

A Decade of National Patient Experience Feedback in Romania (2016–2026): A Harmonized, Reproducible Open Dataset

Dan Valeriu Voinea ^{1,*}  and Ștefan Adrian Voinea ² 

¹ Department of Arts and Media, University of Craiova, 200585 Craiova, Romania

² Faculty of Political, Administrative and Communication Sciences, Babeș-Bolyai University, 400132 Cluj-Napoca, Romania; svoinea@gmail.com

* Correspondence: dan.voinea@gmail.com

Abstract

Romania's Mecanismul de Feedback al Pacientului (MFP; Patient Feedback Mechanism) has collected post-discharge patient experience feedback from public hospitals since 2016 and publishes monthly aggregate workbooks through official channels, but the files are heterogeneous in layout, inconsistently granular, and have never been consolidated into a documented, reusable resource. This data descriptor presents a harmonized, reproducible open dataset built from 117 original workbooks. A Python pipeline integrates them into three analysis-ready artifacts: a national monthly series (113 observed releases over the December 2016–May 2026 span, 6004 tidy answer rows, an 18-row question legend spanning a 2024 instrument revision); a provider release panel (493,450 rows, 466 county plus exact-name provider identities, 42 counties, 39 releases); and a structurally separate legacy 2016 paper-based instrument. Provenance was verified by re-downloading sampled sources and confirming byte-for-byte identity, and national totals reconstructed by hospital summation reconcile exactly with an independent re-aggregation of the published hospital-level details. The release adds SHA-256 source checksums, a machine-readable data dictionary, and a provenance log. The descriptor documents reuse hazards (a ten-to-eight-item instrument break, uneven granularity, aggregate-only records, and string-defined provider identity) and a reuse opportunity: provider- and county-level integrity (informal-payment solicitation) items, usable for rare-event analysis under explicit shrinkage caveats.

Dataset: <https://doi.org/10.5281/zenodo.21144263>

Dataset License: Processed datasets, generated outputs, and documentation: CC BY 4.0. Analysis and build code: MIT License. Original Ministry of Health workbooks are public-sector documents referenced by checksum and are not relicensed.

Keywords: patient feedback; patient-reported experience measures; patient satisfaction; open government data; data harmonization; informal payments; health-system transparency; reproducibility; FAIR data; Romania



Academic Editor: Rüdiger Pryss

Received: 3 July 2026

Revised: 23 July 2026

Accepted: 24 July 2026

Published: 25 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Summary

Patient experience is now treated as a core dimension of healthcare quality, rather than a peripheral measure of hospitality, because patients report on aspects of care that administrative and clinical records do not capture and systematic reviews have associated better reported experience with clinical, safety, and other quality-of-care outcomes [1,2].

The scientific value of patient feedback depends on collecting credible observations and documenting how they were produced so that others can interpret and reuse them appropriately. Contemporary patient-reported experience measures (PREMs) are expected to rest on explicit construct definitions and demonstrated validity and reliability [3]. National feedback systems that publish open data are therefore a valuable but under-documented research resource, and their provenance and limitations must be recorded if the data are to be reused responsibly.

Romania operates one such system. Its Mecanismul de Feedback al Pacientului (MFP) covers public hospitals and combines three policy functions: measurement of patient-reported hospital experience and satisfaction, public disclosure of hospital-level results, and identification of possible integrity incidents, particularly solicitation of money or gifts by medical personnel. The integrity component addresses informal payments, a persistent concern in Central and Eastern European health systems [4,5], and makes the MFP of the most visible administrative data sources on patient-reported solicitation in the region.

The mechanism's documentary history is more layered than the common "introduced in 2016" shorthand suggests. It was piloted in 2014 and first legislated in 2015 as a hospital-administered paper and online system [6,7]. Within the National Anti-Corruption Strategy 2016–2020 [8], it was redesigned as a centralized post-discharge process in which a national platform sends questionnaires by SMS or email independently of the hospital being evaluated [9,10]. A 2021 amendment replaced the data-protection notice without changing the questionnaire [11], and a November 2023 amendment replaced the ten-item instrument with an eight-item form and altered the integrity-reporting pathway [12]. The Ministry of Health publishes the resulting monthly aggregate workbooks on its institutional page and the national open-data portal, and the system remained active in 2026.

These workbooks are openly available, but they are difficult to reuse as published. Layouts differ across months, a large fraction of the legacy .xls files are structurally malformed, per-hospital detail is present in only a minority of releases, the questionnaire changed mid-series, and the files have never been consolidated with documented provenance. Prior research has nonetheless demonstrated its analytical potential: Mosca et al. analyzed reported solicitations and willingness to report over 2017–2021 [13], and Plopeanu used MFP-derived corruption indicators to study spatial and historical structure [14]. Both studies inherited the measurement limitations of the underlying survey and would have benefited from a harmonized, validated base dataset. Large national patient experience programs do publish open data—the United States' HCAHPS program releases hospital-level summary measure scores [15], England's NHS adult inpatient survey publishes trust-level benchmark results [16], and the OECD's Patient-Reported Indicator Surveys report internationally comparable indicators [17]—but these releases are periodic and summary-level. What is uncommon is the conjunction offered here: a monthly-cadence release of full item-by-option response distributions, resolved to named hospitals, carrying an integrity reporting item, spanning a decade, and redistributable under open terms. This conjunction is part of what makes documenting and validating this resource worthwhile.

This descriptor releases that base dataset. The contribution is not an inferential finding, but a reusable research resource: a harmonized national series, a provider release panel, and a separate legacy 2016 instrument, each built by a reproducible pipeline from unmodified official sources, with verified provenance and a documented validation trail. Harmonization here means a consistent tidy schema, verified provenance, and exact reconciliation across heterogeneous source layouts; it does not make the instruments commensurable in constructing meaning across the 2024 questionnaire revision or with the legacy paper-based instrument, which is why the three datasets are released separately rather than being merged. The release is designed to follow the FAIR principles of findable, accessible,

interoperable, and reusable data [18], as well as to be redistributed under open terms while respecting the public sector status and data-protection framing of the original [19].

The dataset is intended to support several distinct lines of secondary research that this descriptor deliberately does not pursue: longitudinal description of participation and satisfaction; channel comparison (SMS versus web) studies; hospital- and county-level heterogeneity analysis with appropriate shrinkage; and analysis of the integrity items as a respondent-reported signal rather than a prevalence estimate. By labeling the data components, documenting their provenance, and stating their hazards and opportunities up front, this descriptor lets later studies start from a transparent and shared foundation rather than re-deriving the cleaning and reconciliation each time.

2. Data Description

2.1. Overview of the Released Resource

The release comprises three processed datasets, a set of derived analysis outputs and figures, and a publication package of provenance and metadata files. The three datasets are deliberately kept separate because they are not mutually commensurable: the national series and the provider release panel are two views of the same modern SMS/web instrument, whereas the legacy 2016 dataset records a different, paper-based questionnaire that must not be concatenated with the modern series. Table 1 summarizes the components.

Table 1. Components of the released resource. Row and identity counts were verified against the processed files on 23 June 2026.

Component	File and Extent	Role
National monthly series	survey_combined.xlsx 4 sheets; 678 response-rate rows; 6004 national answer rows; 18-row question legend	National time series of answer counts and response rates by channel
Provider release panel	answers_by_hospital.csv 493,450 rows; 466 provider identities; 42 counties; 39 releases	Per-hospital answer counts where the source published hospital-level detail
Legacy 2016 instrument	legacy_2016_satisfaction.xlsx 3 sheets; 2067 indicator rows; 48 coverage rows	Standalone historical paper-based satisfaction instrument with demographic cross-tabs
Derived analysis outputs	data/analysis_outputs/ 40+ result tables (CSV/JSON) and 22 figures	Generated descriptive and modeling outputs and figures
Publication package	data/publication_package/ 10 metadata/manifest files	Checksummed source manifest, data dictionary, provenance log, item concordance, license and availability metadata

2.2. National Monthly Series (survey_combined.xlsx)

The principal deliverable integrates the modern feedback mechanism workbooks into a single national time series. It covers 113 observed monthly releases spanning December 2016–May 2026—a 114-month calendar window in which August 2024 is absent (Section 3.5). The workbook contains four sheets, described in Table 2.

Two response scales occur. Four-point ordinal satisfaction items use the set {Foarte nemulțumit, Nemulțumit, Mulțumit, Foarte mulțumit} (very dissatisfied, dissatisfied, satisfied, very satisfied), and binary items use {Da, Nu} (yes/no). The national_source field records the provenance of each national value: reported for months published as national aggregates (3976 rows) and summed_from_hospitals for months reconstructed by summing the per-hospital detail (2028 rows). Of the 113 observed releases, 74 were sourced from the national aggregate layout and 39 from the per-hospital layout.

Table 2. Schema of survey_combined.xlsx. The answers_national sheet is the analytical core; percentages are not stored because they are exactly recoverable from counts.

Sheet	Rows	Fields and Purpose
response_rate	678	Fields: month, stage, channel, respondents, rate_pct. National monthly response-rate statistics for initiated and completed questionnaires, reported as counts and rates by channel. The respondents field contains channel-specific questionnaire counts, not deduplicated persons; SMS and Web together can exceed Total (Section 4.1).
answers_national	6004	Fields: month, survey_version, question_no, question_text, answer_option, channel, count, national_source. Tidy national answer counts, with one row per month, question, answer option, and channel.
questions	18	Fields: survey_version, question_no, question_text. Legend of distinct item tuples used for cross-version alignment by item text.
notes	Per file	Contents: processing log and caveats. Detected source layout, processing status, integrated row count, and methodological notes for each file.

The survey_version field marks the instrument in force. In January 2024, the questionnaire comprised ten items (10Q); from February 2024, it comprised eight (8Q). The revision reworded and renumbered items, so question number is not a stable identifier across the boundary; the 18-row questions legend records every distinct (survey_version, question_no, question_text) tuple and is the recommended key for cross-version alignment by item text. Item text is preserved verbatim in Romanian in the legend, which is the stable alignment key; English glosses for the items discussed in this descriptor are given alongside them (for example, the integrity items in Section 4.2), and a full item-by-item Romanian–English questionnaire concordance is included with the deposited package in the Supplementary Materials. Table 3 condenses that concordance into the cross-version alignment map. The response-option sets are identical across the two instrument versions—the four-point ordinal satisfaction set and the binary Da/Nu set—so the 2024 revision changed the item wording, count, and numbering, but not the response scales. The English glosses in the legend, and the concordance are the authors’ translations; the verbatim Romanian text remains the authoritative alignment key.

Table 3. Cross-version item alignment, condensed from the deposited concordance (item_concordance_ro_en.csv). Scale: 4-pt = four-point ordinal satisfaction set; Y/N = Da/Nu.

Construct (English Gloss)	10Q Item	8Q Item	Scale	Alignment
Overall satisfaction with hospital services	Q1	Q1	4-pt	Aligned (wording shifts from “how satisfied” to “are you satisfied”)
Physician activity and involvement	Q2	—	4-pt	Merged into 8Q Q4; no clean continuation
Hospital cleanliness	Q3	Q2	4-pt	Aligned
Needed to buy medicines or supplies	Q4	Q3	Y/N	Aligned
Nurse activity and involvement	Q5	—	4-pt	Merged into 8Q Q4; no clean continuation
Medical staff activity and involvement	—	Q4	4-pt	Merges 10Q Q2 and Q5; not bridgeable to either separately
Clear explanations of diagnosis and treatment	Q6	Q5	Y/N	Aligned
Would recommend the hospital to someone close	Q7	Q6	Y/N	Aligned
Health better after discharge	Q8	Q7	Y/N	Aligned
Solicitation of money or gifts	Q9	—	Y/N	Standalone item; not comparable to the 8Q conditional item
Willingness to report solicitation	Q10	—	Y/N	Conditional wording, administered to all respondents reaching it; no 8Q equivalent
Combined conditional solicitation-reporting item	—	Q8	Y/N	Not comparable to 10Q Q9 or Q10

2.3. Provider Release Panel (*answers_by_hospital.csv*)

Where a source workbook published hospital-level detail, the per-hospital answer counts are released as a separate UTF-8 (BOM) CSV. This artifact contains 493,450 rows across the 39 observed releases, for which a per-hospital breakdown exists (January 2021–May 2026) covering 465 distinct hospital name strings and 466 county plus exact-name provider identities across 42 counties (Romania’s 41 counties plus Bucharest, i.e., complete territorial coverage). The 39 releases are not contiguous: hospital-level detail was published monthly from January 2021 through December 2022, for June and August 2023, and—after an interruption that spans the February 2024 instrument change—for May 2024, February 2025, and monthly from July 2025 through May 2026. The exact release list is recorded in the month-level provenance log. The artifact is distributed as CSV rather than as a spreadsheet sheet because its size exceeds the practical row capacity of a single worksheet. Table 4 lists the fields.

Table 4. Schema of *answers_by_hospital.csv*. Zero-count cells are omitted (absence denotes zero); this omission does not affect any aggregate sums.

Field	Type	Description
month	string (YYYY-MM)	Reporting month of the release
county	string	Two-letter Romanian county code
hospital	string	Exact hospital-name string as published
provider_id	string	Reproducible key county hospital; not an official facility identifier
question_no	integer	Item number within the instrument in force that month
question_text	string	Verbatim Romanian item text
answer_option	string	Response option (Da/Nu or four-point ordinal label)
answer_order	integer	Ordinal position of the response option
channel	string	SMS or Web
count	integer	Number of responses in the cell

The *provider_id* key is reproducible and stable for a given published name, but it is a string identity rather than an official facility code. It does not by itself prove legal institution continuity through renaming, merger, reorganization, or spelling variation; the accompanying *provider_identity_notes.md* documents the one normalized name duplicate group detected by a conservative diagnostic normalizer (three same-named pneumophthiology hospitals in different counties, counted as three distinct identities) and recommends linkage to an external registry before making claims about legal institutions over time.

2.4. Legacy 2016 Instrument (*legacy_2016_satisfaction.xlsx*)

A structurally distinct paper-based satisfaction questionnaire was administered in 2016 under the first (hospital-administered) regime. It is extracted into its own workbook from three source files (a semester-I 2016 report covering January–June and the September and October 2016 monthly reports), which share an identical 43-column layout. The workbook has three sheets: indicators (2067 tidy rows giving 53 satisfaction indicators with a demographic cross-tabulation by sex, age, education, and residence), coverage (48 rows of administration metrics and the respondent demographic profile at national level), and notes. Each indicator reports the share of patients “satisfied and very satisfied” (or who answered “yes”) as a numerator over a denominator; the underlying four-point distribution is not recoverable from these aggregates.

This dataset is not commensurable with the modern SMS/web series. It differs in collection channel (paper and online versus SMS and web), item set (53 fine-grained indicators versus 8–10 questions), outcome representation (a collapsed satisfaction share versus a full option distribution), and temporal granularity (a semester aggregate for the first half of 2016). It is provided for historical and baseline context only and must not

be concatenated with the national series or the provider panel. It is, however, the only component that carries respondent demographic detail, which the post-2016 centralized instrument does not collect.

2.5. Derived Analysis Outputs and Figures

The release includes a directory of generated descriptive and modeling outputs and a figures subdirectory containing 22 publication-quality figures, produced by the analysis script from the three processed datasets. These outputs support reuse and reproducibility, but are downstream of—and separable from—the core datasets. Their detailed interpretation belongs to companion analytical studies, not to this descriptor. Figure 1 is the one coverage and volume figure reproduced here, showing the monthly participation funnel from the implied published rate base through started and finalized questionnaires. All figures are supplied as PNG files and are regenerable at arbitrary resolutions from the deposited code; the figure reproduced here (Figure 1) is rendered at 600 dpi.

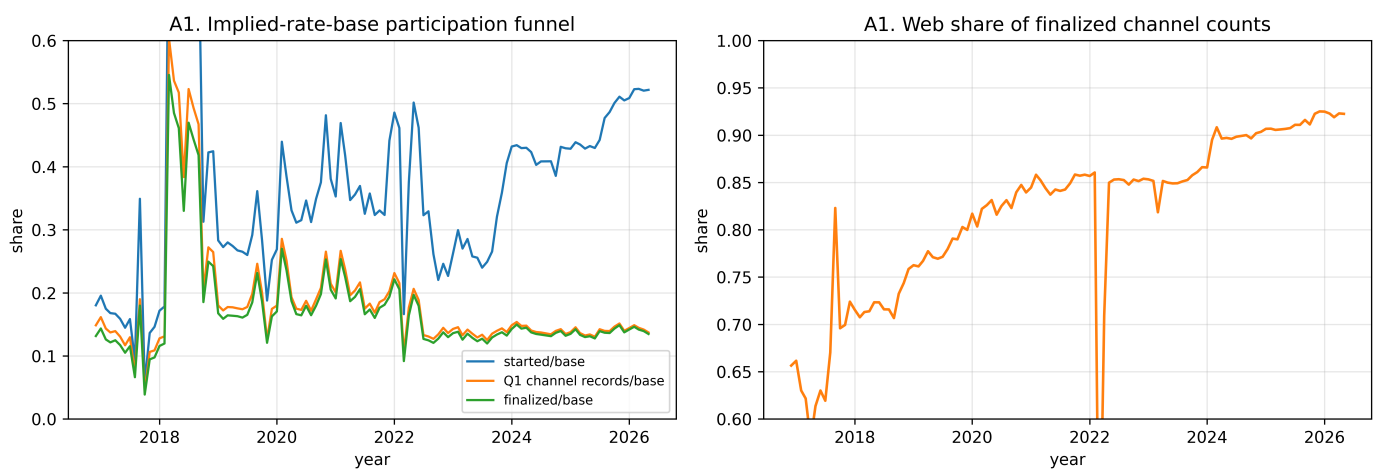


Figure 1. Monthly participation funnel for the national series (December 2016–May 2026). The figure traces the implied published rate base, started questionnaires, and finalized questionnaires over the observed releases. The implied rate base is an inferred denominator (Section 4.1), not a verified count of discharges, eligible patients, invitations, or successfully contacted patients.

2.6. Publication Package and Metadata

A publication package (data/publication_package/) accompanies the datasets and supports the provenance claims without relying on prose. It contains the following: a raw-source manifest of all 117 workbooks with file sizes and SHA-256 checksums, detected layouts, and source hints; a release manifest of files included in the archive; a machine-readable data dictionary covering every distributed table and variable; a month-level provenance and anomaly log; an analysis-output schema inventory; provider-identity notes; an item-by-item Romanian–English questionnaire concordance; a DataCite-style metadata scaffold (dataset_metadata.json); a data-availability statement; and a submission boilerplate. The package was generated using a dedicated builder script so that its contents are themselves reproducible from the processed data.

3. Methods

3.1. Data Source and Provenance

The primary data are monthly Microsoft Excel workbooks published by the Romanian Ministry of Health, each corresponding to one calendar month and named mecanism-feedback-patient-luna-YYYY-MM.xls (or .xlsx). The reports are disseminated through

two official channels: the Ministry of Health institutional page for the Patient Feedback Mechanism, maintained by its Integrity Service [20], and the national open-data portal data.gov.ro, where the reports are cataloged by year as CKAN datasets [21].

File authenticity was confirmed by re-downloading a stratified sample of twelve files—six from ms.ro and six from data.gov.ro, spanning 2018–2026, both .xls/.xlsx formats, and Layouts B and C—and verifying that the locally held copies were byte-for-byte identical to the published originals, including files distributed under variant filenames. Two checksum algorithms appear in the release for two distinct purposes: byte-for-byte source fidelity during corpus assembly and re-download was checked with MD5 digests, while SHA-256 digests of every source workbook are recorded in the archived raw_source_manifest.csv for long-term verification. MD5 is only used as a fidelity check against accidental corruption, not as a security control; any reuser can independently recompute the SHA-256 digests. Feedback is elicited through two response channels, SMS and web, and the monthly workbooks report response counts (and, in the national-aggregate variant, percentages) for a fixed questionnaire together with response rate statistics.

3.2. Source Workbook Layouts

The source workbooks are heterogeneous, with three layouts interleaved across the series rather than following a clean chronological succession. Layout A is the legacy 2016 centralizator: a single dense sheet reporting the paper-based instrument with demographic cross-tabulations, cataloged separately on the open-data portal. Three such files are present and are extracted into the legacy dataset rather than merged. Layout B is the national aggregate: a rata de raspuns (response-rate) sheet followed by one sheet per question, each a small national contingency table of counts and percentages by channel. Layout C is per-hospital: the same response-rate sheet, with each question sheet giving one row per reporting hospital (approximately 420 reporting institutions in a typical release, each prefixed by a two-letter county code) by answer option and channel. Across the full 39-release panel, these cumulate to 466 distinct provider identities (Section 2.3).

In Layouts B and C, the response-rate sheet reports two quantities, questionnaires initiated (chestionare incepute: the patient replied Da to the initial SMS or accessed the URL) and questionnaires completed (chestionare finalizate: the patient answered all questions), each as an absolute count and a rate, broken down by SMS, web, and total.

3.3. Processing Pipeline

Processing was performed in Python 3 using pandas for tabular manipulation, xlrd for legacy .xls files, and openpyxl for .xlsx files, and is reproducible from the original sources via build_combined.py and build_legacy_2016.py. The original workbooks were never modified. The pipeline comprised six steps:

1. File selection and deduplication: Source files were grouped by calendar month. One month (October 2025) was distributed in both .xls and .xlsx form; the two were verified to be data-identical across all sheets, and the redundant (and technically malformed) .xls copy was removed in favor of the .xlsx copy.
2. Recovery of malformed legacy files: A large fraction of the .xls files (all of 2021 and most of 2022–2026) raised a structural error under xlrd. These files were read losslessly with the ignore_workbook_corruption=True option, and the recovered contents were confirmed to be identical to the published files by MD5 checksum.
3. Layout detection: Each workbook was classified as Layout A, B, or C. Files lacking a rata de raspuns sheet were treated as Layout A and excluded from the modern series. Among the remainder, question sheets with five or more columns were classified as per-hospital (Layout C), and otherwise were classed as national aggregate (Layout B).

The detected layout and integrated row count of every file are recorded in the notes sheet; a misclassification would surface at parse time because per-hospital extraction requires county-prefixed hospital rows, and downstream because the summed-from-hospitals months are re-reconciled by the validation gate on every run.

4. Parsing: Sheets were parsed by semantic anchoring on row and column labels rather than fixed cell coordinates to accommodate variable blank spacer rows. Question text was only accepted when it began with the expected item number, preventing date or header labels from entering the question legend. Counts were coerced to integers and reported percentages parsed where present.
5. National aggregation: For the Layout C months, national figures were derived by summing response counts across all hospitals for each (question, answer option, channel) cell. The provenance of each national value is recorded in the `national_source` field.
6. Normalization to tidy form: All answer data were reshaped to long-form with one observation per (month, question, answer option, channel). Percentages were not retained, as counts are the common currency across layouts and percentages are exactly recoverable.

3.4. Technical Validation

The consolidated dataset was subjected to internal-consistency and external-fidelity checks:

- Pipeline-consistency reconciliation of granularity: For all 39 per-hospital releases, the national totals derived by summation were compared against an independent re-aggregation of the raw hospital rows (2028 month $\times$ question $\times$ option $\times$ channel groups); no discrepancies were found. This is an internal implementation check: the source workbooks for these months only publish the per-hospital detail and the response-rate sheet, with no separately reported national aggregates, so the check guards against extraction and aggregation error rather than against errors in the source itself.
- Internal arithmetic consistency: For Layout B months, counts reproduced the independently reported percentages within rounding. For every month, the total number of responses to each question did not exceed the number of initiated questionnaires (maximum observed ratio: 0.84), and answer totals were monotonically non-increasing across the item order, consistent with progressive respondent drop-off.
- External fidelity: All twelve files in the stratified re-download sample (six per official channel) were MD5-identical to the locally held copies, and national totals re-extracted from the freshly downloaded files matched the consolidated output exactly.
- Coverage: The local corpus was cross-checked against the complete open-data catalog; no published month was absent, and the integrated series contains 113 observed monthly releases across the December 2016–May 2026 span, with August 2024 absent.
- Question-legend hygiene: The final legend contains 18 unique (survey_version, question_no, question_text) rows; workbook month labels were excluded from the legend and answer table.
- Legacy dataset: For all 2067 extracted indicator rows, positive responses did not exceed the responses to each question, the recomputed ratio reproduced the source percentage within rounding (≤ 0.6 percentage points), and indicator labels were verified identical across the three source files.

All checks in this section establish fidelity to—and the internal consistency of—the published workbooks; they cannot detect errors made upstream by the publisher. A silent

source-side revision after corpus assembly would surface as a checksum mismatch on re-download rather than being adopted automatically.

3.5. Data Not Available and Anomalies

Four provenance facts must accompany any temporal reuse, and they are encoded in the month-level provenance log:

- August 2024 is missing: The published workbook is an empty template (all data cells are blank), and the data.gov.ro copy is MD5-identical to the Ministry copy and equally empty. This is a genuine gap in the source data and is excluded from both output tables.
- July and September 2024 are disruption-adjacent: A Ministry notice attributed temporary collection and reporting failures between 4 July and 6 September 2024 to IT infrastructure work at the National Health Insurance House (CNAS). The affected data transfer chain passed records from CNAS to the Special Telecommunications Service for questionnaire distribution, and the notice stated that recovery of data from the interval was being pursued [22]. The July and September 2024 releases are present, but incomplete.
- The monthly series begins in December 2016: No feedback mechanism reports are published for November 2016 or earlier; the September and October 2016 files use the distinct legacy instrument and are routed to the legacy dataset.
- Granularity is uneven: A per-hospital breakdown is only available for 39 of the 113 observed national releases; the remaining releases are national-resolution only.

3.6. Computational Environment

Processing used Python 3.12 with pinned package versions (pandas 3.0.0 [23], xlrd 2.0.2 [24], openpyxl 3.1.5 [25]; the full pin set is recorded in data/requirements.txt). The build, analysis, and validation steps are orchestrated by run_all.py and gated by validate_outputs.py, which asserts the primary record counts (113 observed national releases, 466 provider identities, 18 legend rows) and refuses to pass on mismatch. The derived analysis outputs and figures in data/analysis_outputs/ were generated from the three processed datasets by this same pinned environment and are tagged to the v1.0.1-data-descriptor release, so they are able to be regenerated from the deposited code. Source files were accessed over HTTPS from ms.ro and data.gov.ro, and integrity was verified by checksum against the locally held corpus.

4. User Notes

4.1. Known Reuse Hazards

The released data are aggregate administrative counts, and several properties constrain their valid reuse. Table 5 lists the principal hazards and the recommended handling; the first four are structural, and the last two concern denominators and channels.

The implied published rate base deserves emphasis because it anchors any participation analysis. The six monthly response rate rows imply an internally consistent denominator, but the public files do not define it institutionally. To indicate the range of values the response_rate table records rather than to report a trend, the observed extremes across the 113 months are as follows: implied base from about 5500 to about 371,800 (mean $\approx$ 160,300), started questionnaires from 3751 to 180,616, finalized questionnaires from 1805 to 55,675, monthly finalized rate from 3.9% to 54.5%, and conditional finalized-to-started ratio from 0.26 to 0.75. These unordered extremes are properties of the released counts, not population rates; their temporal arrangement is deliberately not narrated here, and interpreting it is left to companion analytical work. The channel overlap hazard can also be quantified: SMS plus

Web exceeds the Total in every one of the 113 observed months at both stages—typically by about 2.7% of the Total for started questionnaires (at most 3.6%) and by about 0.5% for finalized questionnaires (at most 1.2%)—a pattern consistent with a subset of patients engaging through both channels and being counted once per channel but only once in the Total. The per-month overlap is included in the derived outputs (A_funnel_timeseries.csv).

Table 5. Known reuse hazards and recommended handling.

Hazard	What It Is	Recommended Handling
Instrument break	The 10-item to 8-item revision at February 2024 reworded, merged, and renumbered items	Align items by text using the questions legend; do not pool the merged physician/nurse rating; mark the break on every trend
Uneven granularity	Hospital-level detail exists for 39 of 113 observed releases	Restrict hospital-level work to the 39-release provider panel; describe provider_id as a string identity
Aggregate-only records	No respondent-level rows, no free text, no modern-series demographics	Treat all analyses as ecological; no patient-level case-mix adjustment is possible
String provider identity	provider_id is county name, not an official code	Link to an external registry before claims about legal institutions over time
Inferred denominators	The implied published rate base is internally consistent but institutionally undefined	Use it as an inferred published rate base, not as discharges, invitations, eligible patients, or successful contacts
Channel overlap	For started and finalized counts, SMS + Web can exceed Total	Treat web shares as shares of channel-specific counts, not deduplicated proportions of individuals

4.2. A Reuse Opportunity: Provider- and County-Level Integrity Items

Beyond satisfaction, the provider release panel uniquely carries the mechanism’s integrity items at hospital and county resolution, which enables provider- and county-level analysis of patient-reported informal payment solicitation that the national aggregate releases cannot support. The ten-item era contributes a standalone solicitation item (Q9, “were you asked for money or gifts by doctors or nurses?”) and a willingness-to-report item (Q10, conditional in wording but administered to all respondents who reached it) across 26 provider releases (January 2021–August 2023); the eight-item era contributes a single combined item (Q8, likewise conditionally worded and administered to all respondents reaching it) across 13 provider releases (May 2024–May 2026). In total, the panel holds 13,828 provider release integrity observations.

This opportunity comes with a mandatory rare-event caveat. The provider release affirmative rate is approximately 1.9%, and 54.6% of provider releases report zero affirmative events. Single-release league tables are therefore not informative at the provider level, and any reuse should apply shrinkage (for example, beta-binomial empirical Bayes) and funnel presentation against uncertainty limits rather than ranking raw rates. The two integrity eras must also be kept separate: the eight-item conditional item (Q8) is not comparable to the ten-item standalone item (Q9) because the apparent post-2024 change in affirmative responses cannot be attributed to a real change in solicitation: the question structure and the reporting pathway changed at the same boundary. To illustrate what the integrity columns contain rather than to report a finding, in the national monthly series, the ten-item standalone solicitation item declines from 4.5% affirmative in its first ten-item month (December 2016) to 1.7% in its last (January 2024; monthly mean 2.5%), the national willingness item (Q10) averages 2.3% affirmative among all Q10 responses over the same months (its monthly response volume is roughly 98% of Q9’s, so the denominator is the full response base reaching the item, not the subset reporting solicitation), and, within the provider panel itself, the pooled provider release affirmative rate is approximately 1.9%, as noted

above; external sources report related but non-equivalent figures [26]. These are response shares among self-selected respondents, not population prevalence, and any substantive analysis of them belongs to dedicated studies. The released records are aggregate counts with no respondent-level attributes or demographics, and the respondent pool behind any cell is not externally enumerable, so the re-identification risk of these aggregates is low; the residual sensitivity of small-facility integrity counts is reputational rather than personal. This is a further reason—beyond sampling noise—why hospital-level integrity counts should be presented with uncertainty rather than as exact rates, and it applies with particular force to journalistic reuse, where the shrinkage caveat should travel with any republished figures.

4.3. Inappropriate Uses

Several analyses are explicitly not supported by these aggregate data, and reusers should avoid them: (i) Naive single-release hospital league tables: Provider rates from one release are dominated by sampling noise, and over half of provider release integrity cells are zero, so raw ranking is not defensible without shrinkage and uncertainty limits. (ii) Reading the integrity item as a corruption prevalence estimate: The affirmative shares are responses among self-selected post-discharge respondents, not a population rate. (iii) Pooling across the February 2024 instrument break or comparing the ten-item standalone solicitation item with the eight-item conditional item as if they measured the same thing. (iv) Treating the implied published rate base as a verified count of discharges, eligible patients, invitations, or successful contacts. (v) Person-level, case-mix-adjusted, or measurement-invariance analyses, which the aggregate-only records cannot support. These prohibitions complement the recommended handling column of Table 5.

4.4. FAIR Alignment, Access, and Reproduction

The release is designed for FAIR reuse: once the archive is deposited, it will be Findable through a persistent DOI and DataCite-style metadata and Accessible through an open repository without login or paywall (the development repository provides interim access); it is Interoperable through tidy CSV and XLSX tables documented by a machine-readable data dictionary and a consistent vocabulary, and Reusable through open licenses (CC BY 4.0 for data and documentation, MIT for code), complete provenance, and checksummed manifests. This is consistent with the data-sharing and transparency expectations the journal endorses.

The processed datasets, analysis code, generated outputs, documentation, and machine-readable manifests are deposited in an external repository under the all-versions DOI:10.5281/zenodo.21144263, which resolves to the latest release. Reuse of the underlying government content rests on Romania's public sector information regime: The reports are disseminated through the national open-data portal, whose reuse framework implements the EU Open Data Directive [27] as transposed by Law no. 179/2022 [28], and the portal catalog entries for these datasets declare a Creative Commons Attribution 4.0 license, while the ministry page itself declares none; this release nonetheless takes the conservative posture of asserting no ownership over—and applying no new license to—the originals (see the license matrix in the repository LICENSES.md). The original workbooks are not included in the archive; they remain publicly available from ms.ro and data.gov.ro, and are documented in raw_source_manifest.csv with filenames, sizes, SHA-256 checksums, detected layouts, and source channel hints, so the entire pipeline can be re-run end-to-end for as long as the sources remain available from the official channels. The checksums authenticate any copy—including copies recovered from third-party web archives or a future archival deposit of the raw corpus, which the portal's license declaration would permit—but they

cannot substitute for the files themselves; the processed datasets, which are archived, do not share this dependency. The archived release is a frozen snapshot: months published after May 2026 are outside the scope of v1.0.1-data-descriptor and would enter a future tagged release rather than modify this one. Reviewers and reusers can regenerate the artifacts from a fresh checkout with `build_combined.py`, `build_legacy_2016.py`, `run_all.py`, and `validate_outputs.py` under the pinned environment in `data/requirements.txt`.

5. Conclusions

This descriptor releases a harmonized, validated, openly licensed version of a decade of Romania's national patient experience feedback, packaged so that other researchers can trust and reuse it. The resource consists of a national monthly series, a provider release panel, and a separate legacy 2016 instrument, each built in a reproducible manner from unmodified official sources with byte-for-byte provenance verification and exact internal reconciliation. Its value to later work depends on clear documentation of what exists and how it was built, which this descriptor provides, and on honest labeling of the breaks, gaps, and aggregation limits that govern safe reuse. By separating the resource contribution from any inferential claim, the descriptor establishes a shared foundation on which questions about satisfaction, response channels, hospital and county variation, and patient-reported integrity signals can be investigated with appropriate methods and caveats.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/data11080185/s1>. The supplementary file is: `patient_feedback_mechanism-v1.0.1-data-descriptor.zip`. It contains the three processed datasets; reproducible analysis code and generated outputs; the data dictionary; source and release manifests with SHA-256 checksums; the provenance and anomaly log; the schema inventory; provider-identity notes; the Romanian–English questionnaire concordance; dataset metadata; and all 22 generated figures.

Author Contributions: Conceptualization, D.V.V.; methodology, D.V.V. and Ș.A.V.; software, D.V.V.; validation, D.V.V. and Ș.A.V.; data curation, D.V.V.; writing—original draft preparation, D.V.V.; writing—review and editing, D.V.V. and Ș.A.V.; visualization, D.V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The processed datasets, analysis code, generated outputs, documentation, and machine-readable manifests for version v1.0.1-data-descriptor are openly available in Zenodo under the all-versions <https://doi.org/10.5281/zenodo.21144263>, which resolves to the latest release. These data were derived from publicly available official sources: the Romanian Ministry of Health Patient Feedback Mechanism page (<https://ms.ro>, accessed on 23 July 2026) and the Romanian national open-data portal (<https://data.gov.ro>, accessed on 23 July 2026). The original government workbooks are not redistributed in the archive but are documented with filenames, sizes, and SHA-256 checksums, so the pipeline can be reproduced while those sources remain available and any recovered copy can be authenticated.

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Anthropic; Claude Opus 4-series models, accessed via the Claude Code command-line interface) in June–July 2026 to consolidate author-supplied research materials, to assist with developing and documenting the data-build and validation code, and for language drafting. The authors reviewed and edited all output, verified every reported count against the released datasets, and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Doyle, C.; Lennox, L.; Bell, D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* **2013**, *3*, e001570. [CrossRef] [PubMed]
2. Navarro, S.; Ochoa, C.Y.; Chan, E.; Du, S.; Farias, A.J. Will improvements in patient experience with care impact clinical and quality-of-care outcomes? A systematic review. *Med. Care* **2021**, *59*, 843–856. [CrossRef] [PubMed]
3. Bull, C.; Byrnes, J.; Hettiarachchi, R.; Downes, M. A systematic review of the validity and reliability of patient-reported experience measures. *Health Serv. Res.* **2019**, *54*, 1023–1035. [CrossRef] [PubMed]
4. Chereceș, R.M.; Ungureanu, M.I.; Sandu, P.; Rus, I.A. Defining informal payments in healthcare: A systematic review. *Health Policy* **2013**, *110*, 105–114. [CrossRef]
5. Horodnic, A.V. Trends in informal payments by patients in Europe: A public health policy approach. *Front. Public Health* **2021**, *9*, 780337. [CrossRef] [PubMed]
6. Ministry of Health of Romania. Ordinul nr. 146 din 11 Februarie 2015 Privind Aprobarea Implementării Mecanismului de Feedback al Pacientului în Spitalele Publice. 2015. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/165724> (accessed on 21 June 2026).
7. Ministry of Health of Romania. Ordinul nr. 145 din 11 Februarie 2015 Pentru Aprobarea Componentei și a Atribuțiilor Consiliului Etic. 2015. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/165944> (accessed on 21 June 2026).
8. Government of Romania. Hotărârea nr. 583 din 10 August 2016 Privind Aprobarea Strategiei Naționale Anticorupție pe Perioada 2016–2020. 2016. Available online: <https://legislatie.just.ro/Public/DetaliuDocumentAfis/181123> (accessed on 21 June 2026).
9. Ministry of Health of Romania. Ordinul nr. 1.501 din 19 Decembrie 2016 Privind Aprobarea Implementării Mecanismului de Feedback al Pacientului în Spitalele Publice. 2016. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/185535> (accessed on 21 June 2026).
10. Ministry of Health of Romania. Ordinul nr. 1.502 din 19 Decembrie 2016 Pentru Aprobarea Componentei și a Atribuțiilor Consiliului Etic. 2016. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/185429> (accessed on 21 June 2026).
11. Ministry of Health of Romania. Ordinul nr. 1.399 din 30 Iulie 2021 Pentru Modificarea Anexei nr. 2 la Ordinul Ministrului Sănătății nr. 1.501/2016. 2021. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/245224> (accessed on 21 June 2026).
12. Ministry of Health of Romania. Ordinul nr. 3.722 din 2 Noiembrie 2023 Pentru Modificarea Ordinului Ministrului Sănătății nr. 1.501/2016. 2023. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/276239> (accessed on 21 June 2026).
13. Mosca, I.; Radu, C.; Strilciuc, Ș.; Ungureanu, M.-I. A five-year (2017–2021) time series evaluation of patient-reported informal healthcare payments in Romania. *J. Med. Life* **2023**, *16*, 387–393. [CrossRef] [PubMed]
14. Plopeanu, A.-P. Historical origins of corruption in the Romanian public health system—Path dependency and contagion effect. *Health Econ. Rev.* **2024**, *14*, 63. [CrossRef] [PubMed]
15. Centers for Medicare & Medicaid Services. HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems. Available online: <https://hcahpsonline.org> (accessed on 2 July 2026).
16. Care Quality Commission. NHS Adult Inpatient Survey. Available online: <https://www.cqc.org.uk/publications/surveys/adult-inpatient-survey> (accessed on 2 July 2026).
17. OECD. Patient-Reported Indicator Surveys (PaRIS). Available online: <https://www.oecd.org/health/paris/> (accessed on 2 July 2026).
18. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
19. European Parliament and Council of the European Union. Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Off. J. Eur. Union* **2016**, *L 119*, 1–88.
20. Ministry of Health of Romania. Mecanismul de Feedback al Pacientului (Results Page). 2026. Available online: <https://ms.ro/ro/minister/organizare/structura-organizatorica/serviciul-de-integritate/mecanismul-de-feedback-al-pacientului/> (accessed on 21 June 2026).
21. Ministry of Health of Romania. Mecanismul de Feedback al Pacientului 2026 (One CKAN Dataset per Year). Data.gov.ro: 2026. Available online: <https://data.gov.ro/dataset/mecanismul-de-feedback-al-pacientului-2026> (accessed on 21 June 2026).
22. Ministry of Health of Romania. Mecanismul de Feed-Back al Pacientului—2024—August (Dataset Notice). Data.gov.ro: 2024. Available online: <https://data.gov.ro/dataset/mecanismul-de-feed-back-al-pacientului-2023/resource/075723b9-24cc-4257-9b89-84c645816670> (accessed on 21 June 2026).
23. The Pandas Development Team. *Pandas-dev/Pandas: Pandas*, version 3.0.0; Zenodo: Geneva, Switzerland.
24. xlrd Developers. xlrd: Library for Developers to Extract Data from Microsoft Excel Legacy Spreadsheet Files, version 2.0.2. Available online: <https://pypi.org/project/xlrd/> (accessed on 2 July 2026).

25. openpyxl Developers. openpyxl: A Python Library to Read/Write Excel 2010 xlsx/xlsm Files, version 3.1.5. Available online: <https://pypi.org/project/openpyxl/> (accessed on 2 July 2026).
26. OECD. *OECD Reviews of Health Systems: Romania 2025*; OECD Publishing: Paris, France, 2025.
27. European Parliament and Council of the European Union. Directive (EU) 2019/1024 of 20 June 2019 on open data and the re-use of public sector information. *Off. J. Eur. Union* **2019**, L 172, 56–83.
28. Parliament of Romania. Legea nr. 179 din 9 Iunie 2022 Privind Datele Deschise și Reutilizarea Informațiilor din Sectorul Public (Law no. 179/2022 on Open Data and the Re-use of Public-Sector Information). 2022. Available online: <https://legislatie.just.ro/Public/DetaliiDocumentAfis/256414> (accessed on 2 July 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.