PRAETOR WORKING PAPERS

WP n° 2026-01 · Août 2026 · révision v1.1 — septembre 2026

Audit public de la Note de l'Observatoire (A–E) : sensibilité, robustesse au case-mix, stabilité inter-cycles et équité de couverture

A public audit of the Praetor Observatory A–E rating: weight sensitivity, case-mix robustness, inter-cycle stability and coverage equity

Tarik Tarri

Institut Praetor (Nexcellis), France

Institut Praetor — département de recherche de Nexcellis, éditeur du logiciel Praetor-Santé

contact@praetor-santé.com

v1.0 : 20 août 2026 · v1.1 : 16 septembre 2026

Protocole pré-enregistré et déposé avant analyse : DOI 10.5281/zenodo.22077779

Archive de réplication (données scellées, script, registre d'empreintes) : Zenodo, DOI 10.5281/zenodo.22825425

Note de version — v1.1 (16 septembre 2026)

La v1.0 rapportait, pour l'objectif principal, un pire cas de 13,6 % d'établissements changeant de lettre (pondération 60/30/10). Ce chiffre résultait d'un artefact d'arithmétique flottante du script d'analyse : pour 113 établissements couverts par le seul pilier certification — dont la note est, par construction de la règle de redistribution, invariante aux pondérations — la somme des poids 0,6 + 0,3 + 0,1, évaluée en norme IEEE 754, vaut 0,999 999 999 999 999 9 ; un composite exactement égal à 85 passait ainsi sous le seuil de la lettre B, et ces 113 établissements étaient comptés à tort comme changeant de lettre. Le même artefact affectait marginalement le point 45/35/20 (1 établissement). Le script v1.1 neutralise l'artefact par un arrondi à 10^{-9} avant seuillage (une ligne modifiée). Valeurs corrigées : 60/30/10 = 8,6 % (au lieu de 13,6 %), 45/35/20 = 11,16 % (au lieu de 11,20 %) ; le pire cas de la grille s'établit à **12,6 % (pondération 50/30/20, IC 95 % [11,2 ; 14,1])**. Aucune autre valeur du rapport n'est modifiée. Le verdict de l'objectif principal est inchangé — critères pré-enregistrés < 20 % et < 1 % satisfaits, avec une marge accrue. Le script d'analyse v1.1, joint à l'archive de réplication, permet, avec les données scellées qui y figurent, de reproduire chaque valeur et chaque figure du présent rapport ; l'empreinte du script v1.0 est consignée ci-dessous pour traçabilité. La détection provient d'une réimplémentation indépendante de la grille H1 — le dispositif de réplication fonctionnant ici comme prévu.

Empreintes SHA-256 :

script v1.1 — b4c003ea98dd52f8270ad0d1d0f5de76a93084152af8a6a2af1c43ab8e2d002b

script v1.0 — a56e209c8de2091ae2774fc2f48352792e14a0355214c9121ced074244aa35a8

Résumé

La Note de l'Observatoire est un indicateur composite A–E attribué par Praetor aux établissements de santé français du périmètre de certification HAS, construit exclusivement sur des données officielles ouvertes (certification HAS, indicateurs de qualité IQSS, satisfaction e-Satis). La présente étude soumet cet indicateur à un audit de robustesse pré-enregistré, mené par son producteur selon les standards OCDE/JRC applicables aux indicateurs composites, sur le parc entier (2 537 établissements, extraction scellée du 20/08/2026). L'objectif principal — la robustesse de la lettre aux choix de pondération — est atteint : au plus 12,6 % des établissements changent de lettre sous une réallocation de ±10 points des poids (pire cas, pondération 50/30/20), et aucun ne se déplace de plus d'une lettre. Deux objectifs secondaires sont également atteints : la note se révèle insensible à la lourdeur de la patientèle une fois le statut de certification et la catégorie contrôlés (ρ partiel = −0,007, contre −0,22 en brut), et stable entre cycles de certification (88,0 % des établissements à ±1 lettre, cohérence directionnelle de 100 % avec les décisions HAS). Le troisième objectif secondaire révèle en revanche une limite réelle : la couverture des piliers varie fortement selon la catégorie d'établissement (jusqu'à 82,6 points d'écart), reflet des périmètres de mesure des instruments nationaux. Conformément à l'engagement du protocole, ce résultat défavorable est publié et donne lieu à des recommandations d'affichage. L'intégralité du dispositif de preuve (protocole pré-enregistré, données scellées par empreintes SHA-256, code) est jointe en annexes et dans l'archive de réplication.

Mots-clés : indicateur composite ; qualité des soins ; certification HAS ; analyse de sensibilité ; case-mix ; transparence ; science ouverte.

Abstract

The Observatory Rating is an A–E composite indicator assigned by Praetor to French healthcare facilities within the HAS certification perimeter, built exclusively from official open data (HAS certification, IQSS quality indicators, e-Satis patient satisfaction). This study submits the indicator to a pre-registered robustness audit conducted by its producer, following OECD/JRC standards for composite indicators, on the full population (2,537 facilities, sealed extraction of 2026-08-20). The primary objective — robustness of the letter grade to weighting choices — is met: at most 12.6% of facilities change letters under ±10-point weight reallocations (worst case, 50/30/20 weighting), and none moves by more than one letter. Two secondary objectives are also met: the rating proves insensitive to patient case-mix once certification status and facility category are controlled for (partial ρ = −0.007, versus −0.22 unadjusted), and stable across certification cycles (88.0% of facilities within ±1 letter, 100% directional coherence with HAS decisions). The third secondary objective, however, reveals a genuine limitation: pillar coverage varies strongly across facility categories (up to an 82.6-point gap), reflecting the measurement perimeters of the national instruments. As committed in the

protocol, this unfavourable result is published and motivates display recommendations. The full evidence chain (pre-registered protocol, SHA-256-sealed data, code) is provided in the appendices and the replication archive.

Keywords: *composite indicator; healthcare quality; HAS certification; sensitivity analysis; case-mix; transparency; open science. JEL: I11, I18, C18.*

Citation suggérée : Tarri, T. (2026). Audit public de la Note de l'Observatoire (A–E) : sensibilité, robustesse au case-mix, stabilité inter-cycles et équité de couverture. Praetor Working Papers, n° 2026-01, v1.1. Institut Praetor (Nexcellis). Archive de réplication : DOI 10.5281/zenodo.22825425.

Déclaration d'intérêts : cette étude est produite par l'Institut Praetor, département de recherche de Nexcellis, société éditrice du logiciel Praetor-Santé et productrice de la Note de l'Observatoire — l'indicateur audité ici. Ce lien d'intérêt est structurel et assumé ; il est contrebalancé par le pré-enregistrement des hypothèses et seuils avant analyse (annexe A), le scellement des données par empreintes SHA-256 (annexe B), la publication du code et l'engagement — tenu — de publier les résultats défavorables.

PARTIE I · RAPPORT D'ANALYSE

Audit public de la Note de l'Observatoire (A–E) — Rapport d'analyse. 20 août 2026 (v1.0), révisé le 16 septembre 2026 (v1.1, note de version) — analyse conforme au protocole pré-enregistré v1.0 (annexe A, SHA-256 7a2738ff987ff7c800ed4ce853e8e16215bfbc2c5260ddaaa81d3d47a27b072a) et au certificat d'extraction (annexe B, SHA-256 d7d7dfea4d80c9eee3fe1733ece08e088a700e349bd5afff1ceddff1ec00a8ce).

Données : 2 537 établissements du périmètre de certification HAS (parc entier), extraction gelée du 20/08/2026 12:35 UTC, commit b1aebfa4.

Résumé exécutif. L'objectif principal de l'audit est atteint : la lettre A–E est robuste aux choix de pondération, puisqu'au pire cas 12,6 % des établissements changent de lettre sous une réallocation de ±10 points des poids, et qu'aucun ne se déplace de plus d'une lettre. Deux des trois objectifs secondaires sont également atteints : la note est insensible à la lourdeur de la patientèle une fois le statut de certification et la catégorie contrôlés (ρ partiel = −0,007), et elle est cohérente dans le temps (88,0 % des établissements restent à ±1 lettre entre les cycles de certification, avec une cohérence directionnelle de 100 % avec les décisions HAS). Le troisième objectif secondaire, en revanche, met au jour une limite réelle : si le taux de « Non noté » est homogène entre groupes (écarts inférieurs à 10 points), la couverture partielle des piliers varie fortement selon la catégorie d'établissement — 88 % en santé mentale contre 5 % en soins pluridisciplinaires —, reflet des périmètres de mesure des instruments nationaux (IQSS, e-Satis). Conformément à l'engagement du protocole, ce résultat défavorable est publié tel quel et donne lieu à des recommandations d'affichage (section 5).

1. Introduction

Les indicateurs composites de qualité hospitalière — classements, étoiles, lettres — occupent une place croissante dans l'information du public, en France comme à l'étranger. Leur principale faiblesse méthodologique est connue de longue date : le choix des pondérations, des barèmes et des seuils y introduit une part d'arbitraire qui peut, si l'on n'y prend garde, déterminer davantage le classement final que les données elles-mêmes. C'est précisément pour cette raison que l'OCDE et le Centre commun de recherche de la Commission européenne (JRC) recommandent de soumettre tout indicateur composite à une analyse de sensibilité et de robustesse systématique avant publication (OCDE/JRC, 2008).

La Note de l'Observatoire est un indicateur composite de ce type. Elle attribue une lettre de A à E à chaque établissement de santé français du périmètre de certification HAS, à partir de trois piliers exclusivement issus de données officielles ouvertes : la certification HAS (poids publié de 50 %), un socle transversal de trois indicateurs de qualité et de sécurité des soins (IQSS, 40

%) et la satisfaction des patients mesurée par e-Satis (10 %). La formule, les barèmes et les seuils sont intégralement publiés sur la page « Méthodologie » de l'observatoire public Praetor.

La présente étude soumet cet indicateur à l'audit que recommandent les standards internationaux, avec une particularité assumée : l'audit est conduit par le producteur de l'indicateur lui-même. Pour que cette configuration ne vide pas l'exercice de sa valeur probante, l'ensemble du dispositif a été verrouillé à l'avance : les hypothèses, les seuils de décision et les méthodes ont été figés dans un protocole pré-enregistré et haché avant toute analyse (annexe A) ; les données ont été extraites en une seule fois, en lecture seule, et scellées par empreintes SHA-256 (annexe B) ; le moteur de notation exécuté est le code de production lui-même, sans aucune réimplémentation ; et la publication des résultats, favorables ou non, constituait un engagement du protocole — engagement tenu, comme le montre la section 4.4.

2. Objectifs de l'étude

L'étude poursuit un objectif principal et trois objectifs secondaires, qui correspondent respectivement aux hypothèses pré-enregistrées H1 à H4 du protocole (annexe A). La numérotation H1–H4 est conservée dans l'annexe, qui reproduit le document déposé à l'identique ; le corps du présent rapport adopte la terminologie usuelle des objectifs.

- **Objectif principal — sensibilité aux pondérations (hypothèse H1)** : vérifier que la lettre publiée n'est pas un artefact du choix des poids des trois piliers. Critères de jugement pré-enregistrés : sous chaque pondération alternative d'une grille simplexe de ±10 points, (a) moins de 20 % des établissements changent de lettre et (b) moins de 1 % changent de plus d'une lettre.
- **Objectif secondaire n° 1 — robustesse au case-mix (H2)** : vérifier que la note ne pénalise pas les établissements qui prennent en charge les patientèles les plus lourdes. Critères : corrélation partielle $|\rho| < 0{,}10$ et coefficient non significatif en régression logistique ordinale, conditionnellement au statut de certification, à la catégorie et à la taille.
- **Objectif secondaire n° 2 — stabilité inter-cycles (H3)** : vérifier que la note évolue dans le temps avec les décisions de certification, sans les amplifier ni les contredire. Critères : au moins 70 % des établissements à ±1 lettre entre cycles, et au moins 90 % de cohérence directionnelle avec le sens de la décision HAS.
- **Objectif secondaire n° 3 — équité de couverture (H4)** : vérifier que la probabilité d'être « Non noté » ou partiellement couvert ne varie pas indûment selon le statut juridique ou la catégorie. Critère : écarts inférieurs ou égaux à 15 points de pourcentage entre groupes.

3. Données et méthodes

3.1 Population

L'analyse porte sur le parc entier, sans échantillonnage : 2 537 établissements du périmètre de certification HAS, dont 2 241 notés (88,3 %) et 296 « Non notés » (11,7 %). La distribution des lettres — 278 A (12,4 %), 797 B (35,6 %), 796 C (35,5 %), 281 D (12,5 %) et 89 E (4,0 %) — est conforme à la calibration publiée (environ 12,5 / 35 / 36 / 12 / 4 %).

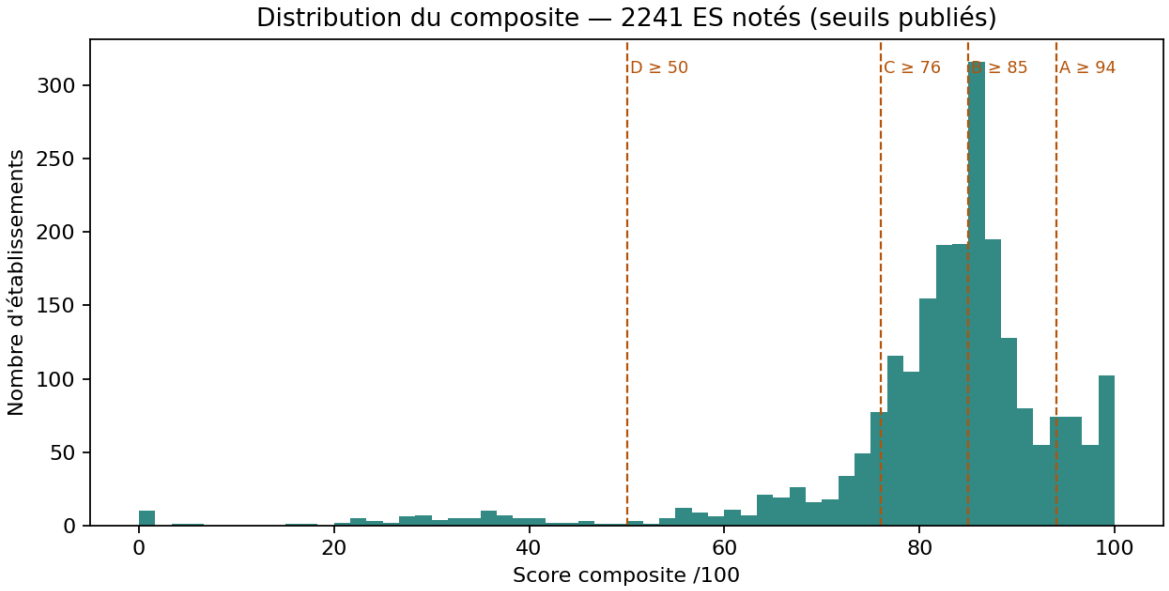


Fig. 0 — Distribution du score composite continu et seuils de lettre publiés.

3.2 Dispositif de preuve et instruments

Trois garanties structurent l'audit. En premier lieu, le protocole pré-enregistré (annexe A) a figé les hypothèses, les seuils et les méthodes le 20 août 2026, avant toute analyse ; son empreinte SHA-256 en interdit toute modification rétroactive. En deuxième lieu, l'extraction des données (annexe B) a été réalisée en lecture seule sur la base de production, le moteur de notation de production étant exécuté tel quel, sans réimplémentation ; les empreintes SHA-256 des fichiers extraits ont été vérifiées identiques entre le serveur et le poste d'analyse. En troisième lieu, la reproductibilité est assurée par l'archivage des scripts dans l'archive de réplication, une graine aléatoire pré-enregistrée (20260820), 1 000 rééchantillonnages bootstrap et la consignation des versions logicielles : Python 3.12.13, pandas 3.0.5, numpy 2.5.2, scipy 1.18.0, statsmodels 0.14.6, matplotlib 3.11.1.

4. Résultats

4.1 Objectif principal — la lettre est robuste aux pondérations

La grille simplexe pré-enregistrée fait varier les poids des piliers par pas de 5 points (certification entre 40 et 60, soins entre 30 et 50, expérience patient entre 0 et 20, somme égale à 100), ce qui engendre 18 pondérations alternatives. Pour chacune d'elles, la note de chaque établissement a été intégralement recalculée, puis comparée à la note officielle.

Les résultats sont sans ambiguïté. La proportion d'établissements changeant de lettre s'échelonne de 3,9 % pour la pondération 55/40/5 à 12,6 % au pire cas, atteint pour la pondération 50/30/20 (IC 95 % [11,2 ; 14,1]) — soit nettement en deçà du seuil pré-enregistré de 20 %. Surtout, aucun établissement ne change de plus d'une lettre, sous aucune des 18 pondérations (0 cas sur 2 241 × 18 recalculs). Les pondérations les plus proches de la pondération officielle (45/45/10, 55/35/10, 50/35/15) ne déplacent que 4 à 5 % des lettres. Les établissements couverts par le seul pilier certification, dont la note est invariante aux pondérations par construction de la règle de redistribution, ne changent de lettre sous aucune pondération (voir la note de version v1.1).

> **Objectif principal atteint** (critères pré-enregistrés de l'hypothèse H1 satisfaits : 12,6 % < 20 % et 0 % < 1 %). La lettre publiée reflète les piliers eux-mêmes, et non un artefact du choix de pondération.

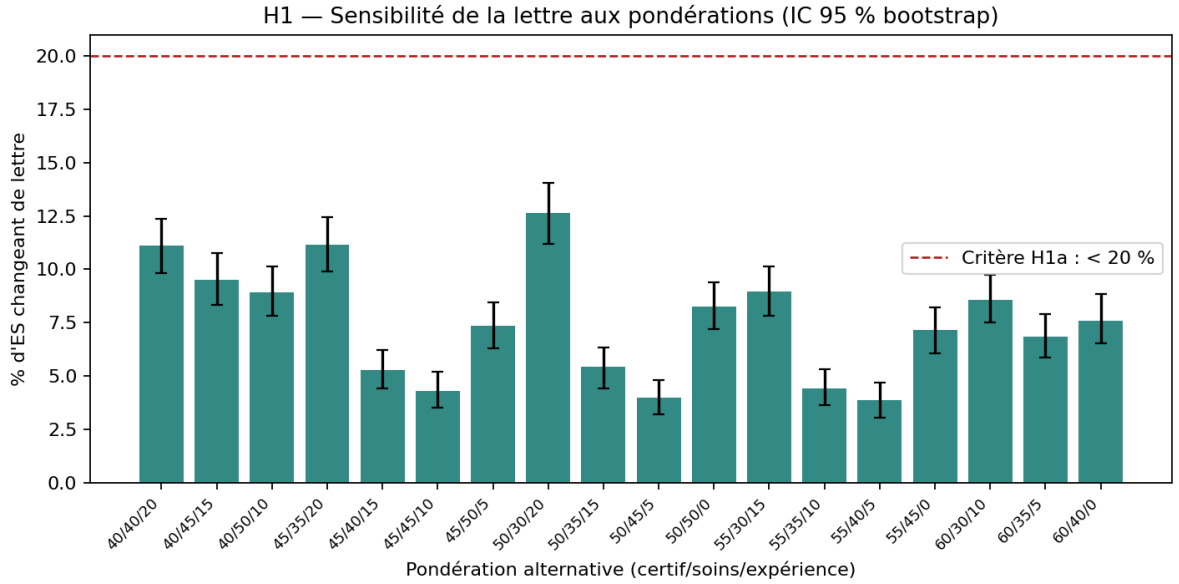


Fig. 1 — Proportion d'établissements changeant de lettre par pondération alternative (IC 95 % bootstrap, 1 000 tirages). Ligne rouge : critère pré-enregistré (20 %). Valeurs v1.1.

4.2 Objectif secondaire n° 1 — la note ne pénalise pas les patientèles lourdes

L'analyse porte sur les 1 037 établissements MCO disposant d'un indice de lourdeur d'activité ScanSanté et d'une capacité en lits exploitable. Prise isolément, la corrélation brute entre la note et le case-mix est négative et significative ($\rho = -0{,}224$, $p < 0{,}001$) : les établissements accueillant les patientèles les plus lourdes semblent moins bien notés. Cette lecture brute serait toutefois trompeuse. Une fois le statut de certification, la catégorie d'établissement et la taille pris en compte, l'association disparaît entièrement : la corrélation partielle de Spearman tombe à −0,007, et le coefficient du case-mix dans la régression logistique ordinale est nul (β standardisé = −0,003, p = 0,974, modèle convergé). L'association apparente était donc un effet de composition — les catégories et les statuts de certification ne se répartissent pas uniformément selon la lourdeur de la patientèle — et non un biais de l'indicateur.

> **Objectif secondaire n° 1 atteint** (critères de l'hypothèse H2 satisfaits : $|-0{,}007| < 0{,}10$ et $p = 0{,}974$). À statut de certification et catégorie donnés, la Note ne pénalise pas les établissements qui prennent en charge les patients les plus lourds.

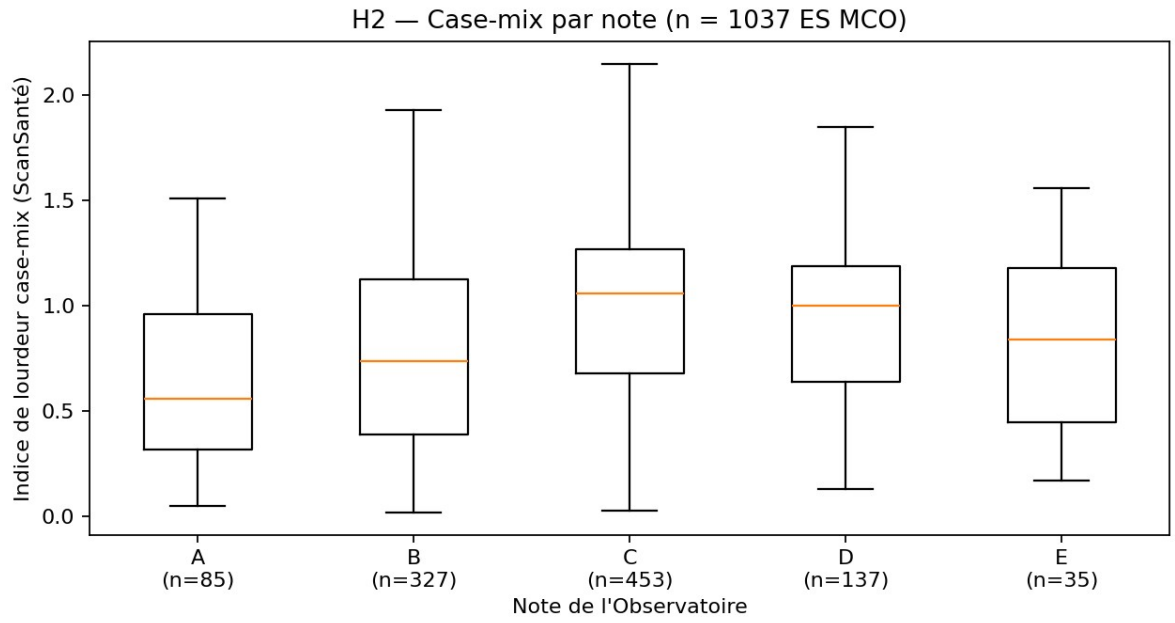


Fig. 2 — Distribution de l'indice de lourdeur par note (n = 1 037 ES MCO).

4.3 Objectif secondaire n° 2 — la note suit les décisions de certification dans le temps

Pour apprécier la stabilité temporelle de l'indicateur, la note de chaque établissement a été recalculée deux fois, en ne faisant varier que la décision de certification : une première fois avec la décision du cycle V2014, une seconde avec celle du référentiel 2021, tous les signaux qualité étant maintenus constants. L'analyse porte sur les 2 055 établissements disposant des deux décisions.

Entre les deux cycles, 38,2 % des établissements conservent exactement la même lettre et 88,0 % restent à ±1 lettre, au-delà du seuil pré-enregistré de 70 %. Les mouvements observés ne traduisent pas une instabilité du composite mais l'évolution réelle des décisions : 1 300 établissements ont vu leur décision HAS s'améliorer entre les cycles et 191 la voir se dégrader, et la cohérence directionnelle est de 100 % dans les deux sens — aucun établissement dont la décision s'améliore ne voit sa note recalculée baisser, et réciproquement. Une lecture structurelle complète ce constat : l'échelle V2014 ne comportant pas de « mention », la lettre A était mécaniquement plus rare sous l'ancien cycle ; les transitions massives de B vers A et de C vers B traduisent ce changement d'échelle décidé par la HAS, non un artefact du composite.

> **Objectif secondaire n° 2 atteint** (critères de l'hypothèse H3 satisfaits : 88,0 % ≥ 70 % et 100 % ≥ 90 %). La note bouge quand — et seulement dans le sens où — la certification bouge.

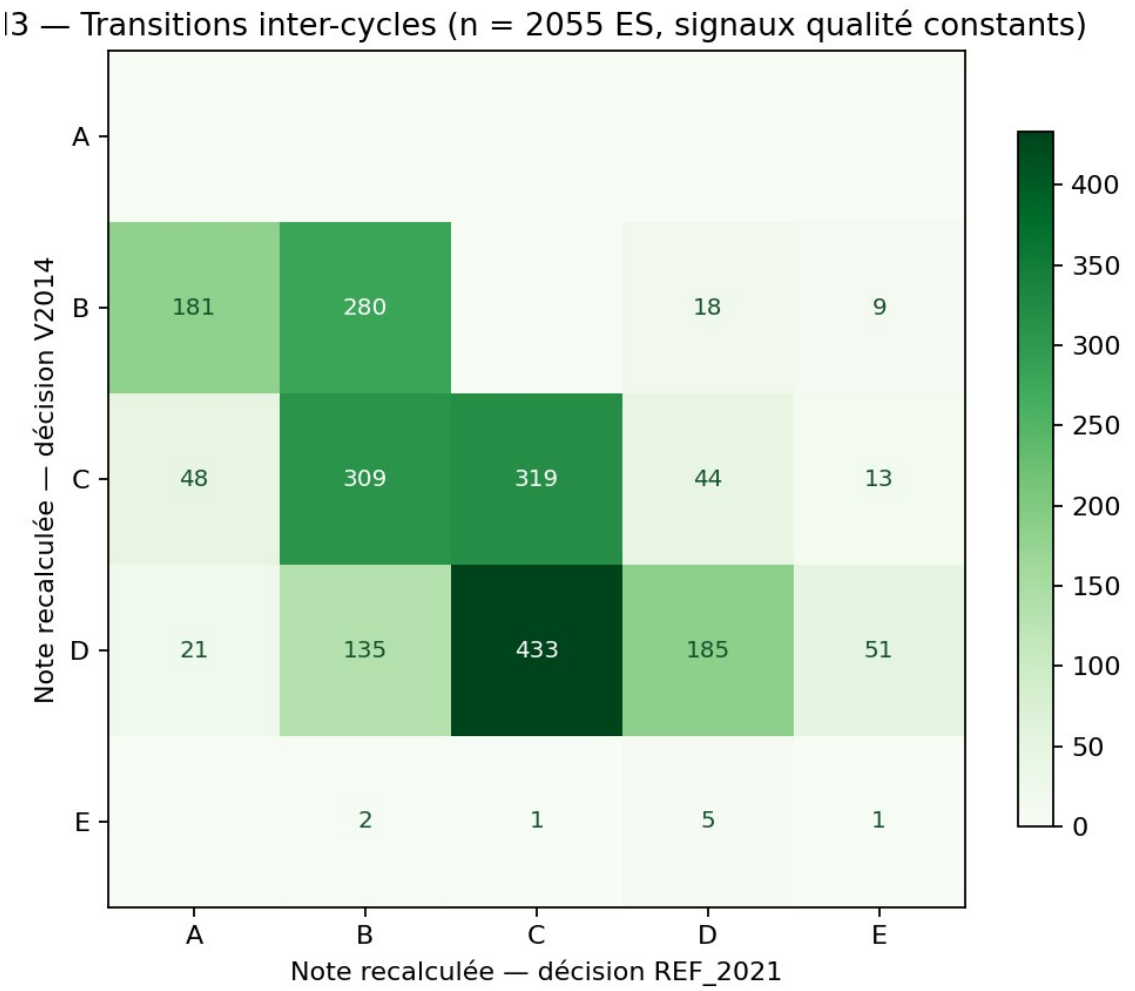


Fig. 3 — Matrice de transition des notes recalculées, décision V2014 (lignes) → décision REF_2021 (colonnes), signaux qualité constants.

4.4 Objectif secondaire n° 3 — la couverture des piliers est inégale selon les catégories

Le dernier objectif portait sur l'équité de couverture : un indicateur public ne doit pas laisser certains groupes d'établissements systématiquement « Non notés » ou notés sur une information appauvrie. Deux grandeurs ont été examinées, avec le même seuil pré-enregistré de 15 points de pourcentage d'écart maximal entre groupes.

Sur la première grandeur, le taux de « Non noté », le critère est respecté : 8,5 % dans le public, 5,3 % dans le privé non lucratif et 14,1 % dans le privé lucratif, soit un écart maximal de 8,8 points ; entre catégories, l'écart maximal est de 9,8 points. Ces écarts sont statistiquement significatifs (χ^2, p Holm < 0,001) mais d'ampleur modérée (V de Cramér ≈ 0,11).

Sur la seconde grandeur, la couverture partielle des piliers, le critère n'est pas respecté, et de loin, comme le détaille le tableau ci-dessous.

Groupe	n	% NN	% couverture partielle
Public	820	8,5	42,2
Privé non lucratif	450	5,3	60,4
Privé lucratif	1 069	14,1	29,9
ES privé autorisé en SSR	567	5,1	34,7
Centre Hospitalier	474	7,8	23,8
ES de Soins Pluridisciplinaire	389	5,7	5,4
Maison de Santé pour Maladies Mentales	234	11,1	88,0
Centre hospitalier, ex Hôpital local	168	14,9	79,8

L'écart de couverture partielle atteint 30,5 points de pourcentage entre statuts juridiques et 82,6 points entre catégories (χ^2 catégorie × couverture : $p < 10^{-127}$, V de Cramér = 0,57, soit un effet fort). L'explication est documentée et tient aux instruments nationaux eux-mêmes : e-Satis ne mesure que les champs MCO et SSR, et une partie des IQSS du socle n'est pas recueillie en psychiatrie ni dans les ex-hôpitaux locaux. La règle publiée de redistribution du poids vers la certification garantit qu'une note reste calculable pour ces établissements, mais elle a pour conséquence que la note des catégories concernées repose principalement sur le seul pilier certification.

Objectif secondaire n° 3 partiellement non atteint (hypothèse H4 partiellement rejetée : taux de NN dans le critère avec des écarts < 15 points, mais couverture partielle hors critère avec 30,5 et 82,6 points > 15). La comparaison de notes entre catégories différentes doit être présentée avec la couverture des piliers.

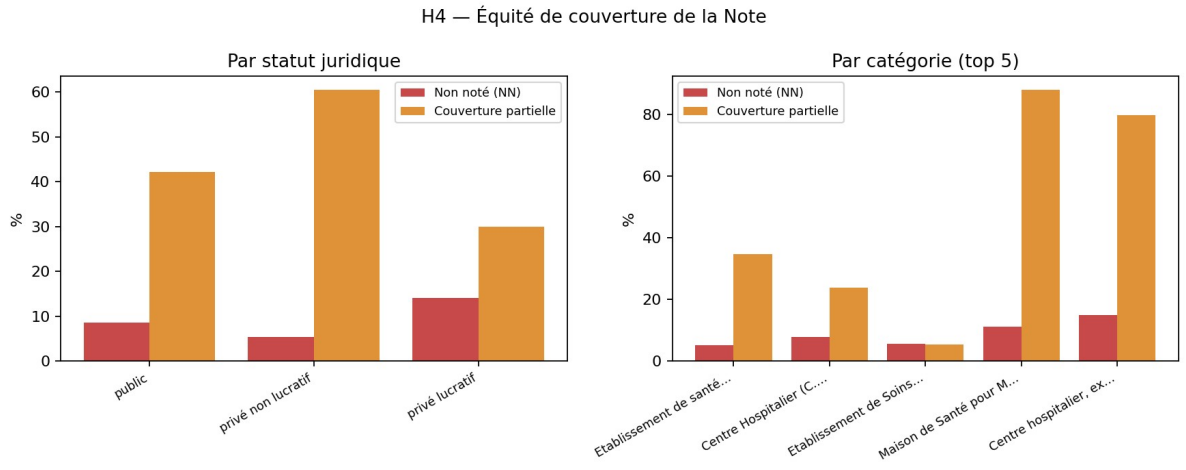


Fig. 4 — Taux de NN et de couverture partielle par statut juridique et par catégorie (top 5).

5. Recommandations issues de l'audit

1. Affichage de la couverture : rendre systématiquement visible, sur la fiche publique et dans les comparateurs, le nombre de piliers couverts (« note fondée sur 1/2/3 piliers »), particulièrement pour la psychiatrie et les ex-hôpitaux locaux.
2. Prudence comparative inter-catégories : privilégier les comparaisons au sein d'une même catégorie ; l'écart de couverture (V = 0,57) rend les comparaisons trans-catégories asymétriques.
3. Documentation méthodologique : intégrer les chiffres de couverture du présent audit à la page « Méthodologie & sources » de l'observatoire public.
4. Ré-audit périodique : rejouer le présent protocole à chaque évolution du moteur (pondérations, seuils, socle) ou à chaque nouveau millésime BQSS.

6. Écarts au protocole

1. Le protocole annonçait « 15 pondérations alternatives » ; la définition pré-enregistrée de la grille (bornes, pas de 5, somme 100) en engendre en réalité 18. La définition formelle fait foi ; les critères s'appliquant à chaque pondération, l'écart est sans incidence sur le verdict.
2. Dans la régression ordinale de l'objectif secondaire n° 1, le statut de certification a été introduit en score numérique standardisé plutôt qu'en indicatrices : les indicatrices provoquent une séparation quasi parfaite (la classe E coïncide avec les non-certifiés) qui empêche la convergence du maximum de vraisemblance. Le ρ de Spearman partiel, calculé comme pré-enregistré, confirme le même résultat.
3. v1.1 (16/09/2026) : correction d'un artefact d'arithmétique flottante affectant deux points de la grille de l'objectif principal (60/30/10 et 45/35/20) — voir la note de version. Aucune autre valeur modifiée ; verdicts inchangés.

7. Discussion et limites

Quatre limites doivent être gardées à l'esprit dans la lecture de ces résultats. D'abord, les données ouvertes utilisées (millésimes BQSS et e-Satis) présentent un décalage temporel de plusieurs mois avec la situation réelle des établissements. Ensuite, l'analyse de robustesse au case-mix est limitée au périmètre MCO, seul champ pour lequel un indice de lourdeur ScanSanté est disponible (1 037 établissements) ; elle ne couvre donc ni le SSR, ni la psychiatrie, ni l'HAD. Par ailleurs, la comparaison inter-cycles repose sur une correspondance d'échelles V2014 ↔ REF_2021 qui relève d'une convention, documentée en section 4.3 ; le score sur 100 de certification n'étant pas publié par cycle, la comparaison porte sur les statuts de décision. Enfin, et surtout, il s'agit d'un auto-audit : l'étude est menée par le producteur de l'indicateur. Les contrepoids mis en place — protocole pré-enregistré et haché, données scellées, code archivé, moteur de production exécuté tel quel — trouvent leur meilleure

garantie dans le fait que le résultat défavorable de l'objectif secondaire n° 3 est publié intégralement, et que l'erreur de calcul de la v1.0, détectée par réplication indépendante, est corrigée publiquement en v1.1.

8. Déclaration d'intérêts

Cette étude est produite par l'Institut Praetor, département de recherche de Nexcellis, société éditrice du logiciel Praetor-Santé et productrice de la Note de l'Observatoire — l'indicateur audité ici. Ce lien d'intérêt est structurel et assumé. Il est contrebalancé par le dispositif de preuve de l'étude : hypothèses et seuils figés avant analyse (annexe A), données scellées par empreintes SHA-256 (annexe B), code publié, moteur de production exécuté tel quel, et engagement — tenu — de publier les résultats défavorables.

9. Conclusion

La Note de l'Observatoire résiste aux trois tests de robustesse classiques des indicateurs composites : elle n'est l'artefact ni des pondérations (objectif principal), ni de la composition de la patientèle (objectif secondaire n° 1), et elle suit fidèlement — sans les amplifier — les décisions de certification dans le temps (objectif secondaire n° 2). L'audit identifie en revanche une limite réelle et structurelle : l'inégalité de couverture des instruments nationaux entre catégories d'établissements (objectif secondaire n° 3), qui appelle un affichage renforcé de la couverture plutôt qu'une correction du composite — corriger reviendrait à imputer des données inexistantes. L'ensemble du dispositif de preuve (protocole, données scellées, code, empreintes) permet à tout tiers de reproduire chaque chiffre du présent rapport — la correction apportée par la v1.1, issue d'une réplication indépendante, en est la démonstration en acte.

Références

- OCDE / JRC (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing, Paris.
- Haute Autorité de Santé. Base sur la qualité et la sécurité des soins (BQSS) ; Certification des établissements de santé (V2014, référentiel 2021) ; Indicateurs de qualité et de sécurité des soins (IQSS) ; e-Satis. Données ouvertes.
- ATIH / ScanSanté. Indice de lourdeur d'activité (case-mix), Open CCAM.
- Praetor (2026). Méthodologie de la Note de l'Observatoire. Observatoire public Praetor, page « Méthodologie & sources ».

ANNEXE A · PROTOCOLE PRÉ-ENREGISTRÉ (REPRODUIT INTÉGRALEMENT)

Protocole d'étude pré-enregistré — version 1.0. Figé le 20 août 2026, avant toute analyse des données. PDF autonome scellé : SHA-256 7a2738ff987ff7c800ed4ce853e8e16215bfbc2c5260ddaaa81d3d47a27b072a. Le document déposé utilise la numérotation H1–H4 ; elle correspond, dans le corps du rapport, à l'objectif principal (H1) et aux objectifs secondaires n° 1 à 3 (H2 à H4).

A.1 Contexte et objectif

La Note de l'Observatoire est un indicateur composite A–E attribué par Praetor à chaque établissement de santé du périmètre de certification HAS, à partir de trois piliers exclusivement officiels : certification HAS (poids publié 50 %), socle transversal de trois IQSS (40 %), expérience patient e-Satis (10 %). La formule, les barèmes et les seuils sont intégralement publiés (page « Méthodologie » de l'observatoire public).

Objectif de l'étude : soumettre cet indicateur à un audit public de robustesse, mené par son propre producteur, selon les standards d'audit des indicateurs composites (OCDE/JRC) : sensibilité aux pondérations, robustesse au case-mix, stabilité temporelle inter-cycles, équité de couverture. La publication des résultats — favorables ou non — est un engagement du protocole.

A.2 Données

A.2.1 Source et périmètre.

- Base : base de production Praetor (PostgreSQL), schéma signal — données exclusivement issues de sources ouvertes officielles (HAS/BQSS, e-Satis, FINESS, ScanSanté/ATIH), retraitées par les pipelines d'ingestion Praetor.
- Périmètre : ensemble des établissements hasEligible = true (démarches de certification HAS). Aucun échantillonnage : parc entier.
- Extraction : une extraction unique, horodatée, exécutée sur le serveur de production en lecture seule. Chaque fichier extrait est scellé par une empreinte SHA-256 consignée dans le certificat d'extraction. Les analyses ne portent que sur cette extraction gelée.

A.2.2 Variables.

Variable	Source	Rôle
Statut de certification courant, score /100, date	HAS (BQSS)	Pilier certification
3 IQSS du socle transversal (douleur, information médicale, hygiène SCAS), normalisés 0-100, dernière valeur par métrique	HAS BQSS	Pilier qualité des soins
4 métriques e-Satis (/100), dernière valeur par métrique	e-Satis	Pilier expérience patient
Catégorie d'établissement, statut juridique, nombre de lits, département, région	FINESS	Covariables H2/H4
Indice de lourdeur d'activité (case-mix, Open CCAM)	ScanSanté/ATIH	Exposition H2 (MCO uniquement)
Décisions de certification historiques par cycle (V2014 et REF_2021)	HAS BQSS (historique)	H3

A.2.3 Critères d'exclusion et données manquantes.

– ES hors périmètre HAS : exclus (jamais notés, par définition).

– Aucune imputation : un pilier non couvert suit la règle publiée de redistribution du poids vers la certification ; un ES sans certification exploitable est « Non noté » (NN) et compte dans H4.

– H2 restreinte au sous-ensemble avec case-mix mesuré (MCO) — l'effectif sera rapporté.

A.3 Hypothèses pré-enregistrées

H1 — Sensibilité aux pondérations. Sous perturbation des poids des piliers sur une grille simplexe au pas de 5 points (certification ∈ [40 ; 60], soins ∈ [30 ; 50], expérience ∈ [0 ; 20], somme = 100), pour chaque pondération alternative : (a) la proportion d'établissements notés changeant de lettre est < 20 % ; (b) la proportion changeant de plus d'une lettre est < 1 %. Statistique descriptive exacte sur le parc entier, intervalles de confiance à 95 % par bootstrap non paramétrique (1 000 rééchantillonnages).

H2 — Robustesse au case-mix. Conditionnellement au statut de certification, à la catégorie et à la taille (lits, log), l'association entre la Note (ordinale A=5…E=1) et l'indice de lourdeur case-mix est négligeable : $|\rho| < 0{,}10$ (corrélation de Spearman partielle) et coefficient non significatif ($\alpha = 0{,}05$) dans une régression logistique ordinale note ~ case-mix + statut certif + catégorie + log(lits). Interprétation pré-engagée : $|\rho| \geq 0{,}10$ significatif = « la note est sensible au case-mix », publié tel quel avec discussion des mécanismes possibles.

H3 — Stabilité inter-cycles. En recalculant la note de chaque ES avec la décision de certification du cycle V2014 puis celle du cycle REF_2021 (signaux qualité constants, seuls les

intrants certification changent), sur les ES disposant des deux décisions : (a) ≥ 70 % des ES restent dans la même lettre ou bougent d'une lettre au plus ; (b) parmi les ES dont la décision HAS s'améliore (resp. se dégrade) d'un cycle à l'autre, la note ne se dégrade pas (resp. ne s'améliore pas) pour ≥ 90 % d'entre eux (cohérence directionnelle). Matrices de transition publiées intégralement.

H4 — Équité de couverture. Le taux de « Non noté » (NN) et le taux de couverture partielle (piliers manquants) ne diffèrent pas de plus de 15 points de pourcentage entre statuts juridiques (public / privé non lucratif / privé lucratif) ni entre les cinq catégories d'établissements les plus fréquentes. Tests du χ^2 d'indépendance ($\alpha = 0{,}05$), correction de Holm pour la famille de tests, tailles d'effet (V de Cramér) rapportées.

A.4 Méthodes et instruments

- Moteur de notation : le code de production lui-même (computePublicRating / computePublicRatingDetailed, module public-rating.ts de signal-service), exécuté au commit git consigné dans le certificat d'extraction — aucune réimplémentation, donc aucun risque de divergence.
- Reconstruction des entrées : même requête « dernier signal par (ES, métrique) » que le service de production.
- Analyses statistiques : Python ≥ 3.11 (pandas, scipy, statsmodels) ; régression logistique ordinale (OrderedModel) ; bootstrap percentile à 1 000 tirages ; graphiques matplotlib. Versions consignées dans le rapport.
- Reproductibilité : scripts d'extraction et d'analyse archivés avec l'étude ; graine aléatoire fixée (seed = 20260820) ; empreintes SHA-256 de chaque fichier de données et de chaque livrable consignées dans un registre.

A.5 Limites connues à l'avance

- Données open data : décalage possible de plusieurs mois avec la situation réelle des ES.
- Case-mix disponible uniquement pour l'activité MCO — H2 ne couvre pas SSR/PSY/HAD.
- e-Satis déclaratif, sensible au taux de réponse ; son poids (10 %) borne l'impact.
- La comparaison inter-cycles (H3) repose sur la correspondance des échelles de décision V2014 ↔ REF_2021, explicitée dans le rapport.
- Étude menée par le producteur de l'indicateur : biais d'auto-évaluation possible, contrepoids = pré-enregistrement des seuils, publication intégrale des matrices et du code.

A.6 Livrables et publication

1. 01 — Le présent protocole (PDF, empreinte SHA-256).
2. 02 — Certificat d'extraction : horodatage, volumétrie, commit git, empreintes des données.
3. 03 — Rapport d'analyse complet avec figures, matrices de transition et code.
4. Publication selon la procédure de référencement Praetor v1.0 : dépôt Zenodo (DOI), dépôt HAL, série Praetor Working Papers (ISSN en cours de demande), ajout aux Publications de référence de l'observatoire public.

ANNEXE B · CERTIFICAT D'EXTRACTION (REPRODUIT INTÉGRALEMENT)

Certificat d'extraction des données. Établi le 20 août 2026. PDF autonome scellé : SHA-256 d7d7dfea4d80c9eee3fe1733ece08e088a700e349bd5afff1ceddff1ec00a8ce.

B.1 Conditions d'extraction

Champ	Valeur
Horodatage (UTC)	2026-08-20 12:35:00.409 → 12:35:01.003 (durée < 1 s)
Serveur	Serveur de production Praetor (nom d'hôte caviardé pour publication) — exécution en lecture seule
Commit git du code de production	b1aebfa4e923822a045120d11cba8336938b1f0e
Moteur de notation	computePublicRatingDetailed() — module public-rating.ts de signal-service, exécuté tel quel
Script d'extraction	etude-01-extract.ts — SHA-256 0553d04c8731b3c9acc64beb71f57f83460dc9a04ca4eda4c97cadb69ce73250

B.2 Volumétrie

Grandeur	Valeur
Établissements du périmètre HAS — parc entier, aucun échantillonnage	2 537
— dont notés (A–E)	2 241 (88,3 %)
— dont Non notés (NN)	296 (11,7 %)
Signaux « dernière valeur » IQSS socle + e-Satis	9 005
Décisions de certification historiques (V2014 / REF_2021)	4 335
Indices case-mix ScanSanté disponibles (périmètre MCO)	1 150

B.3 Constantes du moteur au moment de l'extraction

- Poids des piliers : certification 0,50 · qualité des soins 0,40 · expérience patient 0,10.
- Barème statut HAS : certifié avec mention 100 · certifié 85 · avec recommandations 70 · sous conditions 55 · sursis 25 · non certifié 0.
- Seuils de lettre : A ≥ 94 · B ≥ 85 · C ≥ 76 · D ≥ 50 · E < 50.
- Socle IQSS : 3 indicateurs (douleur, information médicale, hygiène SCAS), minimum 2 pour couvrir le pilier.
- e-Satis : 4 métriques /100 (hosp. > 48 h, ambulatoire, SSR, taux de recommandation).

B.4 Fichiers scellés

Fichier	Empreinte SHA-256
data/etude01_es.csv — 2 538 lignes (en-tête + 2 537 ES), 33 variables	d2f05ea71961a9c8eb24ae5d7c1fdd13f2d9b71c78e556134cb7dcc17fbde2f1
data/etude01_meta.json — métadonnées d'extraction. Version publiée (v1.2) : champ hostname caviardé, tout le reste identique à l'original scellé (empreinte originale : 3497d4d522a8a8c4183c9ed25481dd46b843077fdd9449ce7db0d61c9298cd20)	319b84fd0af5f1ba4b955c2f89eaced330725aff2643b767962802bf8862d25d

Attestation d'intégrité. Les empreintes SHA-256 ont été calculées une première fois sur le serveur de production immédiatement après l'extraction, puis recalculées sur le poste d'analyse après transfert : elles sont strictement identiques. Les analyses de l'étude 2026-01 portent exclusivement sur ces fichiers gelés.

B.5 Nature des données

Les données extraites proviennent exclusivement de sources ouvertes officielles (HAS/BQSS, e-Satis, FINESS, ScanSanté/ATIH) retraitées par les pipelines d'ingestion Praetor. Elles décrivent des établissements (personnes morales identifiées par leur FINESS public) et ne contiennent aucune donnée à caractère personnel.

Praetor Working Papers · WP n° 2026-01 · v1.0 du 20 août 2026, v1.1 du 16 septembre 2026 · Institut Praetor — département de recherche de Nexcellis, éditeur du logiciel Praetor-Santé. Protocole pré-enregistré : DOI 10.5281/zenodo.22077779. Archive de réplication (données scellées, script, registre d'empreintes SHA-256) : Zenodo, DOI 10.5281/zenodo.22825425. Référentiel : OCDE/JRC (2008), Handbook on Constructing Composite Indicators.