

STELLUNGNAHME VERNEHMLASSUNG AUSWERTUNGS- UND PUBLIAKTIONSKONZEPTE

NATIONALE MESSUNG PATIENTENERFAHRUNG (PREMS)

AKUTSOMATIK ERWACHSENE, ERWACHSENENPSYCHIATRIE UND REHABILITATION

Datum: 18.08.2026

Die Auswertungs- und Publikationskonzepte wurden von den Mitgliedern des ANQ genehmigt. Die Rückmeldungen zu den Konzepten sind in die Überarbeitung eingeflossen. Anmerkungen, die nicht berücksichtigt wurden, werden im Folgenden in Form von Fragen und Antworten erläutert.

DATENQUALITÄT

Datenbereinigung

Wie wird die Datenqualität beim Scanning der Papierfragebogen sichergestellt?

w hoch 2 vergleicht die vom Scanninginstitut gelieferten Antworten mit den über den Onlinefragebogen eingegangenen Antworten. Zu diesem Zweck wurde ein spezifisches Programm geschrieben, welches die Codierungen und die Antwortverteilungen überprüft. Dabei wird die Annahme getroffen, dass die Antwortverteilungen der beiden Befragungsmodalitäten im Wesentlichen übereinstimmen sollten. Zudem werden Bilder der gescannten Fragebogen stichprobenartig angeschaut. Dabei wird besonderes Augenmerk auf Fragen gelegt, bei denen mögliche Fehler mit dem Vergleich der Antwortverteilungen nur schwer erkennbar wären (z. B. Korrektheit der Codierung des Geschlechts). Eine weitere mögliche Fehlerquelle, die bei der Swiss PREMs Messung der Akutsomatik 2025 entdeckt wurde und zukünftig speziell kontrolliert wird, sind mögliche Falschcodierungen. Diese können entstehen, wenn der Papierfalz direkt über den Antwortkästchen liegt.

Bei Verdacht auf systematische Fehler besteht die Möglichkeit, die Bilddateien der gescannten Fragebogen mit einer eigenentwickelten Software von w hoch 2 erneut auszulesen, um die vom Scanninginstitut extrahierten Antworten zu plausibilisieren.

Das Scanning der Papierfragebogen wird künftig entfallen, da ab 2027 eine ausschliessliche Online-Befragung erfolgt (siehe [Faktenblatt](#)).

Inwieweit werden die erhobenen Daten bereinigt?

Das Auswertungsinstitut Unisanté nimmt keine inhaltlichen Änderungen an den erhobenen Daten vor. Sollten unplausible Werte auftreten – was erfahrungsgemäss nur beim Geburtsjahr oder der Anzahl der versendeten Fragebogen der Fall ist –, werden diese lediglich als fehlende Werte gekennzeichnet und von Analysen, die das Alter berücksichtigen, ausgeschlossen. Da die Berechnung der Patientenerfahrung ausschliesslich auf den spezifischen PREMs-Fragen basiert, hat dies keinen Einfluss auf die Ergebniswerte. Es handelt sich hierbei um eine notwendige Datenbereinigung zur Qualitätssicherung und nicht um eine Manipulation der Werte.

Werden fehlende Werte ersetzt?

Nein, es findet keine rechnerische Ersetzung (Imputation) fehlender Werte statt. Unvollständige Angaben bleiben leer und werden in den jeweiligen Analysen nicht berücksichtigt. Die Gründe für dieses Vorgehen sowie die methodischen Hintergründe zur Nichtteilnahme und zur Gewichtung werden in den folgenden Abschnitten detailliert erläutert.

Welche Antworten gelten als nicht verwertbar?

Leere Antworten oder Ausweichantworten (z. B. «Trifft nicht zu») werden bei der Berechnung der Themenbereiche nicht berücksichtigt, da diese Ergebniswerte ausschliesslich auf den regulären Antwortkategorien (z.B. «Immer» bis «Nie») basieren. Ausweichantworten sind zwar nicht mit fehlenden Werten im Sinne einer Datenlücke gleichzusetzen, sind jedoch für die Berechnung der Erfahrungswerte nicht verwertbar.

Repräsentativität

Wie wird mit der Nichtteilnahme umgegangen und wie wird die Repräsentativität der Ergebnisse beurteilt?

Da keine Werte rechnerisch ersetzt (imputiert) oder gewichtet werden, basieren die Auswertungen auf der methodischen Annahme, dass fehlende Werte zufällig auftreten («Missing at Random»). Es ist zu berücksichtigen, dass diese Annahme in der Praxis nicht immer vollständig erfüllt ist. Umfassende Korrekturen mittels einer Gewichtung der Stichprobe sind jedoch nicht möglich, da hierfür detaillierte Informationen über die gesamte Referenzpopulation vorliegen müssten.

Das heisst, für eine solche Gewichtung müssten alle teilnehmenden Spitäler und Kliniken soziodemographische Merkmale für sämtliche Patientinnen und Patienten übermitteln, die während der Messperiode ausgetreten sind- unabhängig davon, ob diese an der Befragung teilgenommen haben oder nicht.

Der damit verbundene Aufwand wäre erheblich und für die meisten Institutionen nicht zumutbar.

Einzelne Spitäler/Kliniken, die über die entsprechenden Ressourcen verfügen, können ihre Ergebnisse jedoch jederzeit mit internen Daten vergleichen und vertiefende Analysen durchführen.

Um die Repräsentativität auf nationaler Ebene zumindest teilweise zu adressieren, wird die Struktur der Stichprobe mit Referenzwerten des Bundesamts für Statistik (BFS) verglichen. Dies ermöglicht eine Einschätzung darüber, welche Patientengruppen in der Stichprobe möglicherweise unterrepräsentiert sind.

Die verfügbaren Daten erlauben keine direkten Rückschlüsse auf die Ursache der Nichtteilnahme. Eine sorgfältige Untersuchung der Mechanismen, die zur Nichtteilnahme führen, würde weitergehende Analyse erfordern, beispielsweise durch die gezielte Befragung von Personen, die nicht an der Erhebung teilgenommen haben.

Wie wird mit der Nichtteilnahme in der Psychiatrie umgegangen?

In der Erwachsenenpsychiatrie ist davon auszugehen, dass besonders vulnerable Patientengruppen in der Stichprobe unterrepräsentiert sind.

Die daraus resultierenden Einschränkungen für die Interpretation der Ergebnisse werden im Executive Summary ausdrücklich ausgewiesen.

DATENSCHUTZ

Warum gelten für das Webportal und das Dashboard unterschiedliche Mindestfallzahlen für die Publikation?

Der Unterschied ergibt sich aus der differenzierten Detailtiefe der Daten: Im ANQ-Webportal werden die Ergebnisse aggregiert und ohne Verknüpfung mit soziodemografischen Merkmalen dargestellt. Daher gilt hier eine Publikationsschwelle von mindestens drei auswertbaren Fragebogen. Im Dashboard hingegen stehen den Institutionen detaillierte Filteroptionen zur Verfügung etwa nach Alter, Geschlecht, oder Versicherungsklasse. Da diese zusätzliche Datentiefe das theoretische Risiko einer Re-Identifizierung einzelner Personen erhöht, wurde die Schutzwelle im Dashboard auf mindestens fünf Fälle festgelegt. Dies hat zur Folge, dass Spitäler und Kliniken mit insgesamt weniger als fünf auswertbaren Fragebogen gar keinen Zugang zum Ergebnisdashboard erhalten.

MESSUNG

Messinstrument

Warum umfasst der Fragebogen so viele Fragen und warum und so ähnliche Fragen?

Das Ziel des neuen PREMs-Fragebogens ist es, den Spitälern ein umfassendes und detailliertes Messinstrument zur Verfügung zu stellen. Dieses Instrument wurde bereits mehrfach in unterschiedlichen Kontexten validiert. Dabei hat sich gezeigt, dass seine Länge und

Komplexität die Qualität der erhobenen Daten nicht beeinträchtigen ([siehe psychometrische Validierung](#)). Gleichwohl ist zu berücksichtigen, dass die Ausfüllqualität bei bestimmten Patientengruppen, etwa bei sehr betagten Personen oder Patientinnen und Patienten mit kognitiven Einschränkungen, variieren kann.

Dass sich manche Formulierungen ähneln, ist methodisch beabsichtigt: Um statistisch verlässliche Themenwerte zu generieren, müssen verschiedene Nuancen eines Inhalts beleuchtet werden, ohne dabei redundant zu sein. Bei der [psychometrischen Validierung](#) des Fragebogens konnte nachgewiesen werden, dass dies beim Swiss PREMs der Fall ist.

Auswertungsmethoden

Warum werden Top- und Bottom-Box Werte anstatt einfacher Mittelwerte benutzt?

Die meisten PREM-Fragen werden auf einer vierstufigen Antwortskala gemessen. Diese hat keine «neutrale» Position mehr – die Patientinnen und Patienten müssen sich bei der Beantwortung zwischen «tendenziell gut» oder «tendenziell schlecht» entscheiden. Dies bedeutet, dass die Skala nicht äquidistant und, dass sich damit kein klassischer Mittelwert bilden lässt.

Ein Mittelwert vermischt gute und schlechte Antworten. Dies führt zu einem deutlichen Informationsverlust. Dagegen sind Top- und Bottom-Box Anteile leichter zu interpretieren:

1. **Top Box:** Dieser Wert zeigt den Anteil der Patientinnen und Patienten, die eine bestmögliche Erfahrung gemacht haben.
2. **Bottom-Box:** Dieser Wert zeigt den Anteil der Patientinnen und Patienten, die eine kritische Erfahrung gemacht haben. Er fasst die beiden niedrigsten Antwortkategorien (z. B. «Nie» und «Manchmal») zusammen. Ein hoher Bottom-Box-Wert fungiert für die Institutionen als wichtiges Warnsignal und macht deutlich, in welchen Bereichen dringendes Verbesserungspotenzial besteht.

Warum umfasst die Bottom-Box zwei Antwortkategorien, statt wie die Top-Box nur eine?

Auf Empfehlung des Auswertungsinstituts Unisanté hat der [QA Patientenerfahrung \(PREMs\)](#) bewusst entschieden, bei der Auswertung der Swiss PREMs von der Methodik des [Canadian Institut for Health Information](#) (CIHIs) – die in der Bottom-Box nur die niedrigste Kategorie (z. B. «Nie») vorsieht, abzuweichen. Diese Entscheidung dient primär der statistischen Aussagekraft und basiert auf folgenden Überlegungen:

- 1) Die Anzahl der Antworten in der niedrigsten Kategorie ist sehr gering.
- 2) Die Zusammenlegung der beiden niedrigsten Kategorien führt zu höheren Fallzahlen und damit zu geringeren Schwankungen sowie statistisch robusteren Ergebnissen.
- 3) Im Dashboard können die Spitäler und Kliniken weiterhin zwischen den beiden Kategorien unterscheiden und die detaillierten Ergebnisse einsehen.

Warum wird die dritte Antwortkategorie weder in der Top- noch in der Bottom-Box berücksichtigt?

Der gewählte Ansatz konzentriert sich gezielt auf die Extremwerte, da diese die höchste Aussagekraft für die gezielte Qualitätsentwicklung besitzen: Sie machen unmittelbar sichtbar, in welchen Bereichen Patientinnen und Patienten bereits die bestmögliche Erfahrung machen und wo Probleme bestehen, die konkrete Verbesserungsmaßnahmen erfordern.

Dennoch geht keine Information verloren. Die Spitäler und Kliniken haben im Ergebnisdashboard Zugang zu den detaillierten Antwortverteilungen sowohl auf Ebene der einzelnen Fragen als auch der Themenbereiche.

Wie wurden die Schwellenwerte für die Auswertung der Gesamterfahrung festgelegt?

Die Transformation der 11-stufigen Antwortskala (0 bis 10) bei der Frage «Gesamterfahrung» orientiert sich an der Definition des Canadian Institute for Health Information [CIHI](#), welches den ursprünglichen Fragebogen entwickelt und validiert hat.

Bei der ersten nationalen PREMs-Messung in der Akutsomatik bestätigt die empirische Datelage, die vom CIHI vorgeschlagenen Schwellenwerte.

In anderen Fachbereichen, wie beispielsweise in der Psychiatrie, kann die Verteilung jedoch systematisch abweichen. Da die aktuellen Schwellenwerte primär für die Akutsomatik validiert wurden, ist eine fachspezifische Anpassung dieser Werte im Sinne der methodischen Weiterentwicklung denkbar.

Unabhängig von einer möglichen Anpassung bleibt die Vergleichbarkeit gewahrt. Die Vergleiche erfolgen stets innerhalb eines Fachbereichs (nicht zwischen den Fachbereichen) sodass für alle Spitäler und Kliniken innerhalb desselben Fachbereichs die gleichen Schwellenwerte gelten.

Weshalb wird für die Gesamterfahrung kein einfacher Mittelwert berechnet?

Im Gegensatz zu den übrigen Antwortskalen ist diese 11er-Skala von Natur aus kontinuierlich, was die Berechnung eines Mittelwerts grundsätzlich ermöglichen würde. Aufgrund der stark schiefen Verteilung und des ausgeprägten Deckeneffekts ist die Interpretierbarkeit dieses Kennwerts jedoch eingeschränkt.

Durch die Verwendung von Top- und Bottom-Box Werten wird die Aussagekraft erhöht und es lassen sich leichter Bereiche identifizieren, in denen es besonders gut oder schlechter läuft.

BILDUNG DER THEMENBEREICHE

Warum werden die Themenbereiche nicht nach der Anzahl fehlender Werte gewichtet?

Der ANQ ist der Methodologie des CIHIs gefolgt, die eine solche Gewichtung nicht vorsieht (siehe Abschnitt [4.1 Composite Measures» in Acute Care Patient-Reported Experience Measures: Methodology Notes](#)). Grundsätzlich stellt sich bei einer Gewichtung die Frage nach der

entsprechenden Datengrundlage, beispielweise durch Referenzwerte, ohne diese käme das Verfahren einer «Missing at random»-Annahme gleich.

Eine Gewichtung der Einzelfragen innerhalb eines Themenbereichs wäre methodisch nur dann sinnvoll, wenn grosse Unterschiede in den Antwortraten zwischen den einzelnen Fragen bestünden, die zu Verzerrungen der Ergebnisse führen könnten. Analysen zur Verteilung der fehlenden Werte und Ausweichantworten zeigen jedoch, dass bei den meisten PREMs-Fragen keine signifikanten Diskrepanzen vorliegen.

Dort wo höhere Werte auftreten, handelt es sich überwiegend um Fragen, die nicht in die Themenbildung einfließen (z. B. Frage 11 «Hilfe für die Toilette» oder Frage 12 «Schmerzen Bekämpfung»), oder um Fragen, die alle demselben Themenbereich angehören (Fragen 14 und 15 zu neuen Medikamenten).

Die einzigen beiden Fragen, die innerhalb ihres Themenbereichs eine gewisse Inhomogenität aufweisen, sind Frage 8 «Schnelle Hilfe» im Bereich *Organisation*, der insgesamt vier Fragen umfasst und Frage 29 «Verständnis Medikamente» im Bereich *Austritt*, der insgesamt drei Fragen umfasst.

Insgesamt gibt es daher kaum Hinweise darauf, dass der Verzicht auf eine Gewichtung der einzelnen Fragen zu nennenswerten Verzerrungen bei den Themenbereichswerten führt.

ADJUSTIERUNG

Warum wird auf Risikoadjustierung verzichtet?

Die verfügbaren soziodemografischen Variablen (Alter, Geschlecht etc.) sind für die Patientenerfahrung nicht die relevantesten Einflussfaktoren. Eine mathematische Adjustierung dieser Variablen könnte deshalb neue Verzerrungen erzeugen. Stattdessen wird eine Stratifizierung eingesetzt. Hierbei wird das Kollektiv in Subgruppen unterteilt:

- Akutsomatik: Differenzierung nach Eintrittsart (Notfall vs. Elektiv).
- Psychiatrie: Differenzierung nach Aufenthaltsstatus (Freiwillig vs. Unfreiwillig).
- Rehabilitation: Differenzierung nach Ursache (Unfall- vs. Nicht-Unfall-bedingt).

Dadurch wird die zugrunde liegende Population homogener und ein fairer Vergleich innerhalb der homogenen Gruppe möglich. Sowohl in der Pilotmessung als auch in der ersten nationalen Messung konnte nachgewiesen werden, dass Patientinnen und Patienten, die über die Notfallstation eingetreten sind, alle Themenbereiche signifikant und systematisch schlechter bewerten als jene, die elektiv eingetreten sind.

Der ANQ hat in den Funnel Plots auf dem ANQ-Webportal neue Funktionen eingeführt, die einen Vergleich der unterschiedlichen Kollektive (z.B. Gesamt, Notfall, Elektiv, etc.) ermöglichen.

Reichen die Variablen zur Stratifizierung aus, um die Adjustierung zu ersetzen?

Obwohl eine einzelne Variable den Case-Mix eines Spitals nicht vollständig abbilden kann, belegen die Daten (Pilotmessung sowie erste nationale Messung Akutsomatik), dass die gewählten Stratifizierungsvariablen einen signifikanten Einfluss auf die Ergebnisse haben. Die Stratifizierung ermöglicht somit einen faireren nationalen Vergleich, da sie auf der Bildung homogenerer Untergruppen basiert und so der Heterogenität der Patientenkollektive Rechnung trägt.

Wie werden Case-Mix Unterschiede in der Psychiatrie berücksichtigt?

Die derzeitige Stratifizierung basiert auf den aktuell verfügbaren Informationen. Ein wichtiges Ziel der ersten nationalen Messung wird es sein, zu prüfen, ob sich diese Kategorien- unfreiwilliger vs. freiwilliger Eintritt - als geeignete Diskriminierungsvariable erweisen.

Sind die Ergebnisse trotz der Unterschiede zwischen den Spitälern und Kliniken vergleichbar?

Das primäre Ziel der Funnel Plots und der Färbung der Punkte ist der Vergleich mit dem nationalen Vergleichswert einer einzelnen Institution und nicht der Vergleich zwischen den Institutionen. Der ANQ hat zudem mit dem neuen Fragebogen in den Trichtergrafiken auch die 99-Prozent-Konfidenzintervalle hinzugezogen. Das ermöglicht eine feinere und präzisere Abstufung der Positionen. Während die Institutionen bei den früheren Messungen der Patientenzufriedenheit entweder ausserhalb oder innerhalb der Kontrolllinien lagen, gibt es heute feinere Abstufungen. Die False-Prediction-Rate (siehe Auswertungskonzept) sorgt zudem dafür, dass die Einfärbung konservativ erfolgt – das heisst, jene Spitäler und Kliniken, die eingefärbt werden, weichen mit hoher statistischer Sicherheit vom nationalen Mittelwert ab.

DARSTELLUNG

Wie werden Spitäler und Kliniken eindeutig identifiziert?

Pro Messjahr erhält Jedes Spital/jede Klinik die gleiche messspezifische Nummer (Klinikcode). Neue Nummern werden nur dann vergeben, wenn Spitäler fusionieren oder umfassend restrukturiert werden. Zusätzlich werden eine fachbereichsübergreifende ANQ-ID, und die BUR-Nr. des BFS, wenn vorhanden, verwendet. Mithilfe dieser Identifikatoren lassen sich zeitliche Entwicklungen und Trends visualisieren.

Warum wird keine zeitliche Entwicklung analysiert?

Eine fundierte Längsschnittperspektive ist erst nach Abschluss mehrerer Messperioden möglich. Da aktuell nur die Daten aus der ersten Erhebungswelle vorliegen, sind umfassende Trendanalysen erst ab dem Jahr 2027 (Akutsomatik) bzw. 2028 (Rehabilitation und Psychiatrie) vorgesehen.

Über das Ergebnisdashboard wird es zudem möglich sein, die Qualitätsentwicklung auf Ebene der Spitäler und Kliniken im Zeitverlauf detailliert zu untersuchen.