dans 59 % des tirages (médiane +0,28 QALY pour 1 000 [−2,1 ; +2,5]). Ces distributions sont des plages de plausibilité déclarées, non une incertitude estimée à partir de données (Figures supplémentaires S4 et S5).

3.5 La séquence PREDICARE (monde PE)

Sur 20 ans (v0.19, 10 graines × 60 000), le budget d'invitation a produit en moyenne 18 840 invitations adressées à 5 760 personnes, 3 680 personnes venues au moins une fois, 3 170 personnes équipées au moins une fois et 1 690 reprises du suivi conventionnel ; la sensibilité prescrite pour l'insuffisance cardiaque passait de 0,06 à un an à 0,18 à cinq ans, plus vite que dans PC parce que les patients en déshérence équipés décompensent plus souvent et alimentent donc la courbe plus rapidement. Le scénario dans lequel la télésurveillance n'a aucun effet sur l'excès de risque de déshérence (atténuation 0) est la référence : l'atténuation n'est pas mesurée, et toute valeur positive est une hypothèse qu'aucune donnée actuelle ne soutient. Les décomptes ci-dessus concernent cette référence (le §3.6 utilise une atténuation de moitié, d'où 3 110 équipés) ; une personne peut venir lors de plusieurs épisodes de déshérence, de sorte que la part des participants équipés au moins une fois (86 %) dépasse l'acceptation par venue de 0,80. Les marges par scénario sont données dans le Tableau 10 et la Figure 6 (gauche).

Tableau 10 - Séquence PREDICARE (PE − SQ) par scénario : gain en QALY et marge (20 ans, actualisé ; moyenne ± MCSE sur 10 graines ; télésurveillance à 110 € par mois équipé, autres coûts dans le Tableau supplémentaire S14). Les marges sont une illustration conditionnelle sous des paramètres posés et non calibrés, non des estimations de coûts ni des budgets. Statut : résultat répliqué (10 graines), conditionnel aux paramètres posés du programme.

Scénario PE (20 ans, actualisé)	Δ QALY pour 1 000, PE − SQ	Marge à 50 k€/QALY (€/personne-année couverte)
Référence : télésurveillance sans effet sur l'excès de risque (atténuation 0)	+7.0 ± 0.3	13 ± 1
Atténuation 0.5	+9.4 ± 0.4	28 ± 2
Atténuation 0.5, sans sélection sur la participation	+10.8 ± 0.3	33 ± 2
Attenuation 0.5, Attrition des capteurs de 1% par mois	+11.3 ± 0.4	34 ± 2
Attenuation 1 + traitements renouvelés par télémédecine	+12.5 ± 0.5	47 ± 2
Idem, attrition de 1% par mois	+17.1 ± 0.5	70 ± 2
Comparateur: re-engagement seul (PA, 50% de retour après contact)	+21.3 ± 0.4	88 ± 2

La valeur de la séquence PREDICARE dépend de deux grandeurs que personne n'a mesurées, dans quelle mesure la télésurveillance compense l'excès de risque de déshérence, et si le traitement se poursuit en dehors du suivi conventionnel. PE et PA ne sont pas comparés à capacité égale : PA pose qu'un contact réengage 50 % des personnes jointes, alors que PE enchaîne joignabilité, participation et une reprise de 35 % après bilan, avec des budgets, des taux d'accès et des intensités de suivi différents. La comparaison est une illustration de scénarios, non un classement de stratégies ; harmoniser les capacités, et mesurer ces taux, est précisément ce qu'un démonstrateur doit faire.

Sensibilité probabiliste de la séquence PREDICARE. Les scénarios ci-dessus étant choisis, l'analyse probabiliste a été répétée sur les mondes REF, SQ, PA et PE (300 tirages, 20 000 patients chacun), en ajoutant aux 23 précédentes sept distributions propres à la séquence : participation si joint (bêta, moyenne 0,50), reprise du suivi après bilan (bêta, moyenne 0,35), attrition des capteurs pendant la déshérence (lognormale, médiane 4 % par mois), reconnexion après une vraie alerte (bêta, moyenne 0,50), atténuation de l'excès de risque par la télésurveillance (uniforme 0–1), renouvellement des traitements par l'équipe de télésurveillance (Bernoulli 0,5) et coût du bilan à T0 (uniforme 120–250 €) ; le coût de télésurveillance est tiré uniformément entre 70 et 150 € (chaque tirage publié étant transposé au même quantile, coûts corrigés exactement ; Tableau supplémentaire S14). Sur 20 ans, PE gagnait en médiane +8,8 QALY pour 1 000 [2,9 ; 19,6], et sa marge à 50 k€/QALY avait une médiane de 25 € par personne-année couverte [−1 ; 78] (10 ans : 1 € [−12 ; 26]) ; le bruit de Monte Carlo représentait environ 12 % de sa variance (variances intra-tirage calculées sous la distribution de coût précédente, non réestimées à 110 €). Pour un coût fixe de 10, 20, 30 et 50 €, PE restait en dessous de 50 k€/QALY dans 84 %, 64 %, 39 % et 15 % des tirages. La marge dépendait principalement de l'atténuation de l'excès de risque (ρ de Spearman = 0,58), de la participation (0,39) et de la reprise du suivi (0,32), puis de l'excès de mortalité pendant la rupture (0,27), de l'accélération rénale pendant la rupture (0,20) et du retour spontané (−0,20). Ces résultats quantifient la dépendance de PE à des grandeurs non mesurées ; ils ne rendent pas sa valeur mieux établie que ne le suggèrent les scénarios, et ils identifient ce qui doit être mesuré plutôt qu'un budget cible. Unités : la marge est exprimée par personne atteinte de diabète de type 2 dans le territoire et par an (E4 ; personnes-années actualisées de toute la population simulée, environ 11 par personne sur 20 ans), et non par patient ciblé, invité ou équipé. Les traductions du même total par personne invitée, par participant ou par patient-année équipé sont données dans le Tableau supplémentaire S10 ; elles redistribuent une grandeur conditionnelle sur un dénominateur plus petit et ne sont pas des budgets.

Valeur de l'information. À 50 k€/QALY et pour un coût fixe de 30 € par personne-année couverte, niveau auquel la séquence était efficiente dans 39 % des tirages, l'EVPI était de 86 ± 7 € par personne diabétique sur 20 ans. L'EVPPI était de 52 ± 8 € pour l'atténuation de l'excès de risque sous télésurveillance, 33 ± 8 € pour la participation, 26 ± 7 € pour l'excès de mortalité pendant la rupture, 18 ± 6 € pour la reprise du suivi, et 63 ± 7 € pour l'atténuation, la participation et la reprise conjointement ; la joignabilité (10 ± 6 €), l'accélération rénale et le retour spontané pendant la rupture (14 ± 7 € et 14 ± 6 €) et l'attrition des capteurs (11–12 ± 6–7 €) se situent à environ deux erreurs-types et ne sont pas interprétés, pas plus qu'aucun autre paramètre. À 10 €, la décision n'était plus quasi certaine (efficiente dans 84 % des tirages ; EVPI 13 ± 2 €). Avec la distribution précédente du coût de télésurveillance (10–40 €), l'EVPI était de 41 € et l'EVPPI de l'atténuation de 13 € (Tableau supplémentaire S14) : un kit réaliste rapproche la décision de l'indifférence et double la valeur de l'information. Ce classement plaide pour recueillir d'abord des données sur l'atténuation (par des critères intermédiaires sous télésurveillance, avec les hypothèses d'identification que cela requiert) et sur la participation ; les valeurs sont par personne sur 20 ans, et une EVPI populationnelle exigerait la taille et la durée de la décision, que nous ne posons pas.

La sélection, et la population dans laquelle l'offre est évaluée. Les résultats portent sur la mortalité à 10 ans dans le scénario de référence sans atténuation (moyenne ± MCSE sur 10 graines ; Tableau supplémentaire S10). Parmi toutes les personnes éligibles (E5 ; environ 18 100 par graine, la population

de E1), l'offre réduisait la mortalité à 10 ans de 0,31 ± 0,02 point et augmentait le coût payeur à 10 ans de 764 ± 14 € par personne éligible (337 € avec la valeur par défaut du code de 25 € par mois équipé). Sous le budget mensuel d'invitation, seuls 28,0 % des éligibles ont jamais été invités, après un délai médian de 41 mois, et 18,3 % sont venus ; les éligibles jamais invités avaient, par construction, des issues identiques dans les deux mondes. L'effet causal moyen chez les observants [18] (ITT divisé par la participation, puisque personne ne vient dans le statu quo) retrouvait la vérité parmi les participants, −1,72 ± 0,09 point. Parmi les personnes effectivement invitées (temps zéro : première invitation ; population sélectionnée par la file et le budget), l'effet de l'offre était de −1,25 ± 0,09 point, et 67,1 % sont venues après des invitations répétées. La comparaison naïve des participants aux non-participants parmi les invités donnait −11,1 ± 0,5 points, environ six fois la vérité parmi les participants avec le même temps zéro (−1,87 ± 0,13 point) et neuf fois l'effet parmi les invités (Figure 6, droite), parce que les non-participants étaient plus malades (63 % contre 47 % à charge latente élevée ; 46 % contre 37 % de mortalité à 10 ans dans le statu quo). Avec une atténuation de moitié, l'effet de l'offre était de −0,41 ± 0,02 point parmi tous les éligibles et de −2,28 ± 0,12 points parmi les participants. Le CACE ne retrouve ici la vérité que parce que, par construction, l'offre n'a aucun effet sur les non-participants (restriction d'exclusion) — hypothèse à argumenter sur le terrain, où une invitation peut modifier le recours aux soins même sans venue (effet de rappel) ; le démonstrateur devrait donc mesurer le recours aux soins chez les non-participants. L'effet de l'offre parmi tous les éligibles est par conséquent l'estimand principal : dans la simulation, il correspond au schéma (i) ci-dessous, une évaluation randomisant des territoires avant le début du ciblage avec la même règle d'éligibilité dans les deux bras et une capacité fixée localement. Les invitations répétées font partie de la politique évaluée. Comme le budget dilue l'offre, la mortalité à 10 ans n'est pas un critère principal réaliste. Parmi environ 18 100 éligibles pour 60 000 personnes atteintes de diabète de type 2, une mortalité à 10 ans de 42,5 % dans le statu quo correspond à environ 7 700 décès, dont la politique simulée en évite environ 57. Détecter une différence de 0,31 point avec une puissance de 80 % au seuil bilatéral de 5 % exigerait environ 400 000 éligibles par bras sous randomisation individuelle (approximation normale pour deux proportions), avant toute inflation par l'effet de schéma d'un essai en grappes ; restreindre l'analyse aux personnes effectivement invitées (−1,25 point) en exigerait environ 24 000 par bras, mais dans une population sélectionnée par la file. Ces ordres de grandeur sont conditionnels à des paramètres posés et non calibrés. Ils indiquent qu'un démonstrateur aurait besoin de critères intermédiaires (reprise du suivi, continuité du traitement, décompensations), dont le banc d'essai pose plutôt qu'il n'estime les tailles d'effet, et qu'il devrait rapporter les fractions d'invités et de participants aux côtés de l'ITT. La recommandation est une orientation pour un protocole, non un protocole : l'unité de randomisation, le nombre de grappes, les critères et la taille d'échantillon restent à concevoir et à justifier avec des données réelles. L'effet parmi les participants requiert la restriction d'exclusion (la monotonie est vérifiée par construction ici, puisque la venue est impossible sans invitation) et est secondaire. Le démonstrateur devrait préspécifier l'unité de randomisation, la règle d'éligibilité et le temps zéro, le traitement des invitations répétées, et, pour les schémas en grappes, les grappes, les périodes, la contamination entre bras et les tendances séculaires [33].

Ce qu'un essai de l'offre estimerait sous une capacité partagée. L'offre est délivrée par une file mensuelle à budget fixe, de sorte qu'une invitation envoyée à une personne est une place qui n'est pas envoyée à une autre : les affectations et les issues ne sont pas indépendantes d'une personne à l'autre (interférence). Trois schémas sont possibles, et ils n'estiment pas le même effet. (i) *Randomisation par territoire* (ou par grappe de soins primaires), allouée avant le début du ciblage, avec la même règle d'éligibilité et le même temps zéro dans les deux bras et une capacité de chaque bras fixée localement : ce schéma estime l'effet de la politique dans son ensemble, budget, ordre de file et sélection compris, parmi toutes les personnes éligibles. C'est ce que représente E5 dans la simulation, et le −0,31 point est l'effet de cette politique simulée sous son budget ; ce n'est pas l'effet d'une invitation reçue par chaque personne éligible. (ii) *Randomisation individuelle des places d'invitation* parmi les personnes éligibles à chaque période (une loterie au sein de la file) : ce schéma estime l'effet de se voir proposer une place, parmi ceux qui sont en concurrence pour les places à cette période, conditionnellement à la capacité ;

les personnes non tirées peuvent l'être plus tard, de sorte que le traitement de la réentrée dans la loterie et des invitations répétées doit être préspécifié, et le contraste se dilue à mesure que les témoins sont invités ultérieurement. (iii) *Comparaison de règles d'allocation sous le même budget* (par exemple les deux ordres de file du §3.3, randomisés par grappe) : ce schéma estime la différence entre règles, non l'effet du programme. Quel que soit le schéma, le protocole doit fixer avant l'affectation qui entre dans le dénominateur, à quel temps zéro, comment les réinvitations sont comptées, et comment la capacité est partagée entre les bras.

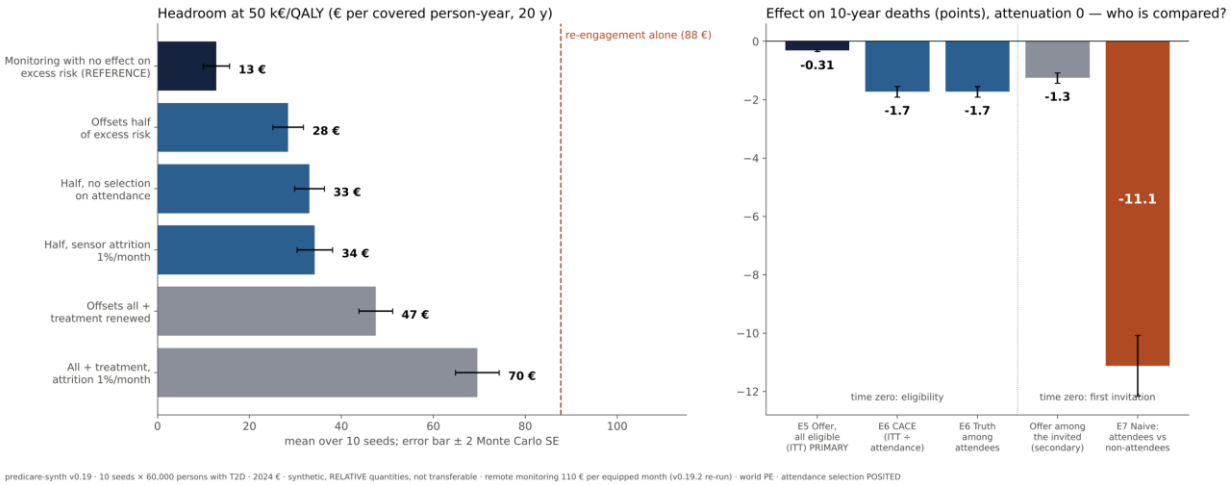

Figure 6 - Monde PE. Gauche : marge à 50 k€/QALY sous les deux grandeurs non mesurées, télésurveillance à 110 € par mois équipé (barres : moyenne ± 2 MCSE sur 10 graines ; barre foncée : référence, atténuation 0). Droite, scénario de référence : l'effet de l'offre parmi toutes les personnes éligibles (E5, principal), sa mise à l'échelle par la participation (CACE, E6) et la vérité parmi les participants (E6), tous avec un temps zéro à l'éligibilité ; puis, avec un temps zéro à la première invitation, l'effet parmi les invités et le contraste naïf participants contre non-participants (E7).

3.6 Un bilan à T0 qui décide, et la capacité clinique derrière les alertes (mécanismes v0.16, exécutions v0.19)

Deux mécanismes ont été ajoutés à PE (atténuation de l'excès de risque 0,5, afin que la télésurveillance ait un effet sur lequel les alertes puissent agir ; 10 graines × 60 000). D'abord, le bilan à T0 mesure désormais la MRC, l'insuffisance cardiaque et une strate microvasculaire élevée avec des erreurs posées et peut piloter l'équipement : sous une politique stratifiée, seuls les participants classés à haut risque sont équipés. Ensuite, les alertes sont traitées par une équipe de capacité limitée (alertes examinées pour 100 patients équipés par mois, face à un budget d'alertes de 5), éventuellement priorisées selon le risque à T0, et l'évitabilité décroît linéairement avec le délai de réponse jusqu'au délai d'anticipation médian de 13 jours du tronc quotidien (Tableau 11).

Tableau 11 - Politique d'équipement à T0, erreurs du bilan, capacité d'alerte et délai de réponse dans le monde PE (atténuation 0,5 ; télésurveillance à 110 € par mois équipé ; moyenne ± MCSE sur 10 graines ; les personnes équipées et les décompensations évitées sont des totaux pour 60 000 personnes sur 20 ans). Statut : résultat répliqué (10 graines), conditionnel aux paramètres posés du programme.

Scénario (PE − SQ, 20 years)	Δ QALY pour 1,000	Marge à 50 k€/QALY	Patients équipés	Décompensations évitées par les alertes
Référence (tous les participants équipés, capacité illimitée, délai de 2 jours)	+9.4 ± 0.4	28 ± 2 €	3,110	97
Équipement stratifié à T0, bilan parfait	+8.4 ± 0.3	29 ± 1 €	2,085	69
Stratifié à T0, erreurs du bilan (sensibilité 0,80 / 0,70 / 0,60, 5 % de faux positifs)	+8.1 ± 0.3	28 ± 1 €	1,861	62
Capacité de 3 alertes pour 100 équipés par mois	+9.2 ± 0.4	28 ± 2 €	3,106	59
Capacité 1.5	+9.3 ± 0.4	28 ± 1 €	3,111	28
Capacité 1.5, , alertes priorisées selon le risque à T0	+9.3 ± 0.4	28 ± 2 €	3,108	35
Délai de réponse de 7 jours	+9.4 ± 0.4	28 ± 2 €	3,111	57
Délai de 7 jours et capacité 1.5	+9.3 ± 0.4	28 ± 1 €	3,108	16

La capacité et le délai supprimaient jusqu'à 84 % des évitements dus aux alertes (97 → 16), et la priorisation selon le risque à T0 récupérait environ un cinquième des évitements perdus lorsque la capacité passait de 3 à 1,5 alerte pour 100 patients équipés par mois (59 → 28 → 35), soit 10 % de la perte par rapport à une capacité illimitée ; pourtant, les QALY et la marge bougeaient à peine : toutes les différences restaient dans environ deux MCSE. Dans ce modèle, la valeur de la séquence ne passe donc pas par les alertes (elles évitent une centaine de décompensations parmi 60 000 personnes sur 20 ans) mais par l'atténuation posée de l'excès de risque pendant la déshérence télésurveillée et par la reconnexion aux soins. C'est en partie structurel, puisque l'atténuation s'applique à chaque patient télésurveillé indépendamment des alertes, et cela fait de ce paramètre, et non du détecteur, la grandeur sur laquelle un démonstrateur devrait recueillir en premier des données intermédiaires ; une offre randomisée n'attribuerait pas, à elle seule, son effet à l'atténuation. Une politique d'équipement stratifiée à T0 équipait un tiers de personnes en moins (2 085 contre 3 110) mais perdait environ un dixième du gain en QALY (8,4 contre 9,4 pour 1 000), parce que les participants à moindre risque bénéficiaient eux aussi de la télésurveillance sous l'atténuation posée ; les erreurs du bilan coûtaient peu de plus. La marge de la stratification dépend du coût d'équipement : avec la valeur par défaut du code de 25 € par mois, elle était inférieure d'environ 4 € à celle de l'équipement de tous les participants (36 contre 40 €) ; à 110 €, les deux politiques sont équivalentes (29 contre 28 €), et à 150 €, la stratification est préférable (25 contre 23 €), parce qu'équiper les participants à moindre risque coûte alors davantage que les QALY qu'il rapporte. La comparaison se déplacerait aussi si la télésurveillance des patients à faible risque n'apportait aucune atténuation.

Sensibilité au plafond du registre. Diviser par deux ou par quatre le plafond pour l'insuffisance cardiaque (0,12 et 0,06 au lieu de 0,24 ; 10 graines chacun) laissait la marge de PE inchangée à 0,2 € près (24,6 → 24,5 € sans atténuation ; 40,1 → 40,0 € avec une atténuation de moitié ; valeurs avec la valeur par défaut du code de 25 € par mois équipé, non rejouées à 110 €) et le gain en QALY inchangé à la première

décimale, tandis que les décompensations évitées par les alertes baissaient de 10–15 % (128 → 109 sans atténuation ; 97 → 83 avec) ; dans PC, la borne du forfait IoMT passait de 9 à 8 € par personne-mois équipée (Tableau supplémentaire S9-C). L'évitement ne baisse pas proportionnellement au plafond parce que les alertes d'insuffisance rénale aiguë et de décompensation diabétique conservent leurs propres plafonds et parce que le plancher de rappel est le hasard au quota (0,05). Dans cette plage, les résultats économiques sont insensibles à la manière dont le rappel du tronc quotidien est transféré à la file mensuelle ; le nombre d'évitements dus aux alertes ne l'est pas, et il faudrait simuler la chaîne d'alerte elle-même (alertes répétées, dédoublonnage, éventail de cas) pour l'interpréter.

3.7 Les schémas fondés sur les données de remboursement sous des mécanismes générateurs alternatifs

Les résultats proviennent de la v0.18 (14 mécanismes × 10 graines × 60 000 personnes ; Figure 7, Tableau supplémentaire S7).

La sélection sur la réapparition biaisait le coût à la hausse sous les 14 mécanismes. Le biais moyen du schéma « réapparus seuls » allait d'environ +2 600 € (M4) à +7 500 € (M3) par personne sur 10 ans. Il était positif dans 139 des 140 couples mécanisme–graine, et la couverture de son intervalle à 95 % était au plus de 40 %. Que l'estimation inverse aussi le signe dépendait du mécanisme : elle l'inversait, en moyenne, lorsque la déshérence était causale et que les patients se reconnectaient après l'hôpital (9 des 13 mécanismes à vérité clairement négative ; dans M4, M7 et M8b, l'estimation se situe à environ 1,5 MCSE de zéro) ; elle ne l'inversait pas lorsque la déshérence était principalement un marqueur de gravité latente (M1–M3) ou lorsque la reconnexion était absente (M8) ; et sous déshérence partielle et intermittente (M6), la vérité est proche de zéro (−70 ± 60 €) alors que le schéma rapporte +4 300 €. Le résultat robuste est le sens du biais, non l'inversion.

Comparer tous les silencieux à des suivis appariés surestimait l'économie sous chaque mécanisme, d'environ 1 300 € (M8) à 3 000 € (M4) en moyenne, et dans 139 des 140 couples mécanisme–graine. L'appariement sur les covariables des données de remboursement ne supprimait ni la perte de soins payés ni les différences latentes.

Le sens du biais de mortalité en intention de suivre dépendait du mécanisme. Lorsque la déshérence était causale (M0, M5–M11), l'intention de suivre diluait l'effet, d'environ 1,7 (M6) à 4,2 (M8) points (sous le mécanisme de référence M0, 3,3 ± 0,1 points contre une vérité de 5,8 ± 0,1 sur 10 graines ; 2,1 contre 5,3 dans l'émulation de cohorte à graine unique du Tableau 7), par contamination et retour aux soins, comme au §3.3 ; lorsqu'un effet causal était combiné à une confusion par une fragilité partagée (M4), elle surestimait l'effet de 1,5 point. Lorsque la déshérence était en partie ou en totalité un marqueur de gravité latente, le même schéma la surestimait : 8,7 points contre 3,4 sous M1, et 6,5 points d'excès de mortalité à 10 ans sous M2 et M3, où l'effet vrai était de 0,1 point. Une étude sur données de remboursement ne peut donc pas savoir, à partir de ses propres données, si son contraste de mortalité est trop faible ou trop fort.

Le rapport des taux de décès selon l'exposition effective ne retrouvait jamais l'effet à 10 ans. Il allait de 1,06 (M10) à 1,57 (M8) sous les mécanismes causaux et valait 0,98–0,99 sous les mécanismes non causaux : il mesure le risque concomitant pendant le silence, non la mortalité différée qui le suit.

L'argument économique est conditionnel à la fraction causale. La marge du réengagement seul (PA, 50 k€/QALY, 20 ans) était de 88 € par personne-année couverte sous M0 ; elle tombait à 28 € lorsque la moitié de l'excès de risque était causale (M1) et à 27 € lorsque 30 % des silences correspondaient à des soins hors système (M7) ; elle devenait négative (jusqu'à −52 €) lorsque la déshérence était un pur

marqueur (M2, M3), parce que le réengagement ramène alors des soins coûteux sans gain de santé ; et elle montait à 173 € lorsque la reconnexion après l'hôpital était absente (M8). La grandeur qui décide si un programme de vigilance peut valoir son coût est donc la part causale de l'excès de risque de déshérence. Le banc d'essai ne peut pas la fournir, et une évaluation randomisée de l'offre à elle seule non plus. Un tel essai identifie l'effet de l'offre — ciblage, invitation, bilan, télésurveillance et toute reprise des soins pris ensemble — dans la population randomisée, quelle que soit la confusion entre déshérence et gravité ; il n'identifie pas la part de l'excès de risque causée par la déshérence elle-même, ni ne sépare les contributions de la reprise du traitement, du bilan à T0, de la télésurveillance et de la réponse aux alertes. Attribuer l'effet d'un essai à l'un de ces mécanismes requiert des mesures supplémentaires et des hypothèses d'identification (critères intermédiaires tels que la continuité du traitement et les décompensations sous télésurveillance, schémas multibras ou factoriels, ou analyses de médiation avec leurs propres conditions).

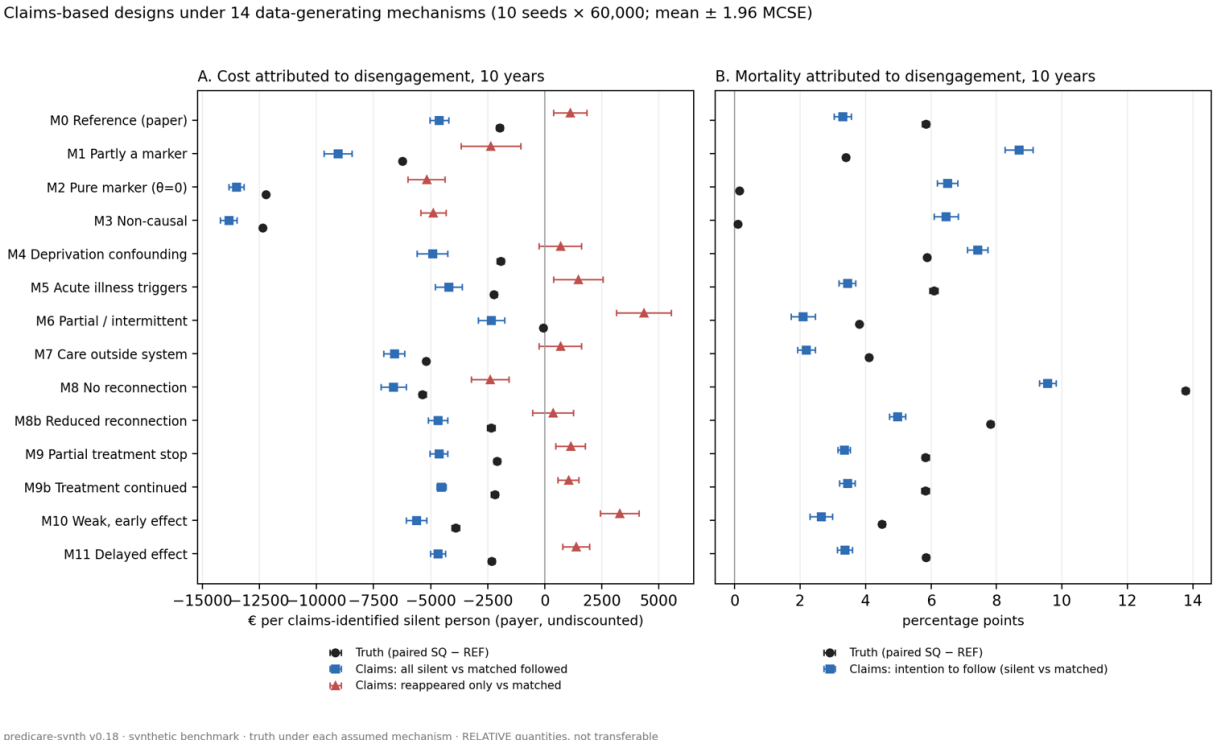

Figure 7 - Schémas fondés sur les données de remboursement sous 14 mécanismes générateurs de données : vérité appariée (SQ − REF sous chaque mécanisme) et estimateurs sur données de remboursement, moyenne ± 1,96 MCSE sur 10 graines. A : coût attribué à la déshérence à 10 ans ; B : mortalité à 10 ans (intention de suivre). Chaque mécanisme a sa propre vérité ; seule la relation entre estimateur et vérité est interprétée.

4. Discussion

Principaux résultats. Dans un banc d'essai synthétique où la vérité est connue par construction, le mécanisme de déshérence posé produisait environ 6 points d'excès de mortalité à 10 ans tout en étant à peu près neutre, voire économe, pour le payeur. Ces ordres de grandeur ne font que reformuler les risques instantanés injectés ; les résultats qui ne dépendent pas de leurs valeurs exactes sont les mécanismes et les biais de schéma qui en découlent. Dans le modèle, l'économie n'est pas un paradoxe : le silence supprime des soins utiles qui sont aussi des soins payés, et son préjudice se paie plus tard en vies et, pour les patients les plus abîmés, en dialyse. Dans les configurations testées, les trois schémas fondés sur les données de remboursement testés échouaient chacun d'une manière caractéristique : la sélection sur la réapparition biaisait le coût à la hausse, au point d'en inverser le signe en moyenne, la

survie en intention de suivre diluait l'effet sur la mortalité par contamination, et l'attribution selon l'exposition effective manquait une mortalité différée au-delà du silence lui-même. Dans la séquence PREDICARE, la courbe prescrite du registre (un gain de performance imposé, non un processus d'apprentissage simulé) montait rapidement parce que les patients en déshérence télésurveillés décompensent souvent, mais la valeur clinique dépendait de deux grandeurs non mesurées — dans quelle mesure la télésurveillance compense l'excès de risque de déshérence, et si le traitement se poursuit en dehors du suivi conventionnel — et la participation sélective posée suffisait à rendre trompeuse toute comparaison fondée sur les participants. Limiter la capacité de l'équipe qui traite les alertes, ou allonger son délai de réponse, supprimait l'essentiel des évitements dus aux alertes sans modifier les QALY : dans le modèle, la valeur passait par l'atténuation de l'excès de risque, non par le détecteur.

Propriétés des schémas, propriétés du mécanisme. L'expérience structurelle sépare deux types de résultats. Le biais à la hausse des cohortes fondées sur la réapparition et l'économie surestimée des comparaisons appariées sur données de remboursement persistaient sous les 14 mécanismes testés ; ce sont des propriétés de ces schémas au sein de cette classe de scénarios. Cette classe est explicite et définie par l'auteur : ce n'est ni un échantillon représentatif des mécanismes plausibles ni une distribution de probabilité sur ceux-ci, et elle ne montre pas que le sens de ces biais vaut pour les études sur le SNDS en général. Parmi les mécanismes non testés figurent une probabilité de reconnexion ou de retour dépendant de la fragilité latente, une confusion variant dans le temps entre précarité, observation et issues, des causes hétérogènes de rupture aux effets différents, et des différences territoriales de pratiques de codage et d'enregistrement. La dilution de la mortalité par l'analyse en intention de suivre, l'ampleur de chaque contraste et la marge économique ne persistaient pas ; ce sont des propriétés du mécanisme supposé, et avant tout de la part causale de l'excès de risque de déshérence.

Résultats algorithmiques.

1. **Dans le monde simulé, l'information nécessaire à la prédiction à 30 jours est absente des données de remboursement**. C'est une propriété de l'architecture d'observation, non du modèle : le même algorithme de gradient boosting sur les mêmes patients passait d'un niveau proche du hasard (AUC-ROC 0,57) à 0,70 lorsque les signaux quotidiens étaient ajoutés dans l'expérience interne — un gain de discrimination sur des prodromes simulés, non une validation prédictive : les ensembles de test poolés comptaient 176 épisodes, la valeur prédictive positive à un quota de 5 % était de 5,3 %, et la calibration était médiocre.

2. **Campagnes historiques (Tableaux supplémentaires S9-A et S9-D).** Dans les prodromes simulés, le capteur utile dépendait de l'organe, et la biologie de routine échouait en raison de sa cadence, non de l'absence de signal ; sous données manquantes informatives, le prédicteur migrait vers un proxy comportemental tandis que l'AUC augmentait ; et un gradient social dans l'équipement divisait par deux la couverture des patients en déshérence sans modifier l'AUC parmi les équipés. L'AUC ne certifie pas un prédicteur, ni son équité : il faut un contrôle de gouvernance sur l'attribution des variables et sur la couverture, non un meilleur score.

3. **Les règles de priorité sont bornées par l'observabilité.** Une file ordonnée par la charge microvasculaire déplaçait le bénéfice vers les patients à haut risque mais ne pouvait surpasser globalement une règle plus simple, parce que les données de remboursement simulées, par construction, n'identifient que 14 % des patients à charge élevée.

4. **Les estimateurs fondés sur les participants ne valent que ce que vaut la sélection que l'on suppose.** Lorsque le retour aux soins était indépendant du risque, le contraste naïf et son ajustement IPTW étaient sans biais ; lorsque la participation dépendait de la charge latente, le contraste naïf surestimait l'effet d'un facteur six environ. Les comparaisons avant/après échouaient, un essai en escalier surestimait d'un

facteur deux environ par confusion calendaire, et un ECR individuel était sans biais mais avait le mauvais signe dans 3 graines sur 10 pour des événements rares — ce qui plaide pour une évaluation randomisée préenregistrée de l'offre, avec des variantes en grappes ou en escalier uniquement sous séquences randomisées et modélisation explicite du temps calendaire, et pour des critères intermédiaires plus fréquents.

Rapport aux travaux antérieurs. Les données de santé synthétiques sont habituellement produites soit par apprentissage à partir de dossiers (réseaux génératifs, modèles de langage), ce qui reproduit leur régime d'observation, soit par des simulateurs à base de règles tels que Synthea [34], qui visent le réalisme des dossiers individuels. Notre générateur appartient à la seconde famille mais sert une autre fin : c'est un *banc d'essai pour des schémas*, plus proche de la simulation plasmode [35] et des études de simulation de méthodes statistiques [3], avec des nombres aléatoires communs pour réduire la variance entre mondes contrefactuels. Le sens de nos effets injectés suit les données publiées (charge microvasculaire et risque cardiovasculaire ; méta-analyses de CVOT pour les AR GLP-1 et les iSGLT2 ; prodromes de télésurveillance dans l'insuffisance cardiaque), mais le banc d'essai ne teste pas ces effets ; il teste ce qui en découle. La version synthétique « top diabète » du Health Data Hub reproduit le format du SNDS, non des patients, et ne porte aucune rupture longitudinale réaliste ; un modèle génératif profond entraîné sur elle apprendrait la distribution d'un autre générateur. Les générateurs appris deviennent pertinents une fois les données réelles du SNDS accessibles dans leur cadre légal (jalon G4) : pour calibrer et vérifier le générateur mécaniste, ou dans un schéma hybride qui conserve le tronc causal pour la vérité terrain et n'apprend que la tête d'observation des données de remboursement.

Limites.

- Circularité. Le banc d'essai prédit ce qu'il a injecté. Tous les résultats sont relatifs (entre bras, schémas ou mondes à vérité identique), jamais des performances, incidences ou coûts absolus. Cela est inscrit dans des domaines de validité lisibles par machine. Les tests automatisés vérifient des invariants internes (séparation de la vérité, déterminisme, identité avant rupture), non l'exactitude clinique des mécanismes ; les différences entre mondes sont plus défendables que leurs valeurs absolues, mais restent dépendantes des hypothèses injectées. Nous avons testé 14 structures alternatives plutôt qu'une seule (§2.7, §3.7) ; elles restent posées, non calibrées et non exhaustives (par exemple, aucun mécanisme ne fait dépendre la probabilité de reconnexion de la fragilité latente).

- **Paramètres.** Presque tous les paramètres sont au statut *hypothèse* ; la calibration sur des sources enregistrées (G2) est en attente. Les plus déterminants et les moins documentés sont la probabilité de retour après contact (réengagement) et, pour la séquence PREDICARE, l'atténuation de l'excès de risque sous télésurveillance, la participation et la reprise du suivi. Le plafond pour l'insuffisance cardiaque de la courbe du registre est tiré de l'expérience de prédiction interne (0,24) ; la valeur antérieure, optimiste (0,39), ne change aucune grandeur économique au-delà de l'erreur de Monte Carlo mais augmente l'évitement dû aux alertes d'environ 12 %. Un premier examen des agrégats simulés (Tableau S2) suggère que plusieurs taux d'événements, notamment la mise en dialyse et l'infarctus du myocarde, sont élevés par rapport aux ordres de grandeur français ; cela gonfle les valeurs absolues et peut amplifier les différences entre mondes.

- **Coûts et utilités.** Les montants remboursés sont des tarifs, non des coûts de production (les recommandations de la HAS privilégient les coûts de production français, tels que l'étude nationale de coûts à méthodologie commune, avec les tarifs en repli [25]) ; la fraction ambulatoire du coût de suivi (0,60) est posée et constitue le paramètre de coût le plus influent sur le signe du coût de la déshérence ; les coûts unitaires du programme sont posés et seront remplacés par la valorisation du démonstrateur ; plusieurs autres postes sont posés ; les utilités ne sont pas des valeurs EQ-5D-5L françaises ; la perspective collective repose sur des transitions vers la dépendance et des heures d'aide informelle posées, sans rétroaction sur la mortalité ni sur les QALY ; l'analyse de sensibilité probabiliste utilise des

plages de plausibilité déclarées, non des distributions estimées à partir de données, et n'élève le statut épistémique d'aucun paramètre.

- **Statut des résultats économiques.** Les marges répondent à une question formelle (quel coût fixe serait compatible avec un seuil de lecture sous ces hypothèses), non à la question pratique de ce que PREDICARE pourrait coûter ; elles sont une illustration secondaire, à recalculer après calibration (jalon G2) et une valorisation du démonstrateur, et ne doivent pas être lues comme un ordre de grandeur opérationnel. Le coût de télésurveillance est en partie posé : seule sa composante capteur de glucose (environ 86 € par mois) repose sur une base de remboursement publiée, qui ne couvre actuellement que les personnes traitées par insuline ; à 25 € par mois, les marges de PE sont supérieures d'environ 7 à 18 € (Tableau supplémentaire S14).

- **Structure.** Une seule condition d'entrée (diabète de type 2) dans le tronc de long terme ; aucun déménagement non signalé (le tronc quotidien en comporte ; Tableau supplémentaire S9-D) ; observance parfaite pendant le suivi ; aucun signal dans les données de remboursement précédant la rupture (les trajectoires pré-index sont plates), de sorte que l'anticipation de la déshérence elle-même ne peut être testée. Le diabète de type 2 est le premier cas d'usage : le phénotype plus large du syndrome métabolique dans les données de remboursement n'est pas encore formalisé, et les résultats ne s'étendent pas aux patients cardio-réno-métaboliques sans ALD diabète. Dans le cas de base, le bilan à T0 révèle le statut de MRC et d'insuffisance cardiaque sans erreur de mesure ni de classement ; la v0.16 ajoute des erreurs de mesure posées et permet au T0 de piloter l'équipement et la priorisation des alertes, mais pas encore l'intensité de la télésurveillance ni les seuils d'alerte. La capacité clinique derrière les alertes est désormais représentée par une capacité, un délai de réponse et une règle de priorisation ; l'hétérogénéité des réponses (appel, téléconsultation, visite, modification de traitement), la lassitude face aux alertes et l'apprentissage des seuils par l'équipe ne sont pas modélisés. Dans les analyses principales, la rupture est un mécanisme unique avec des multiplicateurs communs ; l'expérience structurelle fait varier sa part causale, la confusion, les formes partielles et intermittentes, les soins hors du système observé, la reconnexion et la poursuite du traitement, mais pas la diversité de ses causes (précarité, refus, déménagement, obstacles d'accès, effets indésirables, maladie aiguë intercurrente, dépression ou déclin cognitif), les retours déclenchés par des symptômes, l'observance graduée ou la reprise différée du traitement. Cette simplification peut rendre la chaîne rupture → progression → décès plus nette qu'elle ne l'est. La confusion résiduelle dans les schémas fondés sur les données de remboursement est, à l'inverse, un résultat plutôt qu'un défaut : le banc d'essai montre que la charge microvasculaire latente est largement invisible dans les données de remboursement, et les données réelles portent d'autres déterminants non enregistrés (position sociale, fragilité, préférences) qu'aucun ajustement sur des covariables observées ne peut neutraliser.

- **Chaînes IoMT et couplage des troncs.** Depuis la v0.15, deux chaînes distinctes sont portées par les mondes de long terme : dans la variante IoMT PC, réengagement → équipement → détection → évitement, la télésurveillance s'arrêtant à la rupture suivante ; dans la séquence PREDICARE PE, ciblage → invitation → bilan en unité mobile → télésurveillance pendant la déshérence → alerte et réponse. PC est une variante, non le fonctionnement prévu de PREDICARE. Dans les deux cas, le couplage avec le tronc quotidien est paramétrique (ses performances entrent comme paramètres), non dynamique : aucune physiologie quotidienne n'est simulée sur 20 ans, et la courbe du registre est posée : l'entraînement du modèle, le choix des étiquettes, la validation temporelle et le transfert entre territoires ne sont pas simulés.

- Lacunes de rapport. Les résultats coût-utilité non actualisés et un plan coût-efficacité ne sont pas rapportés ; l'analyse probabiliste ne couvre ni la capacité d'alerte ni le délai de réponse, qui n'ont été explorés que de manière déterministe ; les résultats du tronc quotidien autres que l'expérience de prédiction et les schémas d'évaluation n'ont pas été rejoués pour cette révision et conservent leur étiquette de version d'origine (Tableau supplémentaire S6).

Implications. Pour l'évaluation : le biais d'un schéma devrait être mesuré sur un contrefactuel synthétique avant que le schéma ne soit imposé aux patients, et, puisque les cohortes définies par la réapparition surestimaient le coût de la déshérence sous chaque mécanisme testé, les études sur données de remboursement ne devraient pas s'y fier sans un argument explicite établissant que les mécanismes à l'œuvre diffèrent de ceux testés ; un démonstrateur devrait randomiser l'offre — par grappes (territoires ou cabinets) allouées avant le début du ciblage, chacune avec sa propre capacité, ou par une loterie des places d'invitation à chaque période —, préspécifier quel effet le schéma choisi estime sous la capacité partagée (§3.5), conserver le dénominateur de toutes les personnes éligibles et rapporter l'effet de l'offre séparément de l'effet parmi les participants, puisque la participation est vraisemblablement sélective. Un déploiement en escalier n'est qu'une option conditionnelle : dans le banc d'essai, un essai en escalier surestimait les événements évités d'un facteur deux environ par confusion calendaire, de sorte qu'il exigerait des séquences randomisées, suffisamment de grappes et de périodes, une modélisation explicite des tendances calendaires et un plan pour la contamination, rien de quoi n'est démontré ici. Pour l'économie : un programme de vigilance devrait être défendu comme un coût par QALY sous une marge explicite, conditionnelle à la part causale de l'excès de risque de déshérence, et non comme une économie. Pour la conception : l'organisation compte davantage que la détection dans la variante PC (Tableau supplémentaire S11), équiper les patients réengagés n'était rentable que lorsque leurs pertes de signal étaient traitées hors du budget de réengagement, et la valeur de l'IoMT résidait davantage dans la continuité que dans l'évitement des décompensations ; dans la séquence PREDICARE PE, la valeur dépendait d'abord de la mesure dans laquelle la télésurveillance compense l'excès de risque et de la participation, tandis que la capacité et le délai de traitement des alertes importaient à peine ; quant au choix entre ordres de file, c'est une décision politique sur qui est protégé, qui relève des autorités élues et cliniques, non du modèle.

Technologie et intervention. Le banc d'essai sépare ce que la technologie apporte dans le modèle, la visibilité et l'alerte, de ce qui relève de l'intervention complexe qui l'entoure : l'observance, la continuité du traitement et la reconnexion aux soins. Sous le mécanisme posé, la marge de la séquence PREDICARE variait d'un facteur cinq environ (13 € à 70 € par personne-année couverte) selon l'atténuation de l'excès de risque, le renouvellement des traitements et l'attrition des capteurs, alors que les limites de capacité et les délais de réponse qui supprimaient jusqu'à 84 % des évitements dus aux alertes la laissaient pratiquement inchangée (§3.6). Dans ce modèle, l'économie d'un programme de vigilance dépend donc bien davantage de ce que l'équipe soignante fait avec et pour le patient que de la performance algorithmique. C'est une propriété de la structure simulée, non une démonstration : savoir si la télésurveillance peut seulement produire une telle atténuation, et par quel soutien humain, ne peut se lire dans une offre randomisée seule : l'effet de l'offre agrège toutes les composantes du programme, et son attribution à l'atténuation exigerait des mesures intermédiaires et des hypothèses explicites.

Couverture du registre. Dans le scénario de référence, 18,3 % des personnes éligibles sont venues au moins une fois à l'unité mobile et 86 % des participants ont été équipés au moins une fois : le registre observerait environ une personne éligible sur six. Ce sixième est sélectionné par la joignabilité, par la participation (posée comme décroissante avec la charge latente, de sorte que les plus malades sont sous-représentés) et par l'acceptation et le port du kit ; les personnes qui refusent l'équipement, peut-être celles dont la déshérence est la plus profonde, ne sont pas suivies de la même manière. Le nombre de mesures que collecte un registre ne dit donc rien de sa représentativité de la population ciblée, ni de sa capacité à contenir suffisamment d'événements bien datés pour apprendre des signatures robustes. Ce chiffre d'une personne sur six est une couverture cumulée simulée sous des taux posés de joignabilité, de participation, d'équipement et d'attrition ; il ne dit ni combien de personnes disposeront d'une série suffisamment complète pour être exploitée, ni combien d'événements confirmés chaque sous-groupe contiendra. Ce qu'un registre réel doit rapporter en premier, avant tout résultat prédictif, est sa couverture longitudinale exploitable : durée de port, données manquantes, événements datés et

confirmés, et différences de couverture selon le niveau de risque et selon les caractéristiques sociales et territoriales.

Ce que le banc d'essai dit, et ne dit pas, de la finalité du programme. PREDICARE vise d'abord à rendre observable une population en déshérence et à constituer un registre à partir duquel des motifs de dégradation seraient appris rétrospectivement, et seulement ensuite à orienter les patients vers les soins aigus (§1). Ces étapes constituent un programme de recherche encore à mener, non des capacités testées ici ; lu à l'aune de cette séquence, étape par étape, le banc d'essai éclaire certains maillons et en laisse d'autres intacts.

- **Constitution du registre** (quels patients ciblés sont atteints, équipés et restent mesurables) : partiellement modélisée, par la participation, l'acceptation de l'équipement, une attrition posée des capteurs et le traitement de la perte de signal (§3.4, §3.5) ; les patients ciblés qui refusent l'équipement ne sont pas suivis, alors qu'un registre réel devrait aussi les documenter.

- **Événements reliés aux séries** (décompensations datées, gradées, cliniquement confirmées, et périodes sans événement comparables) : le modèle connaît chaque événement par construction ; ce qui limite l'apprentissage en pratique est le nombre et la diversité des événements bien datés, non le nombre de mesures, et cela n'est pas représenté. Un registre réel a besoin d'une définition opérationnelle de l'événement — critères, source de confirmation, date retenue et procédure d'adjudication. La télésurveillance et le contact avec une équipe peuvent aussi modifier le recours aux soins et la probabilité qu'une décompensation soit documentée, de sorte que davantage d'événements peut refléter une meilleure détection, une dégradation réelle, ou les deux ; faute de les séparer, un modèle peut apprendre le processus de surveillance autant que la trajectoire physiologique.

- **Apprentissage rétrospectif** (quelles trajectoires précèdent les événements, avec quel délai d'anticipation, par rapport aux données de remboursement seules et au bilan à T0) : non simulé dans les mondes de long terme, où la détection est une courbe prescrite ; traité seulement par l'expérience interne sur un prodrome injecté (rappel de 24 %, environ 19 alertes par vrai épisode à un quota de 5 % ; §3.1).

- **Validation avant usage clinique** (période ultérieure, autre territoire, alertes en mode silencieux) : non simulée.

- **Orientation et ses conséquences** (triage, orientation vers le bon niveau de soins, délai avant évaluation, charge de travail des équipes, charge des services d'urgence, orientations inutiles) : non modélisées au-delà d'une capacité et d'un délai de réponse (§3.6) ; les décompensations et les décès sont modélisés.

Une réserve supplémentaire s'applique au registre lui-même : les patients sont observés après invitation, bilan, équipement et contact avec une équipe, autant d'éléments susceptibles de modifier leur comportement et leurs soins. Un registre ainsi constitué apprend des trajectoires sous-programme, non l'histoire naturelle d'une déshérence non prise en charge.

Le banc d'essai soutient donc le principe d'une surveillance plus dense pendant la déshérence sous hypothèses explicites, et non la valeur prédictive de motifs qui restent à apprendre, ni l'efficacité d'un modèle de triage, ni un soulagement des équipes cliniques. L'orientation exigerait de définir ce qui constitue un motif suffisamment fiable et urgent, les seuils cliniques et le circuit de réponse, et de mesurer la charge que génèrent les alertes ; sans cela, une surveillance continue peut déplacer le fardeau vers les équipes d'alerte ou les services aigus au lieu de soulager les professionnels. Avec une valeur prédictive positive d'environ 5 % par alerte dans l'expérience interne, ce risque n'est pas théorique.

5. Conclusion

La contribution principale de cet article est un banc d'essai de simulation dans lequel la vérité est connue par construction, montrant que, dans les mécanismes testés, les schémas d'évaluation fondés sur les données de remboursement faussent l'effet de la déshérence médicale : la sélection sur la réapparition hospitalière en surestimait le coût sous tous les mécanismes, tandis que le sens du biais de mortalité dépendait du mécanisme causal posé. Au sein de ce banc d'essai, et pour le diabète de type 2 comme premier cas d'usage, la déshérence médicale apparaît comme une perte de vies plutôt que comme une économie cachée, conséquence du mécanisme posé que les données réelles doivent encore confirmer. Ce que montre le banc d'essai est méthodologique, et limité aux configurations testées : dans celles-ci, les schémas testés sur données de remboursement surestimaient, diluaient ou différaient un effet connu, et, sur les 14 mécanismes posés testés, le biais à la hausse des cohortes fondées sur la réapparition persistait alors que le sens du biais de mortalité ne persistait pas. Les mondes appariés synthétiques rendent le biais d'un schéma d'évaluation mesurable avant qu'aucun patient n'y soit exposé, et déplacent l'argument économique en faveur de la vigilance de « combien cela fera-t-il économiser ? » vers « combien cela peut-il coûter ? ». La marge calculée ici répond à une question formelle — quel coût fixe de fonctionnement serait compatible avec un seuil de lecture sous les hypothèses posées — et non à la question pratique de ce que PREDICARE pourrait coûter : elle repose sur des paramètres cliniques non calibrés, une atténuation non mesurée de l'excès de risque, une participation et une reprise des soins non documentées, des utilités rénales indirectes, des transitions vers la dépendance posées et des coûts de fonctionnement inconnus. Une évaluation randomisée et préenregistrée de l'offre peut estimer l'effet du programme dans son ensemble, et non, à elle seule, la manière dont cet effet se répartit entre ses mécanismes. Pour le programme prévu — ciblage dans les données de remboursement, bilan en unité mobile, puis télésurveillance pendant la déshérence —, ce pilote devrait randomiser l'offre et conserver chaque personne éligible dans le dénominateur, car, si la participation est sélective comme posé ici, ceux qui viennent ne sont pas ceux qui restent à l'écart.

Disponibilité des données et du code

Le générateur (predicare-synth v0.19), ses fichiers de configuration, domaines de validité, jalons de décision, registre de provenance, 25 registres de décision d'architecture, tests automatisés, scripts d'expérience et le journal d'exécution de chaque résultat rapporté (Tableau 12 ; Tableau supplémentaire S6) sont versionnés et reproductibles par graine. Les tables de type SNDS (20 000 patients, environ 10 millions de lignes), les cohortes émulées et le répertoire de vérité terrain sont disponibles sous la forme d'un jeu de données synthétique séparé. Les paramètres, le statut de calibration, les distributions et les détails d'implémentation sont donnés dans les Tableaux supplémentaires S1–S14, et la valeur de chaque tableau par tâche et par graine dans un fichier de résultats du dépôt de code. Le logiciel est archivé sur Zenodo (DOI de concept https://doi.org/10.5281/zenodo.23474999 ; cette version, v0.19 : https://doi.org/10.5281/zenodo.23475000) sous licence Apache 2.0, et le jeu de données synthétique — tables de type SNDS, cohortes émulées, répertoire de vérité terrain, journaux d'exécution par graine et matériel supplémentaire — sous la forme d'un dépôt séparé (DOI de concept https://doi.org/10.5281/zenodo.23485751 ; cette version https://doi.org/10.5281/zenodo.23485752) sous CC-BY-4.0 ; l'extraction des dépenses Cnam utilisée pour les coûts unitaires est incluse dans ce dépôt sous sa licence ODbL d'origine. Les résultats du monde PE utilisent un coût de télésurveillance de 110 € par mois équipé : les Tableaux 10–11 et la Figure 6 ont été rejoués avec predicare-synth v0.19.2, qui reproduit la v0.19.1 bit pour bit sous les réglages par défaut (archive predicare-synth-v0.19.2.zip, SHA-256 b14994c3ef6cfb461f634d3f07ddc633b45eb8f4fa7f0f44ea49738098edfb45), et l'analyse probabiliste B a été corrigée exactement à partir de ses tirages rejoués ; scripts et journaux d'exécution

figurent dans la version 0.19.2 du jeu de données [DOI de version à ajouter au dépôt]. Aucune donnée de patient réel n'a été utilisée à aucune étape.

Tableau 12 - Provenance et statut des résultats rapportés : répliqués, à graine unique ou historiques ; version, réplication, population et script (carte complète dans le Tableau supplémentaire S6). « Conditionnel » : dépend de paramètres posés du programme (participation, atténuation, courbe du registre, coûts du programme).

Result (where reported)	Status	Version	Replication	Population / estimand	Script (code deposit)
Paired contrasts and claims designs (Abstract; Tables 5–6; Figures 2–4)	Replicated	v0.17	10 seeds × 60,000	E1, E2, E3	`revision/r1_runner.py`, job `base` (exp09, exp14)
Parametric sensitivity, 13 scenarios (§3.2–3.3)	Replicated	v0.19	10 seeds × 60,000 per scenario	E1, E4	`revision/sens_param_r1.py`, jobs `p9`
Cohort emulation and survival analyses (Table 7)	Single seed	v0.14 = v0.19	1 seed × 20,000 (seed 42)	E3	`export_snds_long` → `ciblage_cohortes_snds` → `evaluer_cohortes`
Cost-utility of re-engagement (Table 8; Figure 5)	Replicated	v0.17	10 seeds × 60,000	E4	job `base` (exp10)
IoMT variant PC (Supplementary Table S11; Figure S3)	Replicated; conditional	v0.19	10 seeds × 60,000	E4	jobs `e11`
Cost by population and perspective (Table 9; Supplementary Table S12)	Replicated	v0.17 / v0.19	10 × 60,000 (payer); 10 × 30,000 (collective)	E2, E4	jobs `base`, `e15`

Probabilistic analyses A and B (§3.4–3.5; Supplementary Figures S4–S5)	Replicated (300 draws); conditional	v0.19 (B: v0.19.2, 110 €)	300 draws × 20,000	E4	`revision/psa_r1.py`, `revision/psa_analyse_r1.py`
PREDICARE sequence and its estimands (Table 10; Figure 6; §3.5)	Replicated; conditional	v0.19.2 (110 €)	10 seeds × 60,000	E4–E7	jobs `e16`, `e16b`; `analyses/kit_realiste/R1/r1_kit.py`
T0 assessment, capacity, delay; registry ceiling (Table 11; §3.6)	Replicated; conditional	v0.19.2 (110 €; smax: v0.19, 25 €)	10 seeds × 60,000	E4	jobs `e17`, `smax`; `analyses/kit_realiste/R1/r1_kit.py`
Structural experiment (§3.7; Figure 7)	Replicated	v0.18	14 mechanisms × 10 seeds × 60,000	E1, E3, E4	`revision/exp_struct_r1.py`, jobs `s2`
Internal prediction experiment (§3.1)	Replicated; internal	v0.18	10 seeds (daily trunk)	E9	`revision/exp13b_prediction_r1.py`
Evaluation designs (§3.1; Supplementary Table S8)	Replicated	v0.18	10 seeds × 40,000 (daily trunk)	avoided events (ITT)	`revision/exp05b_designs_r1.py`
Historical daily-trunk results (Supplementary Table S9-D)	Historical, not re-run	v0.1–v0.14	1–5 seeds	E8, E9	`experiments/exp02`, `exp06–exp08`, `tools/`

Éthique

L'étude a utilisé exclusivement des données synthétiques ; aucun avis éthique ni aucune autorisation relative à la protection des données n'étaient requis. Tout usage du script de ciblage sur données réelles relève du cadre légal du SNDS et sort du périmètre du présent article.

Rapport

L'évaluation économique est rapportée selon CHEERS 2022 [36] (liste de contrôle dans le Tableau supplémentaire S4). Le modèle de simulation est décrit selon les recommandations de bonnes pratiques ISPOR–SMDM sur la transparence des modèles, leur validation et l'analyse d'incertitude [28,31] ; son statut de calibration est rapporté dans le Tableau supplémentaire S2.

L'étude de simulation est rapportée selon les lignes directrices STRESS [37] (variante à base d'agents, avec l'item à événements discrets sur les files d'attente ; Tableau supplémentaire S5), et son schéma suit la structure ADEMP des études de simulation [3].

Matériel supplémentaire

Les Tableaux supplémentaires S1 (paramètres, valeurs, sources et statut épistémique ; S1-C, inventaire des coûts en euros 2024), S2 (statut de calibration), S3 (analyses de sensibilité probabilistes et détails d'implémentation), S4 (liste de contrôle CHEERS 2022), S5 (liste de contrôle STRESS), S6 (carte des résultats : version, script, commande et réplication de chaque résultat), S7 (expérience structurelle), S8 (schémas d'évaluation), S9 (expériences de détection : expérience de prédiction interne, sensibilité au plafond du registre et campagnes historiques), S10 (estimands de l'offre : toutes les personnes éligibles contre personnes invitées), S11 (variante IoMT PC), S12 (perspective collective) et S13 (tests automatisés), ainsi que les Figures supplémentaires S1 (parcours PREDICARE avec ses points d'entrée de coûts et d'effets sur la santé), S2 (calibration dans l'expérience de prédiction interne), S3 (courbe du registre du monde PC), S4 et S5 (analyse probabiliste A), accompagnés des résultats de chaque tableau par tâche et par graine, sont fournis dans un fichier séparé et seront déposés avec le code.

Financement

Financé par le Twingital-Institute.

Liens d'intérêts

Jérôme Vétillard est vice-président R&D et Chief Product Officer de Qualees (TweenMe), qui développe des solutions de jumeaux numériques et de vigilance, dont le volet opérationnel du programme PREDICARE (Sentinelle IA), et il est le fondateur du Twingital Institute, qui porte le programme doctrinal PREDICARE. Afin de permettre un examen indépendant des résultats produits sous ce lien d'intérêts, le code complet, la configuration, les journaux d'exécution et les données synthétiques sont déposés en accès ouvert (Disponibilité des données et du code), et chaque résultat de cet article peut être régénéré à partir d'eux.

Contributions de l'auteur

Jérôme Vétillard : conceptualisation, méthodologie, supervision, validation, rédaction, relecture et révision.

Usage de l'IA générative

La plateforme TweenMe® a été utilisée pour implémenter et orchestrer les composants de simulation à base d'agents du démonstrateur PREDICARE. L'auteur a fait et revu les choix de conception, les résultats et le texte, et assume l'entière responsabilité du contenu.

Remerciements

Aucun

Références

Les références sont numérotées dans l'ordre de première citation (style Vancouver). Les notices bibliographiques et les DOI ont été vérifiés auprès de PubMed et des pages des éditeurs ou des agences le 16 octobre 2026 ; les trois références ajoutées dans cette révision (Morris et al., O'Hagan et al. et INSEE) le 17 octobre 2026.

1. Inspection générale des affaires sociales, Inspection générale des finances. Revue de dépenses relative aux affections de longue durée. Paris: IGAS-IGF; 2024 (erratum juillet 2025). Disponible sur : https://igas.gouv.fr/sites/igas/files/2025-07/Rapport%20Igas-IGF%20affections%20longue%20dur%C3%A9e%20(revue%20de%20d%C3%A9penses).pdf

2. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758-64. doi:10.1093/aje/kwv254

3. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074-102. doi:10.1002/sim.8086

4. Ndumele CE, Rangaswami J, Chow SL, et al. Cardiovascular-kidney-metabolic health: a presidential advisory from the American Heart Association. Circulation. 2023;148(20):1606-35. doi:10.1161/CIR.0000000000001184

5. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int. 2024;105(4S):S117-S314. doi:10.1016/j.kint.2023.10.018

6. Brownrigg JR, Hughes CO, Burleigh D, et al. Microvascular disease and risk of cardiovascular events among individuals with type 2 diabetes: a population-level cohort study. Lancet Diabetes Endocrinol. 2016;4(7):588-97. doi:10.1016/S2213-8587(16)30057-2

7. Yen FS, Wei JC, Shih YH, Hsu CC, Hwu CM. Impact of individual microvascular disease on the risks of macrovascular complications in type 2 diabetes: a nationwide population-based cohort study. Cardiovasc Diabetol. 2023;22(1):109. doi:10.1186/s12933-023-01821-8

8. Giugliano D, Scappaticcio L, Longo M, et al. GLP-1 receptor agonists and cardiorenal outcomes in type 2 diabetes: an updated meta-analysis of eight CVOTs. Cardiovasc Diabetol. 2021;20(1):189. doi:10.1186/s12933-021-01366-8

9. McGuire DK, Shih WJ, Cosentino F, et al. Association of SGLT2 inhibitors with cardiovascular and kidney outcomes in patients with type 2 diabetes: a meta-analysis. JAMA Cardiol. 2021;6(2):148-58. doi:10.1001/jamacardio.2020.4511

10. Plateforme des données de santé (Health Data Hub). Données synthétiques – top diabète [dataset]. data.gouv.fr; 2024. Licence Ouverte 2.0. Disponible sur : https://www.data.gouv.fr/datasets/donnees-synthetiques-top-diabete

11. Caisse nationale de l'Assurance Maladie. Pathologies : dépenses remboursées affectées à chaque pathologie, 2015–2024 [dataset]. data.gouv.fr; version du 6 juillet 2026 (millésime 2024 ; CSV complet,

SHA-256 8b05ec57…6816505). Disponible sur : https://www.data.gouv.fr/datasets/pathologies-depenses-remboursees-affectees-a-chaque-pathologie

12. Institut national de la statistique et des études économiques (INSEE). Indice des prix à la consommation – ensemble, base 2015, moyennes annuelles. Paris: INSEE. Disponible sur : https://www.insee.fr/fr/statistiques/6036866

13. Agence de la biomédecine. Rapport annuel REIN 2024 (Réseau épidémiologie et information en néphrologie). Saint-Denis: ABM; 2026. Disponible sur : https://www.agence-biomedecine.fr/en/observatoire-de-la-maladie-renale-chronique-reports

14. Chaudhry SI, Wang Y, Concato J, Gill TM, Krumholz HM. Patterns of weight change preceding hospitalization for heart failure. Circulation. 2007;116(14):1549-54. doi:10.1161/CIRCULATIONAHA.107.690768

15. Boehmer JP, Hariharan R, Devecchi FG, et al. A multisensor algorithm predicts heart failure events in patients with implanted devices: results from the MultiSENSE study. JACC Heart Fail. 2017;5(3):216-25. doi:10.1016/j.jchf.2016.12.011

16. Clarke P, Gray A, Holman R. Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). Med Decis Making. 2002;22(4):340-9. doi:10.1177/0272989X0202200412

17. Stout NK, Goldie SJ. Keeping the noise down: common random numbers for disease simulation modeling. Health Care Manag Sci. 2008;11(4):399-406. doi:10.1007/s10729-008-9067-6

18. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91(434):444-55. doi:10.2307/2291629

19. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30. Disponible sur : http://jmlr.org/papers/v12/pedregosa11a.html

20. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009;28(25):3083-107. doi:10.1002/sim.3697

21. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457-81. doi:10.2307/2281868

22. Aalen OO, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand J Stat. 1978;5(3):141-50.

23. Hagenmeyer EG, Koltermann KC, Dippel FW, Schädlich PK. Health economic evaluations comparing insulin glargine with NPH insulin in patients with type 1 diabetes: a systematic review. Cost Eff Resour Alloc. 2011;9:15. doi:10.1186/1478-7547-9-15

24. Beaudet A, Clegg J, Thuresson PO, Lloyd A, McEwan P. Review of utility values for economic modeling in type 2 diabetes. Value Health. 2014;17(4):462-70. doi:10.1016/j.jval.2014.03.003

25. Haute Autorité de santé. Choix méthodologiques pour l'évaluation économique à la HAS [Choices in methods for economic evaluation]. Saint-Denis La Plaine: HAS; 2020. Disponible sur : https://has-sante.fr/upload/docs/application/pdf/2020-11/methodological_guidance_2020_-choices_in_methods_for_economic_evaluation.pdf

26. Boneschi S, Miron de l'Espinay A. Aides à l'autonomie des personnes âgées : qui paie quoi ? L'apport du modèle Autonomix – résultats 2019. Les Dossiers de la DREES n° 99. Paris: DREES; 2022. Disponible sur : https://drees.solidarites-sante.gouv.fr/sites/default/files/2022-07/dd99.pdf

27. Besnard X, Brunel M, Couvert N, Roy D. Les proches aidants des seniors et leur ressenti sur l'aide apportée – résultats des enquêtes « CARE » auprès des aidants (2015-2016). Les Dossiers de la DREES n° 45. Paris: DREES; 2019. Disponible sur : https://drees.solidarites-sante.gouv.fr/sites/default/files/2020-07/dd45.pdf

28. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Making. 2012;32(5):722-32. doi:10.1177/0272989X12458348

29. O'Hagan A, Stevenson M, Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. Health Econ. 2007;16(10):1009-23. doi:10.1002/hec.1199

30. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med Decis Making. 2014;34(3):311-26. doi:10.1177/0272989X13505910

31. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Med Decis Making. 2012;32(5):733-43. doi:10.1177/0272989X12454579

32. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002;58(1):21-9. doi:10.1111/j.0006-341x.2002.00021.x

33. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials. 2007;28(2):182-91. doi:10.1016/j.cct.2006.05.007

34. Walonoski J, Kramer M, Nichols J, et al. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J Am Med Inform Assoc. 2018;25(3):230-8. doi:10.1093/jamia/ocx079

35. Franklin JM, Schneeweiss S, Polinski JM, Rassen JA. Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. Comput Stat Data Anal. 2014;72:219-26. doi:10.1016/j.csda.2013.10.018

36. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. BMJ. 2022;376:e067975. doi:10.1136/bmj-2021-067975

37. Monks T, Currie CSM, Onggo BS, Robinson S, Kunc M, Taylor SJE. Strengthening the reporting of empirical simulation studies: introducing the STRESS guidelines. J Simul. 2019;13(1):55-67. doi:10.1080/17477778.2018.1442155