

PRISES DE POSITION RELATIVES AUX CONCEPTS D'ÉVALUATION ET DE PUBLICATION

MESURE NATIONALE DE L'EXPÉRIENCE DES PATIENT-E-S (PREMS) SOINS AIGUS POUR ADULTES, PSYCHIATRIE POUR ADULTES ET RÉADAPTATION

Date: 18.08.2026 (Ce document a été traduit à l'aide de DeepL Pro)

Les concepts d'évaluation et de publication ont été approuvés par les membres de l'ANQ. Les commentaires reçus sur ces concepts ont été pris en compte lors de leur révision. Les remarques qui n'ont pas été retenues sont expliquées ci-après sous forme de questions et réponses.

QUALITÉ DES DONNÉES

Nettoyage des données

Comment la qualité des données est-elle garantie lors de la numérisation des questionnaires papier ?

w hoch 2 compare les réponses fournies par l'institut de numérisation avec celles reçues via le questionnaire en ligne. À cette fin, un programme spécifique a été développé pour vérifier les codages et les répartitions des réponses. On part ici du principe que les répartitions des réponses des deux modalités d'enquête devraient correspondre pour l'essentiel. De plus, des images des questionnaires numérisés sont examinées par échantillonnage. Une attention particulière est portée aux questions pour lesquelles d'éventuelles erreurs seraient difficiles à détecter par la seule comparaison des répartitions des réponses (par exemple, l'exactitude du codage du sexe). Une autre source d'erreur potentielle, découverte lors de la mesure Swiss PREMs concernant les soins aigus de 2025 et qui fera l'objet d'un contrôle spécifique à l'avenir, concerne les éventuelles erreurs de codage. Celles-ci peuvent survenir lorsque le pli du papier se trouve directement au-dessus des cases de réponse.

En cas de suspicion d'erreurs systématiques, il est possible de relire les fichiers image des questionnaires numérisés à l'aide d'un logiciel développé en interne par w hoch 2, afin de vérifier la plausibilité des réponses extraites par l'institut de numérisation.

La numérisation des questionnaires papier sera supprimée à l'avenir, car à partir de 2027, l'Enquête se déroulera exclusivement en ligne (voir [la fiche descriptive](#)).

Dans quelle mesure les données collectées sont-elles ajustées ?

L'institut d'analyse Unisanté n'apporte aucune modification au contenu des données collectées. Si des valeurs non plausibles apparaissent – ce qui, d'après l'expérience, n'est le cas que pour l'année de naissance ou le nombre de questionnaires envoyés –, celles-ci sont simplement marquées comme valeurs manquantes et exclues des analyses tenant compte de l'âge. Le calcul de l'expérience des patient-e-s reposant exclusivement sur les questions spécifiques aux PREMs, cela n'a aucune incidence sur les résultats. Il s'agit là d'un nettoyage des données nécessaire à l'assurance qualité et non d'une manipulation des valeurs.

Les valeurs manquantes sont-elles remplacées ?

Non, il n'y a pas de remplacement (imputation) des valeurs manquantes. Les données incomplètes restent vides et ne sont pas prises en compte dans les analyses correspondantes. Les raisons de cette approche ainsi que les fondements méthodologiques relatifs à la non-participation et à la pondération sont expliqués en détail dans les sections suivantes.

Quelles réponses sont considérées comme non exploitables ?

Les réponses vides ou les réponses d'évitement (par exemple « Ne s'applique pas ») ne sont pas prises en compte dans le calcul des thèmes, car ces résultats se basent exclusivement sur les catégories de réponse habituelles (par exemple « Toujours » à « Jamais »). Si les réponses évasives ne sont pas assimilables à des valeurs manquantes au sens d'une lacune dans les données, elles ne sont toutefois pas exploitables pour le calcul des valeurs empiriques.

Représentativité

Comment la non-participation est-elle traitée et comment la représentativité des résultats est-elle évaluée ?

Étant donné qu'aucune valeur n'est remplacée (imputée) ou pondérée mathématiquement, les analyses reposent sur l'hypothèse méthodologique selon laquelle les valeurs manquantes apparaissent de manière aléatoire (« Missing at Random »). Il convient de noter que cette hypothèse n'est pas toujours pleinement vérifiée dans la pratique. Des corrections exhaustives par pondération de l'échantillon ne sont toutefois pas possibles, car cela nécessiterait de disposer d'informations détaillées sur l'ensemble de la population de référence.

Cela signifie que, pour procéder à une telle pondération, tous les hôpitaux et cliniques participants devraient transmettre les caractéristiques sociodémographiques de l'ensemble des patients sortis au cours de la période d'observation, qu'ils aient participé ou non à l'Enquête.

La charge de travail que cela impliquerait serait considérable et ne saurait être raisonnablement exigée de la plupart des établissements.

Les hôpitaux ou cliniques qui disposent des ressources nécessaires peuvent toutefois à tout moment comparer leurs résultats avec leurs données internes et effectuer des analyses approfondies.

Afin de remédier, au moins en partie, au problème de la représentativité au niveau national, la structure de l'échantillon est comparée aux valeurs de référence de l'Office fédéral de la statistique (OFS). Cela permet d'évaluer quels groupes de patients sont éventuellement sous-représentés dans l'échantillon.

Les données disponibles ne permettent pas de tirer des conclusions directes sur les causes de la non-participation. Une étude approfondie des mécanismes conduisant à la non-participation nécessiterait une analyse plus poussée, par exemple par le biais d'une enquête ciblée auprès des personnes n'ayant pas participé au relevé.

Comment la non-participation est-elle gérée en psychiatrie ?

En Psychiatrie pour adultes, on peut supposer que les groupes de patients particulièrement vulnérables sont sous-représentés dans l'échantillon.

Les limites qui en découlent pour l'interprétation des résultats sont expressément mentionnées dans l'Executive Summary.

PROTECTION DES DONNÉES

Pourquoi les seuils minimaux de cas requis pour les publications diffèrent-ils entre le portail web et le tableau de bord ?

Cette différence s'explique par le niveau de détail des données : sur le portail web de l'ANQ, les résultats sont agrégés et présentés sans lien avec les caractéristiques sociodémographiques. C'est pourquoi le seuil de publication est fixé à au moins trois questionnaires exploitables. Dans le tableau de bord, en revanche, les établissements disposent d'options de filtrage détaillées, par exemple par âge, sexe ou catégorie d'assurance. Étant donné que ce niveau de détail supplémentaire augmente le risque théorique de réidentification de personnes individuelles, le seuil de protection a été fixé à au moins cinq cas dans le tableau de bord. Il en résulte que les hôpitaux et les cliniques totalisant moins de cinq questionnaires exploitables n'ont pas accès au tableau de bord des résultats.

MESURE

Instruments de mesure

Pourquoi le questionnaire comporte-t-il autant de questions, et pourquoi ces questions sont-elles si similaires ?

L'objectif du nouveau questionnaire PREMs est de mettre à la disposition des hôpitaux un instrument de mesure complet et détaillé. Cet instrument a déjà été validé à plusieurs reprises

dans différents contextes. Il est apparu que sa longueur et sa complexité n'affectaient pas la qualité des données collectées ([voir validation psychométrique](#)). Il convient toutefois de tenir compte du fait que la qualité des réponses peut varier chez certains groupes de patients, par exemple chez les personnes très âgées ou chez les patients présentant des troubles cognitifs.

Le fait que certaines formulations se ressemblent est un choix méthodologique délibéré : afin de générer des valeurs thématiques statistiquement fiables, il est nécessaire de mettre en lumière différentes nuances d'un contenu sans pour autant être redondant. La [validation psychométrique](#) du questionnaire a permis de démontrer que c'est le cas pour le Swiss PREMs.

Méthodes d'analyse

Pourquoi utilise-t-on les valeurs « Top Box » et « Bottom Box » plutôt que de simples moyennes ?

La plupart des questions PREM sont évaluées sur une échelle de réponse à quatre niveaux. Celle-ci ne comporte plus de position « neutre » : pour répondre, les patients doivent choisir entre « plutôt bien » et « plutôt mal ». Cela signifie que l'échelle n'est pas équidistante et qu'il n'est donc pas possible d'en calculer une moyenne classique.

Une moyenne mélange les bonnes et les mauvaises réponses. Cela entraîne une perte d'information considérable. En revanche, les pourcentages des catégories « Top » et « Bottom » sont plus faciles à interpréter :

1. **Top Box** : cette valeur indique la proportion de patients ayant vécu la meilleure expérience possible.
2. **Bottom-Box** : cette valeur indique la proportion de patients ayant vécu une expérience critique. Elle regroupe les deux catégories de réponse les plus basses (par exemple « Jamais » et « Parfois »). Une valeur élevée de la Bottom-Box constitue un signal d'alerte important pour les établissements et met en évidence les domaines dans lesquels il existe un potentiel d'amélioration urgent.

Pourquoi la « Bottom-Box » comprend-elle deux catégories de réponse, alors que la « Top-Box » n'en comporte qu'une seule ?

Sur recommandation de l'institut d'analyse Unisanté, le [groupe qualité expérience des patient-e-s \(PREMs\)](#) a délibérément décidé de s'écarter, dans l'évaluation des Swiss PREMs, de la méthodologie de [l'Institut canadien d'information sur la santé \(ICIS\)](#) – qui ne prévoit que la catégorie la plus basse (par exemple « Jamais ») dans la « Bottom Box ». Cette décision vise avant tout à renforcer la pertinence statistique et repose sur les considérations suivantes :

- 1) Le nombre de réponses dans la catégorie la plus basse est très faible.
- 2) Le regroupement des deux catégories les plus basses permet d'augmenter le nombre de cas et, par conséquent, de réduire les fluctuations et d'obtenir des résultats statistiquement plus fiables.

- 3) Dans le tableau de bord, les hôpitaux et les cliniques peuvent toujours distinguer les deux catégories et consulter les résultats détaillés.

Pourquoi la troisième catégorie de réponse n'est-elle prise en compte ni dans la « Top Box » ni dans la « Bottom Box » ?

L'approche choisie se concentre spécifiquement sur les valeurs extrêmes, car celles-ci sont les plus pertinentes pour le développement ciblé de la qualité : elles mettent immédiatement en évidence les domaines dans lesquels les patients bénéficient déjà de la meilleure expérience possible et ceux où des problèmes nécessitent des mesures d'amélioration concrètes.

Pour autant, aucune information n'est perdue. Dans le tableau de bord des résultats, les hôpitaux et les cliniques ont accès à la répartition détaillée des réponses, tant au niveau des questions individuelles que des domaines thématiques.

Comment les seuils d'évaluation de l'expérience globale ont-ils été fixés ?

La transformation de l'échelle de réponse à 11 niveaux (de 0 à 10) pour la question « Expérience globale » s'appuie sur la définition de l'Institut canadien d'information sur la santé ([ICIS](#)), qui a élaboré et validé le questionnaire initial.

Lors de la première mesure nationale des PREMs en soins aigus, les données empiriques confirment les seuils proposés par l'ICIS.

Dans d'autres spécialités, comme la psychiatrie par exemple, la répartition peut toutefois s'en écarter systématiquement. Les seuils actuels ayant été validés principalement pour les soins aigus en médecine générale, une adaptation interdisciplinaire de ces valeurs est envisageable dans le cadre d'un perfectionnement méthodologique.

Indépendamment d'une éventuelle adaptation, la comparabilité est préservée. Les comparaisons s'effectuent toujours au sein d'une même spécialité (et non entre les spécialités), de sorte que les mêmes seuils s'appliquent à tous les hôpitaux et cliniques relevant de cette même spécialité.

Pourquoi ne calcule-t-on pas une simple moyenne pour l'expérience globale ?

Contrairement aux autres échelles de réponse, cette échelle à 11 niveaux est de nature continue, ce qui permettrait en principe de calculer une moyenne. Cependant, en raison de la distribution fortement asymétrique et de l'effet de plafond marqué, l'interprétabilité de cet indicateur est limitée.

L'utilisation des valeurs « top box » et « bottom box » renforce la pertinence des résultats et permet d'identifier plus facilement les domaines dans lesquels les performances sont particulièrement bonnes ou mauvaises.

DÉFINITION DES DOMAINES THÉMATIQUES

Pourquoi les domaines thématiques ne sont-ils pas pondérés en fonction du nombre de valeurs manquantes ?

L'ANQ a suivi la méthodologie du CIHI, qui ne prévoit pas une telle pondération (voir la section [4.1 « Composite Measures » dans Acute Care Patient-Reported Experience Measures: Methodology Notes](#)). D'une manière générale, la pondération soulève la question de l'absence de la base de données correspondante, par exemple au moyen de valeurs de référence ; sans cela, la procédure reviendrait à une hypothèse de « données manquantes aléatoires ».

Une pondération des questions individuelles au sein d'un domaine thématique n'aurait de sens sur le plan méthodologique que s'il existait de grandes différences dans les taux de réponse entre les différentes questions, susceptibles d'entraîner des biais dans les résultats. Les analyses de la répartition des valeurs manquantes et des réponses d'évitement montrent toutefois qu'il n'y a pas de divergences significatives pour la plupart des questions PREMs.

Lorsque des valeurs plus élevées apparaissent, il s'agit principalement de questions qui n'entrent pas en ligne de compte dans la définition des thèmes (par exemple, la question 11 « Aide pour aller aux toilettes » ou la question 12 « Traitement de la douleur »), ou de questions qui appartiennent toutes au même domaine thématique (questions 14 et 15 sur les nouveaux médicaments).

Les deux seules questions qui présentent une certaine hétérogénéité au sein de leur domaine thématique sont la question 8 « Aide rapide » dans le domaine « *Organisation* », qui comprend au total quatre questions, et la question 29 « Compréhension des médicaments » dans le domaine « *Sortie* », qui comprend au total trois questions.

Dans l'ensemble, il n'y a donc pratiquement aucun indice suggérant que l'absence de pondération des différentes questions entraîne des distorsions notables dans les valeurs des domaines thématiques.

AJUSTEMENT

Pourquoi renonce-t-on à l'ajustement au risque ?

Les variables sociodémographiques disponibles (âge, sexe, etc.) ne constituent pas les facteurs d'influence les plus pertinents pour l'expérience des patient-e-s. Un ajustement mathématique de ces variables pourrait donc générer de nouveaux biais. On recourt plutôt à une stratification. Dans ce cadre, l'échantillon est divisé en sous-groupes :

- Soins aigus : différenciation selon le type d'admission (urgence vs. admission élective).
- Psychiatrie : différenciation selon le statut du séjour (volontaire vs involontaire).
- Réadaptation : différenciation selon la cause (liée à un accident vs non liée à un accident).

La population de référence devient ainsi plus homogène, ce qui permet une comparaison équitable au sein de ce groupe plus homogène. Tant lors de la mesure pilote que lors de la première mesure nationale, il a été démontré que les patients admis via les urgences évaluent tous les domaines de manière significativement et systématiquement moins favorable que ceux admis en admission programmée.

L'ANQ a introduit de nouvelles fonctionnalités dans les graphiques en entonnoir disponibles sur le portail web de l'ANQ, qui permettent de comparer les différents groupes (par exemple : total, urgences, admissions programmées, etc.).

Les variables de stratification sont-elles suffisantes pour remplacer l'ajustement ?

Bien qu'une variable isolée ne puisse pas refléter entièrement le case-mix d'un hôpital, les données (mesure pilote ainsi que première mesure nationale en soins aigus) montrent que les variables de stratification choisies ont une influence significative sur les résultats. La stratification permet ainsi une comparaison nationale plus équitable, car elle repose sur la constitution de sous-groupes plus homogènes et tient ainsi compte de l'hétérogénéité des populations de patients.

Comment les différences de case-mix sont-elles prises en compte en psychiatrie ?

La stratification actuelle repose sur les informations disponibles à ce jour. L'un des objectifs majeurs de la première mesure nationale sera de vérifier si ces catégories – admission involontaire ou volontaire – s'avèrent être des variables de discrimination appropriées.

Les résultats sont-ils comparables malgré les différences entre les hôpitaux et les cliniques ?

L'objectif premier des graphiques en entonnoir et de la coloration des points est la comparaison d'un établissement donné avec la valeur de référence nationale, et non la comparaison entre les établissements. L'ANQ a par ailleurs intégré, grâce au nouveau questionnaire, les intervalles de confiance à 99 % dans les graphiques en entonnoir. Cela permet une gradation plus fine et plus précise des positions. Alors que, lors des précédentes mesures de la satisfaction des patients, les établissements se situaient soit en dehors, soit à l'intérieur des lignes de contrôle, on observe aujourd'hui des nuances plus fines. Le taux de fausses prédictions (voir le concept d'évaluation) garantit en outre que la coloration est prudente – c'est-à-dire que les hôpitaux et cliniques colorés s'écartent de la moyenne nationale avec un haut degré de certitude statistique.

PRÉSENTATION

Comment les hôpitaux et les cliniques sont-ils identifiés de manière univoque ?

Pour chaque année de mesure, chaque hôpital/clinique se voit attribuer le même numéro spécifique à l'évaluation (code de clinique). De nouveaux numéros ne sont attribués qu'en cas de fusion ou de restructuration en profondeur des hôpitaux. En outre, un identifiant ANQ

interdisciplinaire et le numéro BUR de l'OFS, s'il est disponible, sont utilisés. Ces identifiants permettent de visualiser les évolutions dans le temps et les tendances.

Pourquoi n'analyse-t-on pas l'évolution dans le temps ?

Une perspective longitudinale solide n'est possible qu'après l'achèvement de plusieurs périodes de mesure. Étant donné que seules les données de la première vague d'enquête sont actuellement disponibles, des analyses de tendances exhaustives ne sont prévues qu'à partir de 2027 (soins aigus) ou 2028 (réadaptation et psychiatrie).

Le tableau de bord des résultats permettra en outre d'examiner en détail le développement de la qualité au niveau des hôpitaux et des cliniques au fil du temps.