

Chile-ED-Resp: A Curated and Reproducible Weekly Hospital Dataset for Respiratory Emergency-Demand Forecasting in Chile

Wilman Balcázar-Quimi ¹, Carla Angulo ^{2,*} , Cristian Cornejo-Gaete ³, Brandon Salinas-Neira ², Jorge Fernández-Bastías ², Miguel Tupac-Yupanqui ⁴ and Cristian Vidal-Silva ^{5,*}

¹ Facultad de Salud y Servicios Sociales, Universidad Estatal de Milagro, Milagro 092301, Ecuador; lwbalcazarq@unemi.edu.ec

² Escuela de Ciencias Digitales y Tecnologías Avanzadas, Universidad Bernardo O'Higgins, Santiago 8370846, Chile; brandon.salinas@pregrado.ubo.cl (B.S.-N.); jorge.fernandez@pregrado.ubo.cl (J.F.-B.)

³ Facultad de Ingeniería, Ciencia y Tecnología, Universidad Bernardo O'Higgins, Santiago 8370846, Chile; cristian.cornejo@ubo.cl

⁴ EAP Ingeniería de Sistemas e Informática, Universidad Continental, Huancayo 12000, Peru; mtupac@continental.edu.pe

⁵ Facultad de Ingeniería y Negocios, Universidad de Las Américas, Manuel Montt 948, Providencia, Santiago 7500975, Chile

* Correspondence: carla.angulo@ubo.cl (C.A.); cristian.vidal.silva@edu.udla.cl (C.V.-S.)

Abstract

Respiratory emergencies create recurrent seasonal pressure on hospital services, yet reproducible forecasting research is often hindered by the effort required to retrieve, interpret, harmonize, and document public administrative data. Chile-ED-Resp is a curated secondary dataset and reproducible processing pipeline derived from the official Chilean Ministry of Health/DEIS dataset “Atenciones de urgencias de causas respiratorias por semana epidemiológica”, produced by the Sistema de Atención Diaria de Urgencias (SADU). Version 1.1.0 uses the closed 2022–2025 period and was built from the official Parquet snapshot retrieved on 30 August 2026. After restriction to hospitals and hospital emergency units (UEH), the curated long artifact contains 449,844 rows and 32 variables, representing 180 hospitals and 12 official respiratory cause labels. A complementary forecasting artifact contains 37,620 hospital-week rows and 38 variables on a fixed 209-position statistical-week calendar (52 weeks in 2022–2024 and 53 in 2025); 178 of 180 hospitals (98.9%) satisfy the predefined $\geq 95\%$ target-coverage criterion. Acute upper respiratory infection is retained under its official Spanish label and acronym, IRA Alta. Missing source records remain missing and are distinguished from explicitly reported zeros. Pre-computed target-derived predictors use outcomes available through the preceding week and are intended for rolling-origin one-week-ahead evaluation, not an unqualified fixed-origin forecast of the entire test year. Reproducibility is supported through source meta-data, SHA-256 checksums, machine-readable descriptive audits, 15 validation checks, unit tests, and an example that recomputes predictors at each forecast origin. The resource provides a transparent foundation for forecasting, benchmarking, health-informatics teaching, and secondary analyses while preserving source provenance and licensing conditions.

Dataset: Chile-ED-Resp, version 1.1.0. Creator: Balcázar-Quimi, W.; Angulo, C.; Cornejo-Gaete, C.; Salinas-Neira, B.; Fernández-Bastías, J.; Tupac-Yupanqui, M.; Vidal-Silva, C. Publisher/location: Zenodo. Version-specific DOI: <https://doi.org/10.5281/zenodo.22209795>. Development repository: <https://github.com/cvidalmsu/chile-ed-resp-benchmark> (accessed on 2 September 2026).



Academic Editors: Daniel Rojas-Valverde, Aldo Alfonso Vasquez-Bonilla and Rodrigo Yáñez-Sepúlveda

Received: 8 August 2026

Revised: 1 September 2026

Accepted: 3 September 2026

Published: 7 September 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.

Dataset License: Curated data adaptations: Creative Commons Attribution-NonCommercial 2.0 Generic (CC BY-NC 2.0), requiring non-commercial use, source attribution, identification of changes, and retention of the license notice; software: MIT License.

Keywords: emergency department; respiratory illness; hospital demand; time series; forecasting; data curation; DEIS; SADU; reproducibility; FAIR data; Chile

1. Summary

Hospital emergency departments must absorb demand that changes with epidemiological conditions, season, age structure, and local service organization. Respiratory illness is particularly relevant because recurrent winter epidemics can concentrate consultations and hospitalizations within a limited number of weeks. Forecasting approaches have consequently been used to anticipate emergency attendance, bed demand, and severe respiratory hospitalizations [1–4]. Methodological work on emergency-department demand forecasting spans early systematic reviews of statistical models [5] through unified and multivariable forecasting pipelines for hourly and daily arrivals and occupancy [6–10]. Recent systematic reviews show both the growing use of machine-learning methods in emergency-demand prediction and persistent challenges related to external validation, generalizability, and operational integration [11,12]. Pathogen-specific forecasting systems have likewise targeted influenza, respiratory syncytial virus, and combined winter respiratory pressure on hospitals [13–16]. In Chile, empirical analyses have documented temporal variation in emergency consultations and pronounced respiratory-season effects [17,18]. Complementary Chilean clinical guidance and recent pediatric commentary describe the evolving respiratory-syncytial-virus and severe acute respiratory-infection context relevant to seasonal preparedness [19,20].

Within this setting, the Chilean Ministry of Health makes weekly respiratory emergency data publicly available through the Departamento de Estadísticas e Información de Salud (DEIS). The source dataset, produced from the Sistema de Atención Diaria de Urgencias (SADU), reports consultations and hospitalizations by respiratory cause, establishment, statistical week, and age group [21]. Although the source is highly valuable, downstream forecasting studies must repeatedly solve the same data-engineering tasks: retrieving a large Parquet file, harmonizing field names and types, isolating hospital emergency units, separating consultations from hospitalization rows, checking temporal coverage, pivoting causes to hospital-week features, constructing lagged predictors without leakage, and documenting provenance.

From a data-publication perspective, the principal contribution of Chile-ED-Resp is therefore not a new predictive algorithm. Instead, the resource provides a transparent and reusable research layer over the official source. The design follows the FAIR principles of findability, accessibility, interoperability, and reusability [22], including implementation practices such as machine-readable metadata, versioned provenance, and automated validation [23–28]. Table 1 now compares the release with the closest Chilean and Latin American emergency or respiratory resources rather than with datasets from unrelated domains.

Accordingly, Chile-ED-Resp releases two complementary representations of the same official 2022–2025 respiratory emergency information. The first is a curated long-format table that preserves all qualifying hospital emergency-unit records and official count values. The second is a hospital-week forecasting artifact that reorganizes selected consultation causes and adds documented deterministic features. An explicit coverage criterion identi-

fies establishments suitable for benchmarking rather than hard-coding a small set of hospitals. The repository records dimensions, inventories, descriptive audits, source checksum, retrieval time, and validation results, allowing later versions to remain auditable if the upstream portal revises historical records.

Table 1. Closest Chilean and Latin American emergency or respiratory data resources and the distinct scope of Chile-ED-Resp.

Resource	Unit and Scope	Preparation for Reuse	Difference from Chile-ED-Resp
DEIS weekly respiratory-emergency source, Chile [21]	Establishment-week counts by respiratory cause and age group	Mutable official Parquet/XLSX resource and data dictionary	Direct upstream source; not versioned as a forecasting benchmark
DEIS public-network emergency data, Chile [29]	Annual emergency-attendance files derived from DAU records	Year-specific ZIP files and official dictionary	Broader emergency scope; no curated fixed hospital-week respiratory panel
SIVEP-Gripe, Brazil [30]	Case-level severe acute respiratory syndrome surveillance	Public annual microdata and documentation	Clinical surveillance cases rather than hospital-level emergency-demand series
PAHO respiratory-virus open data [31]	Country/territory surveillance indicators for the Americas	Regional harmonized portal	Country-level surveillance rather than Chilean hospital-week demand
Chile-ED-Resp (this work)	180 hospitals on a fixed 209-week calendar	Versioned long and forecasting artifacts; checksums, audits, tests, and rolling-origin example	Reproducible hospital-level demand benchmark with explicit zero/missing semantics

2. Data Description

2.1. Official Source and Scope

The upstream resource is the government dataset “Atenciones de urgencias de causas respiratorias por semana epidemiológica”, published through datos.gob.cl by the Chilean Ministry of Health/DEIS [21]. The active Parquet resource has the identifier `ae6c9887-106d-4e98-8875-40bf2b836041`. At the time this descriptor was prepared, the portal identified the resource under a Creative Commons Non-Commercial license; the linked terms correspond to the Creative Commons Attribution-NonCommercial 2.0 Generic license (CC BY-NC 2.0) [21,32]. The source field is named `SemanaEstadistica`; consequently, the manuscript uses the term statistical week for the analytical coordinate and retains epidemiological week only when reproducing the official Spanish dataset title. Because the current reporting year is continuously updated, the curated release intentionally uses the closed four-year interval 2022–2025. Restricting the versioned release to completed years avoids treating a partially reported current week as a genuine demand decline. This concern for completeness and timeliness mirrors broader open-government-data quality frameworks [33,34].

The source schema contains 25 variables after normalization to snake-case names. These fields describe establishment, region, municipality, health service, establishment type, administrative dependency, level of care, emergency-care type, coordinates, complexity level, year, statistical week, cause code/order, cause label, total count, and five age-specific count fields. Table 2 summarizes the variable families retained from the source. The official DEIS data dictionary documents the source schema and field definitions used to map the Parquet resource into the curated release [35].

Table 2. Variable families retained from the official DEIS/SADU source. The complete machine-readable dictionary is distributed as `metadata/data_dictionary.csv`.

Family	Main Fields	Interpretation
Establishment	establecimiento_codigo, establecimiento_glosa, tipo_establecimiento, nivel_atencion, nivel_complejidad	Official identifiers and service characteristics.
Geography	region_*, comuna_*, latitud, longitud	Administrative and geographic descriptors.
Health service	servicio_salud_*, dependencia_administrativa, tipo_urgencia	Network and emergency-unit descriptors.
Time	anio, semana_estadistica	Official DEIS statistical year and week.
Respiratory event	orden_causa, causa, num_total	Cause label and total consultations or hospitalizations in the stratum.
Age strata	num_menor1anio, num1a4anios, num5a14anios, num15a64anios, num65o_mas	Official age-specific counts.

The principal dimensions, coverage results, and validation outcomes of Version 1.1.0 are summarized in Table 3.

Table 3. Version 1.1.0 release snapshot generated from the official DEIS/SADU resource.

Release Property	Value
Official source snapshot	Retrieved 30 August 2026 at 23:38:36 UTC; 3,629,980 rows; 67,929,078 bytes; resource ID ae6c9887-106d-4e98-8875-40bf2b836041.
Curated long artifact	449,844 rows $\times$ 32 variables; 180 hospitals; 12 official cause labels.
Forecasting artifact	37,620 rows $\times$ 38 variables; 180 hospitals $\times$ 209 expected statistical-week positions.
Benchmark coverage	178/180 hospitals (98.9%) satisfy $C_h \geq 0.95$; 172 have complete 209-week target coverage.
Validation	15 PASS, 0 WARN, 0 FAIL reported validation statuses.
Source integrity	SHA-256 recorded in <code>metadata/source_manifest.json</code> and fixed for the archived release.

2.2. Released Artifacts

Version 1.1.0 generates the artifacts summarized in Table 4. The long artifact contains 449,844 rows and 32 variables, whereas the forecasting artifact contains 37,620 rows and 38 variables. Both represent 180 hospitals; the long table retains 12 official respiratory cause labels. Exact dimensions and descriptive audits are stored with the release so that a mutable government source can be traced to the precise snapshot used for the build.

Table 4. Principal files created by the reproducible release pipeline.

File	Format	Role
data/chile_ed_resp_hospital_long.parquet data/chile_ed_resp_hospital_long.csv.gz	Parquet; CSV.gz	Official hospital emergency-unit records plus provenance and quality-control fields.

Table 4. *Cont.*

File	Format	Role
data/chile_ed_resp_ ira_alta_forecasting.parquet data/chile_ed_resp_ ira_alta_forecasting.csv	Parquet; CSV	Fixed-grid hospital-week consultation features and deterministic predictors.
metadata/source_ manifest.json	JSON	Retrieval time, identifiers, source URL, byte size, SHA-256, and license.
metadata/filter_audit.csv metadata/cause_mapping.csv	CSV	Filtering counts and auditable cause-to-feature mapping.
metadata/variable_ completeness.csv metadata/forecast_ variable_coverage.csv	CSV	Missing/zero frequencies and hospital-level coverage for every forecasting cause.
metadata/target_summary.csv metadata/temporal_summary.csv	CSV	Distributional and week-level IRA Alta summaries.
results/validation_ checks.csv	CSV	Machine-readable structural and quality-control results, including records examined and flagged.
LICENSE-DATA CITATION.cff .zenodo.json	Text; YAML; JSON	Redistribution terms, citation metadata, and permanent-archive deposit metadata.

2.3. Curated Long Table

The long table restricts the official source to the years 2022–2025, establishments classified as Hospital, and records whose emergency type contains the hospital-emergency designation UEH. Table 5 reports the row count after every filtering step. The final 449,844 rows represent 180 hospitals and 12 official respiratory cause labels; no cause is removed from the long artifact. Consultation codes 3–11 and hospitalization codes 33–35 are identified from the official `orden_causa/causa` combinations and stored in `measure_type`. This order-aware rule prevents hospitalization totals from entering consultation predictors.

Table 5. Sequential filtering audit for Version 1.1.0. Counts are also distributed in `metadata/filter_audit.csv`.

Step	Retained Records	Rows Retained	Rows Removed at Step
0	Official Parquet snapshot	3,629,980	Not applicable
1	Statistical years 2022–2025	1,547,292	2,082,688
2	<code>tipo_establecimiento = Hospital</code>	455,784	1,091,508
3	<code>tipo_urgencia</code> contains UEH	449,844	5940

Beyond normalization of column names and coordinate types, official count fields are retained unchanged. Derived fields include a deterministic row identifier, `statistical_period`, the fixed source-resource identifier, the sum of the five official age strata, the difference between that sum and `num_total`, and a quality flag. All six official count variables are complete in the selected long table. Any future missing source value remains missing rather than being imputed. Such treatment supports transparent downstream handling of administrative health data [36].

2.4. Forecasting-Ready Hospital-Week Table

For reuse in time-series forecasting, consultation rows are mapped to canonical variables using the auditable rules in Table 6. The official label *IRA Alta (J00–J06)* is translated

as *acute upper respiratory infection*; the Spanish label and acronym IRA Alta are preserved in the data and manuscript because they are the official source terminology. Aggregate consultation categories, crisis obstructiva bronquial, and all hospitalization rows are retained in the long artifact but excluded from the forecasting features. No source value is reinterpreted during the pivot.

Table 6. Exact cause-by-cause mapping from original DEIS/SADU labels to forecasting fields. Word-ing and capitalization reproduce the source; `metadata/cause_mapping.csv` preserves the verbatim strings, including spacing and punctuation.

ID	Official Spanish Label	Measure	Forecast Field
3	TOTAL CAUSA SISTEMA RESPIRATORIO (J00–J98)	Consult.	Excluded aggregate
4	IRA Alta (J00–J06)	Consult.	<code>ira_alta</code>
5	Influenza (J09–J11)	Consult.	<code>influenza</code>
6	Neumonía (J12–J18)	Consult.	<code>pneumonia</code>
7	Bronquitis/bronquiolitis aguda (J20–J21)	Consult.	<code>bronchitis_</code> <code>bronchiolitis</code>
8	Crisis obstructiva bronquial (J40–J46)	Consult.	Excluded non-target category
9	Otra causa respiratoria no contenidas en las categorías anteriores (J22; J30–J39, J47, J60–J98)	Consult.	<code>other_</code> <code>respiratory</code>
10	TOTAL ATENCIONES POR COVID-19 Virus no Identificado U07.2	Consult.	<code>covid19_</code> <code>unidentified</code>
11	TOTAL ATENCIONES POR COVID-19, Virus Identificado U07.1	Consult.	<code>covid19_</code> <code>identified</code>
33	- Causas sistema respiratorio (J00–J98)	Hosp.	Excluded hospitalization
34	- Por COVID-19, virus no identificado U07.2	Hosp.	Excluded hospitalization
35	- Por COVID-19, virus identificado U07.1	Hosp.	Excluded hospitalization

The mapping is order-aware rather than a fuzzy text substitution: only consultation orders 4, 5, 6, 7, 9, 10, and 11 receive canonical forecasting fields. Consequently, labels that contain similar respiratory or COVID-19 wording but belong to hospitalization orders 33–35 cannot enter the forecasting variables.

An absent hospital–week–cause record is interpreted as unavailable reporting, not as zero events. Explicit zeros are values for which the official source contains a record with `num_total` = 0. Accordingly, the pivot retains absent records as NA and preserves recorded zeros as 0; these states are never treated as interchangeable. Table 7 shows that every forecasting cause has 133 missing grid positions (0.35%), while zero frequencies vary materially by cause.

A mandatory interpretation rule applies to the released tables: NA means that no qualifying source record was available for a given hospital–week–cause combination, whereas 0 means that a qualifying official row explicitly reported zero events. Users must not replace NA with 0 unless an external reporting model provides a justified basis for that recoding.

Hospital selection for benchmarking is coverage-based. The denominator is the fixed 209-position release calendar, not the set of weeks in which IRA Alta happens to appear somewhere nationally. Of the 180 hospitals, 172 report IRA Alta in all 209 positions; six additional hospitals remain above the 95% threshold with 204–208 observed target weeks. Consequently, 178 hospitals (98.9%) receive `recommended_for_benchmark` = True. The other two hospitals have 178/209 (85.2%) and 121/209 (57.9%) observed positions. Across the complete 180×209 grid, 37,487 of 37,620 IRA Alta values are observed (99.65% com-

pleteness). Within the recommended cohort, 37,188 of 37,202 values are observed (99.96%), leaving 14 missing targets. Table 8 summarizes the distribution.

Table 7. Missingness and explicit-zero frequencies for released count variables. Zero percentages use non-missing observations as the denominator.

Artifact	Variable	Missing, <i>n</i> (%)	Explicit Zero, <i>n</i>	Zero Among Observed (%)
Long	num_total	0 (0.00)	153,394	34.10
Long	num_menor1anio	0 (0.00)	320,982	71.35
Long	num1a4anios	0 (0.00)	280,005	62.24
Long	num5a14anios	0 (0.00)	263,106	58.49
Long	num15a64anios	0 (0.00)	199,422	44.33
Long	num65o_mas	0 (0.00)	222,460	49.45
Forecast	ira_alta	133 (0.35)	625	1.67
Forecast	influenza	133 (0.35)	16,091	42.92
Forecast	pneumonia	133 (0.35)	2679	7.15
Forecast	bronchitis_ bronchiolitis	133 (0.35)	1934	5.16
Forecast	other_respiratory	133 (0.35)	3529	9.41
Forecast	covid19_ unidentified	133 (0.35)	23,867	63.67
Forecast	covid19_ identified	133 (0.35)	23,274	62.09

Table 8. Distribution of IRA Alta temporal coverage C_h across the 180 hospitals in Version 1.1.0.

Coverage Bin (C_h)	Hospitals	Share of 180	Benchmark Status
209/209 target weeks (100%)	172	95.6%	Recommended
204–208/209 target weeks (97.6–99.5%)	6	3.3%	Recommended
178/209 and 121/209 target weeks (85.2% and 57.9%)	2	1.1%	Not recommended
Total (all establishments retained)	180	100%	178/180 (98.9%) recommended

Table 9 extends the hospital-level coverage audit to every forecasting cause. All seven variables share the same missingness mask because the selected official cause rows occur together at the retained hospital–week coordinates; their explicit-zero frequencies nevertheless differ, as shown in Table 7. The complete machine-readable audit is distributed as `metadata/forecast_variable_coverage.csv`.

The 180 hospitals cover all 16 Chilean regions (Table 10). Hospital–week metadata are joined on hospital, statistical year, and statistical week, so a latest record is no longer propagated backward. A separate stability audit found no within-hospital changes across the 12 released metadata fields during 2022–2025, but the time-specific implementation avoids relying on that empirical stability.

IRA Alta has 37,487 observed hospital–week values, a mean of 76.63 (SD 73.14), median 58 (interquartile range 29–102), range 0–4415, and aggregate total 2,872,473. Table 11 shows the year-specific distribution and national weekly peak. The peak shifts between statistical weeks 20 and 43 across years, illustrating the temporal heterogeneity that motivates year-aware validation.

Table 9. Missingness and hospital-level coverage for every forecasting cause on the $180 \times 209 = 37,620$ fixed grid.

Variable	Observed, <i>n</i>	Missing, <i>n</i> (%)	Complete Hospitals (209/209)	Hospital Coverage Groups
ira_alta	37,487	133 (0.35)	172	6 with 204–208; 2 below 199
influenza	37,487	133 (0.35)	172	6 with 204–208; 2 below 199
pneumonia	37,487	133 (0.35)	172	6 with 204–208; 2 below 199
bronchitis_ bronchiolitis	37,487	133 (0.35)	172	6 with 204–208; 2 below 199
other_respiratory	37,487	133 (0.35)	172	6 with 204–208; 2 below 199
covid19_ unidentified	37,487	133 (0.35)	172	6 with 204–208; 2 below 199
covid19_ identified	37,487	133 (0.35)	172	6 with 204–208; 2 below 199

Table 10. Hospitals by region in Version 1.1.0.

Region	Hospitals	Region	Hospitals
Metropolitana de Santiago	24	La Araucanía	23
Los Lagos	20	Biobío	20
Valparaíso	19	O’Higgins	15
Maule	13	Coquimbo	9
Los Ríos	9	Ñuble	7
Antofagasta	5	Atacama	5
Aysén	5	Magallanes	4
Arica y Parinacota	1	Tarapacá	1

Table 11. Distribution and temporal summary of acute upper respiratory infection (official label: IRA Alta).

Period	Observed Missing	Zeros	Mean SD	Median IQR	Peak Week (Total)
2022	9352/8	163	86.38 ± 95.13	61 [28–116]	W43 (26,615)
2023	9298/62	151	80.11 ± 70.31	61 [32–106]	W20 (26,279)
2024	9307/53	161	75.24 ± 63.39	59 [33–101]	W20 (26,031)
2025	9530/10	150	65.01 ± 56.35	51 [26–88]	W43 (16,280)
Overall	37,487/133	625	76.63 ± 73.14	58 [29–102]	2022-W43 (26,615)

Temporal features are constructed only after sorting by hospital and continuous `time_index`. The release provides one-, two-, and four-week lags, `ira_alta_prev_year_same_week`, trailing four- and 12-week means, and within-year sine/cosine encodings. A chronological split labels 2022–2024 as training and 2025 as testing. The information available for a one-week-ahead origin t comprises target observations through $t - 1$; therefore, the stored target-derived predictors support rolling-origin one-step evaluation. They must not be used unchanged for a fixed-origin multi-step forecast of all 2025, because later 2025 rows incorporate outcomes observed during 2025. For fixed-origin evaluation, users must generate predictors recursively or restrict models to information available at the fixed origin [37].

3. Methods

3.1. Reproducible Processing Workflow

Figure 1 summarizes the curation workflow. The process begins from the stable government resource identifier rather than from a manually downloaded local copy. For Version 1.1.0, the official Parquet snapshot was retrieved on 30 August 2026 at 23:38:36 UTC, contained 3,629,980 rows, had a size of 67,929,078 bytes, and produced the SHA-256 digest reported in the source manifest and Data Availability Statement. The downloader verifies the Parquet signature, computes the digest, and writes provenance before transformation. The raw government file is excluded from Git version control; the release stores its manifest and redistributes only the curated 2022–2025 artifacts under the source-compatible license.

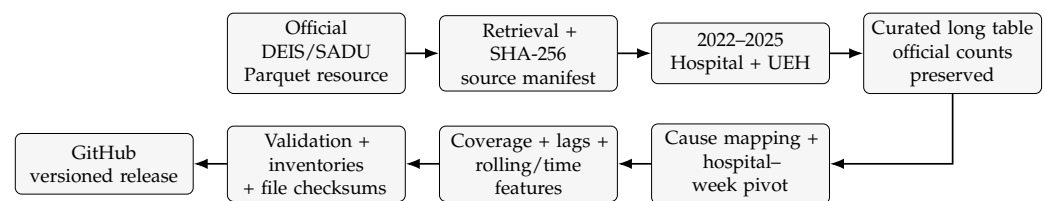


Figure 1. Reproducible Chile-ED-Resp curation workflow. All demand counts originate in the official DEIS/SADU resource; the pipeline adds deterministic organizational, provenance, quality-control, and forecasting fields.

3.2. Schema Harmonization and Type Treatment

Column names are normalized to lower-case snake case. Because public-data schemas can appear with either CamelCase or previously cleaned headings, the build script matches fields through an underscore-insensitive canonical key before checking for the complete expected 25-field schema. If an expected source field is absent, the build stops rather than guessing a replacement. Numeric count, year, week, and cause-order columns are explicitly parsed; coordinates are converted from decimal-comma text to numeric decimal-point representation. No other source values are standardized in a way that changes their substantive meaning.

The curation script also keeps an explicit cause inventory. Consultation versus hospitalization is assigned from the official cause-order group, and `canonical_feature` is assigned only to consultation orders 4, 5, 6, 7, 9, 10, and 11. The inventory records every official label, measure class, number of source rows, aggregate reported total, canonical field, inclusion status, and exclusion reason. Maintaining the original label alongside canonical features allows users to audit the mapping and rebuild alternative targets.

3.3. Fixed Statistical-Week Calendar and Time Index

The release calendar is constructed before inspecting hospital-specific target availability. The official DEIS coordinates present in the archived 2022–2025 source define weeks 1–52 for 2022, 2023, and 2024 and weeks 1–53 for 2025, yielding 209 expected positions. The build checks the observed national IRA Alta coordinate set against this complete grid, validates the week-53 transition, and crosses all 180 hospitals with the grid. A zero-based `time_index` runs continuously from 0 to 208, and `weeks_in_statistical_year` records whether the corresponding year contains 52 or 53 positions. Consequently, a week absent everywhere in a future source snapshot would remain an explicit all-missing grid position rather than causing later row-shift lags to cease representing consecutive calendar positions.

3.4. Hospital Coverage and Selection

Rather than selecting hospitals on the basis of peak demand, geographic convenience, or prior model performance, the release applies a transparent completeness criterion. Let W^* denote the fixed set of 209 expected statistical-week positions, and let $W_h \subseteq W^*$ denote the positions with an observed IRA Alta value for hospital h . Coverage is calculated as follows:

$$C_h = \frac{|W_h|}{|W^*|} = \frac{|W_h|}{209}. \quad (1)$$

The fixed denominator prevents a nationally absent source record from inflating every hospital's coverage. The default benchmark recommendation is $C_h \geq 0.95$. On the 209-position calendar, this rule requires at least 199 observed target weeks and permits at most 10 missing weeks (4.8%). It was specified as a completeness convention independently of forecasting outcomes. The observed coverage distribution provides an additional empirical check: the least-complete retained hospital has 204/209 weeks (97.6%), whereas the most-complete excluded hospital has 178/209 weeks (85.2%). Thus, the 0.95 cutoff lies inside an observed separation gap, and any threshold in $(178/209, 204/209]$ yields the same 178-hospital benchmark cohort. This threshold is not a clinical quality criterion or a claim that 0.95 is universally optimal; it limits avoidable missingness while retaining 98.9% of hospitals, and the full 180-hospital grid remains available for sensitivity analyses using alternative criteria.

3.5. Feature Engineering Without Leakage

For a hospital h and target series $y_{h,t}$ ordered by `time_index`, short-term lag features are $y_{h,t-k}$ for $k \in \{1, 2, 4\}$. The year-over-year feature is not a fixed 52-row shift. For statistical year a and statistical week w , it matches $y_{h,a,w}^{(\text{prev-year})} = y_{h,a-1,w}$ and is therefore named `ira_alta_prev_year_same_week`. This definition handles the 2025 week-53 transition explicitly; where the preceding year lacks week 53 or the matched source value is unavailable, the feature is missing. For prior rolling means, we use

$$\bar{y}_{h,t}^{(m)} = \frac{1}{m} \sum_{j=1}^m y_{h,t-j}, \quad m \in \{4, 12\}, \quad (2)$$

whenever all required prior values exist. Because every hospital is crossed with the complete calendar before feature computation, $t - k$ denotes the exact preceding grid position rather than the k -th preceding observed value. Thus, a missing $y_{h,t-k}$ produces a missing lag and is never bypassed. Rolling windows likewise operate on the preceding four or 12 grid positions with the required observation count equal to the window length; if any position in that window is missing, the rolling mean remains missing rather than shortening the window or skipping the gap. Boundary positions without sufficient history also remain missing. The current observation $y_{h,t}$ is excluded from each rolling feature.

Seasonal position is encoded as

$$s_t = \sin\left(\frac{2\pi(w_t - 1)}{m_t}\right), \quad c_t = \cos\left(\frac{2\pi(w_t - 1)}{m_t}\right), \quad (3)$$

where w_t is the statistical-week number and $m_t \in \{52, 53\}$ is the number of positions in that statistical year. In a 53-week statistical year, week 53 is retained as a distinct observation and is encoded at angle $2\pi(52/53)$, the last unique phase before week 1 of the next year returns to $(s_t, c_t) = (0, 1)$. It is neither dropped nor duplicated as week 1. In a 52-week year, the denominator is 52; consequently, the same numbered week can have a slightly different normalized annual phase across 52- and 53-week years by de-

sign. Therefore, the implementation does not use a fixed denominator of 52: it reads `weeks_in_statistical_year` for every row. The information set for rolling-origin one-week-ahead evaluation at origin t is $\mathcal{I}_{h,t} = \{y_{h,s} : s < t\}$, together with non-target co-variables available by t . The repository example reconstructs all target-derived predictors from $\mathcal{I}_{h,t}$ at every origin and appends $y_{h,t}$ to the history only after the forecast has been scored. For a fixed-origin horizon $H > 1$, users must recursively update target-derived features or omit predictors unavailable at the origin; the precomputed 2025 columns do not by themselves establish leakage-free fixed-origin evaluation.

3.6. Automated Build and Versioning

The public repository contains scripts implemented in Python 3.12.13 (Python Software Foundation, Wilmington, DE, USA) for source retrieval, curation, validation, descriptive audits, and rolling-origin evaluation. Unit tests cover schema normalization, cause/measure separation, the 209-position calendar, and continuous time indexing. The automated workflow is executed using GitHub Actions, a continuously delivered cloud service for which a single conventional version number is not applicable (<https://github.com/features/actions>, accessed on 5 September 2026). The workflow uses Python 3.12.13 and installs `requirements.lock` with the `-require-hashes` option. This lock file records the exact versions and verified hashes of all direct and transitive software dependencies used for the reproducible build. The human-maintained `requirements.txt` records only development-level minimum constraints and is not used for the reproducible workflow. The workflow then executes the unit tests, downloads the official resource, builds both data artifacts, runs validation, and writes the versioned outputs and metadata. The evaluation script implements a one-week horizon for 2025 and recomputes lag-1, lag-2, lag-4, four- and 12-week rolling means, and the previous-year same-week value at every forecast origin without reading the precomputed predictor columns. This approach makes the publication package independently rebuildable without distributing the upstream raw file [38,39].

Version-specific provenance is strengthened through cryptographic hashes. The source manifest records retrieval time, file size, and SHA-256 of the official Parquet snapshot; the final checksum file records hashes of the curated Parquet/CSV artifacts and principal metadata files. Similar checksum-based provenance has recently been used in a national health-data Data Descriptor to distinguish an immutable research release from mutable upstream public files [40].

3.7. Technical Validation

Validation is designed to detect structural and source-quality issues without rewriting official observations. Table 12 lists the automated checks. A FAIL status stops the release when the forecasting table cannot be trusted structurally; a WARN status exposes a potentially relevant anomaly while preserving the source values for transparent secondary analysis.

Table 12. Validation controls and record counts for the 30 August 2026 Version 1.1.0 build. The same fields are distributed in `results/validation_checks.csv`.

Check	Records Examined	Observed Result	Status
Official source counts preserved	2,699,064 count cells	0 modified cells	PASS
Analysis years	449,844 long rows	0 rows outside 2022–2025	PASS
Non-negative counts	2,699,064 count cells	0 negative cells	PASS
Age-total reconciliation	449,844 long rows	0 mismatches	PASS

Table 12. Cont.

Check	Records Examined	Observed Result	Status
Unavailable age reconciliation	449,844 long rows	0 unavailable checks	PASS
Duplicate source strata	449,844 long rows	0 duplicated-key rows	PASS
Unique hospital-week forecast key	37,620 forecast rows	0 duplicated keys	PASS
Chronological ordering	37,620 forecast rows	0 order violations	PASS
Chronological split	37,620 forecast rows	0 invalid split rows	PASS
Fixed 209-week calendar	209 national positions	0 missing/unexpected positions	PASS
Continuous time index	209 positions	0 discontinuities	PASS
Complete hospital-calendar grid	37,620 expected rows	0 missing/extra rows	PASS
Consultation-hospitalization separation	449,844 long rows	0 misclassified rows	PASS
Zero/missing state separation	37,620 target cells	133 missing; 625 explicit zeros; 0 conflation	PASS
No count imputation	2,699,064 count cells	0 imputed cells	PASS

For the 30 August 2026 build of Version 1.1.0, the validation report recorded 15 PASS, 0 WARN, and 0 FAIL statuses. Table 12 reports the exact scope of each check and retains zero-valued exception counts rather than replacing them with a generic pass indicator. It confirms 449,844 long-format rows, 37,620 forecasting rows, 180 hospitals, 12 official labels, the fixed 209-position grid, a continuous time index, and order-aware separation of consultation and hospitalization measures. The rolling-origin example writes origin-level predictions and establishment-level MAE, RMSE, and MAPE; these outputs demonstrate the evaluation protocol and are not presented as evidence that one forecasting model is superior.

4. User Notes

4.1. Reproducing the Release

Users can reproduce the complete release from a clean checkout by executing the commands provided in Listing 1. The reproducible workflow uses Python 3.12.13 and the hash-verified `requirements.lock`; `requirements.txt` remains the editable declaration of minimum development requirements rather than the environment specification used for the release.

Listing 1: Commands for reproducing the complete release from a clean checkout.

```
# Repository accessed on 5 September 2026
git clone https://github.com/cvidalmsu/chile-ed-resp-benchmark.git
cd chile-ed-resp-benchmark
python -m venv .venv
# Activate the environment according to the operating system
python -m pip install --require-hashes -r requirements.lock
pytest -q
python scripts/make_release.py --start-year 2022 --end-year 2025 --force-download
```

Alternatively, repository maintainers can run *Actions* → *Build DEIS-SADU dataset* → *Run workflow*. The action records the newly retrieved source checksum before rebuilding the release. For reuse and citation, users should reference the permanent Chile-ED-Resp Version 1.1.0 archive at <https://doi.org/10.5281/zenodo.22209795>, report the dataset version used in their analysis, and consult the public development repository at <https://github.com/cvidalmsu/chile-ed-resp-benchmark> (accessed on 2 September 2026) for subsequent updates.

4.2. Recommended Forecasting Use

For comparable model experiments, the recommended starting point is the forecasting artifact `data/chile_ed_resp_ira_alta_forecasting.parquet`. Users seeking a high-completeness benchmark should initially restrict the table to the 178 establishments carrying the benchmark recommendation flag. The supplied `split` field provides a 2022–2024/2025 partition, but the valid interpretation of target-derived predictors depends on the forecast design. For rolling-origin one-week-ahead evaluation, users should run `scripts/rolling_origin_example.py`; the script reconstructs the information set at each origin and exposes each recomputed predictor in `results/rolling_origin_predictions.csv`. For a fixed origin preceding 2025, later 2025 target-derived columns must not be used as known inputs; users should generate them recursively or select only origin-available covariates. Randomly shuffled weekly observations are unsuitable for either design [37]. High-dimensional models should be interpreted cautiously because model complexity and sampling frequency can materially affect performance in emergency-demand forecasting [11,41].

For descriptive analyses, the long table is preferable because it retains official age strata, cause wording, hospitalizations, and establishment metadata. Users should avoid summing consultation and hospitalization rows into a single outcome unless the analytical question explicitly requires such an aggregate. The `measure_type`, `causa`, and `orden_causa` fields make that distinction inspectable. Likewise, absent records must not be recoded as zeros without a justified external reporting model.

4.3. Licensing, Attribution, and Provenance

The official resource page labels the source under a Creative Commons Non-Commercial license [21]; the release documents the corresponding CC BY-NC 2.0 licensing terms [32]. The curated tables are adaptations of those records and may be copied, shared, and further adapted only for non-commercial purposes under those terms. Redistribution must retain reasonable attribution to the Chilean Ministry of Health/DEIS, the official dataset title and source link, the CC BY-NC 2.0 license URI, and applicable license/disclaimer notices; it must also identify the transformations made by Chile-ED-Resp and any further changes made by the redistributor. Redistributors may not add legal or technological restrictions that prevent recipients from exercising the licensed rights. Use primarily intended for commercial advantage or private monetary compensation requires separate permission from the applicable rights holder. The repository file `LICENSE-DATA` provides a practical redistribution checklist and the authoritative license link. The MIT license applies only to the software and does not override the non-commercial condition attached to the source-derived data; the original DEIS/SADU records are not relicensed by the authors.

4.4. Limitations and Reuse Hazards

Several limitations should guide interpretation. First, the resource contains aggregate administrative counts rather than individual patient records; it cannot support patient-level risk prediction, causal inference, or analyses requiring comorbidity trajectories. Second, the meaning and completeness of administrative reporting depend on the upstream SADU system and local establishment practices. The pipeline reports anomalies but cannot independently verify whether every source count reflects complete clinical activity. Third, retrospective corrections by the public agency can change the current upstream file; source manifests and versioned releases are therefore essential when reproducing a published analysis.

Further limitations arise from temporal and modeling choices. The 2022–2025 window was selected to create a closed contemporary period, not to claim that earlier years are unusable. The 95% IRA Alta coverage threshold is a reproducible benchmarking rule rather than a universal hospital-inclusion standard. Derived lags and rolling means remain missing when history is insufficient or when an exact required grid position is unavailable; gaps are never skipped or collapsed. Moreover, precomputed target-derived fields encode observations through the preceding row and therefore support sequential one-week origins, not a fixed-origin 53-week claim. No weather, staffing, bed-occupancy, holiday, or socioeconomic variable is added to the core release because such variables require separate provenance and licensing documentation.

Finally, the data should not be interpreted as a direct measure of latent population disease incidence. Emergency consultations reflect both morbidity and healthcare-seeking behavior, local service organization, referral patterns, and reporting practice. Forecasting models built from the resource are therefore best understood as models of reported emergency demand within the represented public-service context. Table 13 summarizes these limitations alongside recommended mitigations for downstream users.

Table 13. Principal limitations of Chile-ED-Resp and recommended mitigations for reuse.

Limitation	Recommended Mitigation/User Guidance
Aggregate administrative counts, not patient-level records	Do not use for patient-level risk prediction or causal inference; restrict to hospital- or population-level demand analysis
SADU reporting completeness depends on the upstream system and local establishment practice	Prioritize the 178 hospitals flagged <code>recommended_for_benchmark = True</code> ; consult <code>hospital_coverage.csv</code> before pooling establishments
Upstream DEIS/SADU file may be retrospectively corrected by the public agency	Cite the Chile-ED-Resp GitHub repository, report the dataset version used, and preserve the SHA-256 checksum of the corresponding source snapshot rather than relying only on the mutable upstream portal URL
The 95% coverage threshold is a reproducibility convention, not a clinical quality standard	Recompute C_h with an alternative threshold using the complete 180-hospital grid if a different inclusion rule is required
Lags and rolling means are missing where history is insufficient or an exact required grid position is unavailable	Preserve the fixed time axis; do not skip gaps or substitute zero for absence
Target-derived 2025 fields become available sequentially	Use the rolling-origin script for one-week forecasts; reconstruct recursively for fixed-origin multi-week evaluation
Absent hospital-week-cause record is not an event count	Preserve as missing; use only explicit source records with <code>num_total = 0</code> as zero events
Consultations reflect healthcare-seeking behavior as well as morbidity	Interpret forecasts as models of reported emergency demand, not of latent disease incidence

5. Conclusions

In summary, Chile-ED-Resp converts a valuable but operationally oriented Chilean public-health resource into a documented and reproducible research dataset without replacing official count observations. Version 1.1.0 comprises a 449,844-row, 32-variable long artifact and a 37,620-row, 38-variable hospital-week artifact covering 180 hospitals, 12 official labels, and 209 expected statistical weeks during 2022–2025. A fixed-denominator coverage rule identifies 178 hospitals for benchmarking while retaining all 180 establish-

ments. The release provides a verbatim cause-to-variable mapping, reports coverage for all seven forecasting causes, distinguishes missing records from explicit zeros, preserves time-specific metadata, excludes hospitalization codes from consultation features, and provides a one-week rolling-origin example. Fifteen validation checks pass with no warnings or failures, and their record scopes and zero-valued exception counts are reported explicitly. Source manifests, checksums, descriptive audits, tests, licensing guidance, and automated rebuilding support traceability from DEIS/SADU to the versioned release.

Author Contributions: Conceptualization, W.B.-Q., C.A. and C.V.-S.; methodology, C.A., C.C.-G., B.S.-N., J.F.-B. and C.V.-S.; software, B.S.-N., J.F.-B. and M.T.-Y.; validation, C.C.-G., B.S.-N., J.F.-B. and M.T.-Y.; formal analysis, C.A., C.C.-G., B.S.-N. and J.F.-B.; data curation, B.S.-N., J.F.-B. and C.C.-G.; writing—original draft preparation, W.B.-Q., C.A., B.S.-N. and J.F.-B.; writing—review and editing, all authors; visualization, B.S.-N., J.F.-B. and M.T.-Y.; supervision, C.A. and C.V.-S.; project administration, W.B.-Q. and C.V.-S. 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. The dataset is derived from publicly available, aggregate administrative health statistics and contains no patient-level records or direct personal identifiers.

Informed Consent Statement: Not applicable. No individual participants were recruited and no identifiable individual-level data were accessed.

Data Availability Statement: The curated dataset, processing code, machine-readable descriptive audits, validation outputs, and documentation for Version 1.1.0 are permanently archived in Zenodo at <https://doi.org/10.5281/zenodo.22209795>. The public development repository is available at <https://github.com/cvidalmsu/chile-ed-resp-benchmark> (accessed on 2 September 2026). Version 1.1.0 was built from the official DEIS/SADU Parquet snapshot retrieved on 30 August 2026 at 23:38:36 UTC (67,929,078 bytes). The source SHA-256 is ef21c621527091dc109ecda8a66f11199d8680b50435d4d77cf69626a08d90e0. The upstream records are publicly available from the Chilean Ministry of Health/DEIS dataset “Atenciones de urgencias de causas respiratorias por semana epidemiológica”, resource ID ae6c9887-106d-4e98-8875-40bf2b836041.

Acknowledgments: The authors acknowledge the Departamento de Estadísticas e Información de Salud (DEIS), Ministerio de Salud de Chile, for maintaining the public SADU respiratory-emergency dataset used as the source of this curated resource.

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

References

1. Lucini, F.R.; dos Reis, M.A.; da Silveira, G.J.C.; Fogliatto, F.S.; Anzanello, M.J.; Andrioli, G.G.; Nicolaidis, R.; Beltrame, R.C.F.; Neyeloff, J.L.; Schaan, B.D. Man vs. Machine: Predicting Hospital Bed Demand from an Emergency Department. *PLoS ONE* **2020**, *15*, e0237937. [CrossRef] [PubMed]
2. Duarte, D.; Walshaw, C.; Ramesh, N. A Comparison of Time-Series Predictions for Healthcare Emergency Department Indicators and the Impact of COVID-19. *Appl. Sci.* **2021**, *11*, 3561. [CrossRef]
3. Albrecht, S.; Broderick, D.; Dost, K.; Cheung, I.; Nghiem, N.; Wu, M.; Zhu, J.; Poonawala-Lohani, N.; Jamison, S.; Rasanathan, D.; et al. Forecasting Severe Respiratory Disease Hospitalizations Using Machine Learning Algorithms. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 293. [CrossRef] [PubMed]
4. Ackermann, A.E.F.; Fani, V.; Bandinelli, R.; Sellitto, M.A. Short-Term Prediction in an Emergency Healthcare Unit: Comparison Between ARIMA, ANN, and Logistic Map Models. *Forecasting* **2025**, *7*, 52. [CrossRef]
5. Wargon, M.; Guidet, B.; Hoang, T.D.; Hejblum, G. A Systematic Review of Models for Forecasting the Number of Emergency Department Visits. *Emerg. Med. J.* **2009**, *26*, 395–399. [CrossRef] [PubMed]
6. Vollmer, M.A.C.; Glampson, B.; Mellan, T.; Mishra, S.; Mercuri, L.; Costello, C.; Klaber, R.; Cooke, G.; Flaxman, S.; Bhatt, S. A Unified Machine Learning Approach to Time Series Forecasting Applied to Demand at Emergency Departments. *BMC Emerg. Med.* **2021**, *21*, 9. [CrossRef] [PubMed]

7. Rostami-Tabar, B.; Browell, J.; Svetunkov, I. Probabilistic Forecasting of Hourly Emergency Department Arrivals. *Health Syst.* **2024**, *13*, 133–149. [[CrossRef](#)] [[PubMed](#)]
8. Tuominen, J.; Pulkkinen, E.; Peltonen, J.; Kannianen, J.; Oksala, N.; Palomäki, A.; Roine, A. Forecasting Emergency Department Occupancy with Advanced Machine Learning Models and Multivariable Input. *Int. J. Forecast.* **2024**, *40*, 1410–1420. [[CrossRef](#)]
9. Porto, B.M.; Fogliatto, F.S. Enhanced Forecasting of Emergency Department Patient Arrivals Using Feature Engineering Approach and Machine Learning. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 377. [[CrossRef](#)] [[PubMed](#)]
10. Alenezi, A.M.; Sameh, M.; Aljohani, M.; Alharbi, H. Predicting Emergency Department Patient Arrivals at Hospitals Using Machine Learning Techniques. *Healthcare* **2026**, *14*, 1191. [[CrossRef](#)] [[PubMed](#)]
11. Blanco, J.; Ferreras, M.; Cosido, O. Predictive Modeling of Hospital Emergency Department Demand Using Artificial Intelligence: A Systematic Review. *Int. J. Med. Inform.* **2026**, *207*, 106215. [[CrossRef](#)] [[PubMed](#)]
12. Hosseini, B.; Patel, A.; Landes, M.; Vaillancourt, S.; Mamdani, M.; Maruthananth, K.; Matharu, N.; Pathan, Z.; Sivapragasam, K.; Ruangsomboon, O.; et al. Implementation of Machine Learning in Emergency Departments: A Systematic Review. *Digit. Health* **2026**, *12*, 20552076251411209. [[CrossRef](#)] [[PubMed](#)]
13. Mellor, J.; Christie, R.; Overton, C.E.; Paton, R.S.; Leslie, R.; Tang, M.; Deeny, S.; Ward, T. Forecasting Influenza Hospital Admissions Within English Sub-Regions Using Hierarchical Generalised Additive Models. *Commun. Med.* **2023**, *3*, 190. [[CrossRef](#)] [[PubMed](#)]
14. Mellor, J.; Tang, M.L.; Jones, O.; Ward, T.; Riley, S.; Deeny, S.R. Forecasting COVID-19, Influenza, and RSV Hospitalizations over Winter 2023–4 in England. *Int. J. Epidemiol.* **2025**, *54*, dyaf066. [[CrossRef](#)] [[PubMed](#)]
15. Cutrera, R.; Ciofi degli Atti, M.L.; Dotta, A.; D'Amore, C.; Ravà, L.; Perno, C.F.; Villani, A. Epidemiology of Respiratory Syncytial Virus in a Large Pediatric Hospital in Central Italy and Development of a Forecasting Model to Predict the Seasonal Peak. *Ital. J. Pediatr.* **2024**, *50*, 65. [[CrossRef](#)] [[PubMed](#)]
16. Chen, H.; Xiao, M. Seasonality of Influenza-Like Illness and Short-Term Forecasting Model in Chongqing from 2010 to 2022. *BMC Infect. Dis.* **2024**, *24*, 432. [[CrossRef](#)] [[PubMed](#)]
17. Toro Castillo, C.; Parraguez Delgado, B.; Pérez, J. Tasa de Consultas Médicas en el Servicio de Urgencias del Hospital Base de Linares, entre los Años 2019–2023 en Chile. *Rev. ANACEM* **2024**, *18*, 96–104. [[CrossRef](#)]
18. Díaz, F.; Carvajal, C. Análisis de la Epidemia Estacional por Virus Respiratorios y Virus Respiratorio Sincial (VRS) del Año 2023 Comparado con Cohorte Prepandemia (2015–2019) en Chile. *Andes Pediatr.* **2025**, *96*, 13–22. [[CrossRef](#)]
19. Zepeda, F.G.; Díaz, A.P.; Ortiz, F.P.; Palomino, M.M.A.; Tapia, F.L.; Puppo, G.H.; Rodríguez, B.H.; Prado, A.F.; Núñez, C.M.; Rivera, K.N.; et al. Bronquiolititis viral: Recomendaciones sobre su manejo en Chile 2024. *Rev. Chil. Enferm. Respir.* **2024**, *40*, 261–273. [[CrossRef](#)]
20. Díaz Rubio, F. Infecciones Respiratorias Agudas Graves en Niños: Desafíos Actuales en un Panorama Epidemiológico en Evolución. *Andes Pediatr.* **2026**, *97*, 10–13. [[CrossRef](#)]
21. Ministerio de Salud de Chile, Departamento de Estadísticas e Información de Salud (DEIS). Atenciones de Urgencias de Causas Respiratorias por Semana Epidemiológica. Portal de Datos Abiertos del Estado de Chile. Dataset ID: 606ef5bb-11d1-475b-b69f-b980da5757f4; Resource ID: ae6c9887-106d-4e98-8875-40bf2b836041; Parquet Resource. 2026. Available online: <https://datos.gob.cl/dataset/606ef5bb-11d1-475b-b69f-b980da5757f4/resource/ae6c9887-106d-4e98-8875-40bf2b836041> (accessed on 30 August 2026).
22. 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](#)]
23. Inau, E.T.; Sack, J.; Waltemath, D.; Zeleke, A.A. Initiatives, Concepts, and Implementation Practices of the Findable, Accessible, Interoperable, and Reusable Data Principles in Health Data Stewardship: Scoping Review. *J. Med. Internet Res.* **2023**, *25*, e45013. [[CrossRef](#)] [[PubMed](#)]
24. Bahim, C.; Casorrán-Amilburu, C.; Dekkers, M.; Herczog, E.; Loozen, N.; Repanas, K.; Russell, K.; Stall, S. The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Sci. J.* **2020**, *19*, 41. [[CrossRef](#)]
25. Devaraju, A.; Huber, R. An Automated Solution for Measuring the Progress Toward FAIR Research Data. *Patterns* **2021**, *2*, 100370. [[CrossRef](#)] [[PubMed](#)]
26. Devaraju, A.; Mokrane, M.; Cepinskas, L.; Huber, R.; Herterich, P.; de Vries, J.; Akerman, V.; L'Hours, H.; Davidson, J.; Diepenbroek, M. From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. *Data Sci. J.* **2021**, *20*, 4. [[CrossRef](#)]
27. Groth, P.; Cousijn, H.; Clark, T.; Goble, C. FAIR Data Reuse—The Path through Data Citation. *Data Intell.* **2020**, *2*, 78–86. [[CrossRef](#)]
28. Fenner, M.; Crosas, M.; Grethe, J.S.; Kennedy, D.; Hermjakob, H.; Rocca-Serra, P.; Durand, G.; Berjon, R.; Karcher, S.; Martone, M.; et al. A Data Citation Roadmap for Scholarly Data Repositories. *Sci. Data* **2019**, *6*, 28. [[CrossRef](#)] [[PubMed](#)]

29. Ministerio de Salud de Chile, Departamento de Estadísticas e Información de Salud (DEIS). Atenciones de Urgencia de la Red Pública de Salud. Portal de Datos Abiertos del Estado de Chile. Annual Aggregated Emergency-Care Files Derived from DAU Records; Files Available for 2017–2024. 2024. Available online: <https://datos.gob.cl/dataset/atenciones-de-urgencia> (accessed on 30 August 2026).
30. Ministério da Saúde do Brasil. SRAG 2019 a 2026: Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe). Portal de Dados Abertos do SUS. Public Case-Level Severe Acute Respiratory Syndrome Surveillance Data. 2026. Available online: <https://dadosabertos.saude.gov.br/dataset/srag-2019-a-2026> (accessed on 30 August 2026).
31. Pan American Health Organization. PAHO/EIH Open Data: Influenza, COVID-19 and Respiratory Viruses. Regional Country-Level Surveillance Indicators. 2026. Available online: <https://opendata.paho.org/en> (accessed on 30 August 2026).
32. Creative Commons. Attribution-NonCommercial 2.0 Generic (CC BY-NC 2.0). Canonical License Deed. 2004. Available online: <https://creativecommons.org/licenses/by-nc/2.0/> (accessed on 30 August 2026).
33. Vetrò, A.; Canova, L.; Torchiano, M.; Minotas, C.O.; Iemma, R.; Morando, F. Open Data Quality Measurement Framework: Definition and Application to Open Government Data. *Gov. Inf. Q.* **2016**, *33*, 325–337. [\[CrossRef\]](#)
34. Bhandari, S.; Ranjan, N.; Kim, Y.C.; Park, J.D.; Hwang, K.I.; Kim, W.H.; Hong, Y.S.; Kim, H. An Automatic Data Completeness Check Framework for Open Government Data. *Appl. Sci.* **2021**, *11*, 9270. [\[CrossRef\]](#)
35. Ministerio de Salud de Chile, Departamento de Estadísticas e Información de Salud (DEIS). Diccionario de Datos: Atenciones de Urgencias Respiratorias Semanal. Portal de Datos Abiertos del Estado de Chile. Official Data-Dictionary Resource. 2024. Available online: <https://datos.gob.cl/dataset/atenciones-de-urgencia-causas-respiratorias/resource/27fbe340-0e74-465f-b9f9-488c3bf6533c> (accessed on 30 August 2026).
36. Lin, J.; De Silva, A.P.; Falster, M.; Dobbins, T.; Chen, Y.; Zhang, Z.; Cao, Y.; Davidson, P.M. Addressing Missing Data in Real-World Administrative Health Datasets. *Eur. J. Cardiovasc. Nurs.* **2026**, *25*, 632–639. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021.
38. Kim, A.Y.; Herrmann, V.; Barreto, R.; Calkins, B.; Gonzalez-Akre, E.; Johnson, D.J.; Jordan, J.A.; Magee, L.; McGregor, I.R.; Montero, N.; et al. Implementing GitHub Actions Continuous Integration to Reduce Error Rates in Ecological Data Collection. *Methods Ecol. Evol.* **2022**, *13*, 2572–2585. [\[CrossRef\]](#)
39. Braga, P.H.P.; Hébert, K.; Hudgins, E.J.; Scott, E.R.; Edwards, B.P.M.; Sánchez Reyes, L.L.; Grainger, M.J.; Foroughirad, V.; Hillemann, F.; Binley, A.D.; et al. Not Just for Programmers: How GitHub Can Accelerate Collaborative and Reproducible Research in Ecology and Evolution. *Methods Ecol. Evol.* **2023**, *14*, 1364–1380. [\[CrossRef\]](#)
40. Voinea, D.V.; Voinea, S.A. A Decade of National Patient Experience Feedback in Romania (2016–2026): A Harmonized, Reproducible Open Dataset. *Data* **2026**, *11*, 185. [\[CrossRef\]](#)
41. Tuominen, J.; Lomio, F.; Oksala, N.; Palomäki, A.; Peltonen, J.; Huttunen, H.; Roine, A. Forecasting Daily Emergency Department Arrivals Using High-Dimensional Multivariate Data: A Feature Selection Approach. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 134. [\[CrossRef\]](#) [\[PubMed\]](#)

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.