Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# An integrated framework for public health laboratory, epidemiological, and territorial data supports a denominator-aware approach to RSV surveillance in Minas Gerais, Brazil

**Luiz Alcantara**

luiz.alcantara@fiocruz.br

Instituto Rene Rachou, Fundacao Oswaldo Cruz

**André Leal Bernardes**

Ezequiel Dias Foundation/Central Public Health Laboratory https://orcid.org/0000-0002-2814-5696

**Bárbara Andrade**

**Gilmar Rodrigues**

**Janaína Almeida**

**Josiane Piedade**

**René Carvalho**

**Eva Medeiros**

**Pedro Alves**

Fiocruz https://orcid.org/0000-0002-9659-5621

**Marcelo Gomes**

**Rafaela Andrade**

**Walquíria Almeida**

**Miriam Livorati**

**Felipe Iani**

Laboratório Central de Saúde Pública, Instituto Octávio Magalhães, FUNED, Belo Horizonte, Minas Gerais, Brazil https://orcid.org/0000-0002-3629-8529

**Marta Giovanetti**

Instituto Oswaldo Cruz, FIOCRUZ https://orcid.org/0000-0002-5849-7326

**Luiz Carlos Junior Alcantara**

Instituto Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil; Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz

**José Lourenço**

Universidade Católica Portuguesa, Católica Medical School, Católica Biomedical Research Center, Lisboa, Portugal https://orcid.org/0000-0002-9318-2581

**GRIPE-MG Team**

Central Public Health Laboratory of the State of Minas Gerais, Ezequiel Dias Foundation, Brazil;

**Article**

**Keywords:** Respiratory Syncytial Virus, Laboratory Surveillance, Public Health Surveillance, Denominator-Aware Surveillance, Data Integration, Bayesian Analysis, Brazil

**Posted Date:** September 15th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10843925/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** There is **NO** Competing Interest.

# Abstract

### Background

Brazil's public health laboratory network comprises 27 Central Public Health Laboratories sharing a national infrastructure. These laboratories receive specimens through surveillance, diagnostic, hospital and reference-laboratory workflows. Records outside epidemiological surveillance are not always represented in, or linkable to, national notification databases. We assessed whether integrating these records could extend RSV surveillance beyond aggregate detection counts.

### Methods

We retrospectively analysed RSV testing records from Funed/Lacen-MG, Minas Gerais, Brazil, during 2021–2025. Records were deduplicated, standardized, assigned epidemiological timing and territory context, and linked, when available, to influenza-like illness (ILI) and severe acute respiratory infection (SARI) records. Analyses distinguished testing volume from valid RSV-tested denominators; beta-binomial Bayesian models estimated age-specific positivity trajectories and peak timing.

### Results

The panel comprised 134,730 records from 441 municipalities across all 14 health macroregions; 134,235 had valid RSV outcomes and 11,713 were RSV-detectable (8.7%). 81,818 records (60.7%) were linked to ILI and/or SARI surveillance, whereas 52,912 (39.3%) had no linkage and were retained for laboratory analyses. Annual detection-count peaks shifted from the epidemiological-year boundary in 2021–2022 to weeks 11, 15 and 21 in 2023–2025. Testing intensity and laboratory-detected activity varied across macroregions. From 2023 onward, estimated positivity peaks among children younger than 5 years preceded those among adults and older adults by 1–3 weeks, with posterior probabilities of 0.935–1.000.

### Conclusions

Denominator-aware integration generated temporal, territorial and age-specific RSV surveillance evidence beyond aggregate counts. If prospectively validated as an early signal of broader RSV activity, the 1–3-week lead among young children could support seasonal preparedness; its utility requires evaluation against reporting delays, hospital demand and intervention timelines.


# Introduction

Respiratory syncytial virus (RSV) is a major cause of acute respiratory infection across the life course, with the greatest burden occurring among infants and young children, older adults, and clinically vulnerable populations[1,2]. In 2019, RSV was estimated to cause approximately 33 million episodes of acute lower respiratory infection, 3.6 million hospital admissions, and 101,400 deaths among children younger than 5 years globally, with mortality disproportionately concentrated in low- and middle-income

countries[3]. RSV is also increasingly recognized as an important cause of severe respiratory disease and hospitalization among older adults, particularly those with chronic cardiopulmonary disease, frailty, or immunosuppression[4]. These age-specific patterns of disease burden underscore the need for surveillance systems capable of characterizing not only how much RSV is detected, but also when and where RSV activity occurs and which population groups are most affected.

The COVID-19 pandemic reshaped both respiratory virus epidemiology and public health laboratory operations. Public health laboratories substantially expanded molecular diagnostic capacity and adapted testing workflows for surveillance and response[5–7]. At the same time, non-pharmaceutical interventions, changes in healthcare-seeking behaviour, mobility and school attendance, together with shifting laboratory priorities, altered the timing and detectability of seasonal respiratory virus activity[8,9]. The expansion of laboratory capacity increased the volume and potential analytical value of routinely generated data. However, changes in testing volume, eligibility criteria and the composition of the tested population also complicated their interpretation over time. Together, these developments highlighted an opportunity to extend laboratory surveillance beyond testing and reporting by incorporating a more systematic analytical function. Realizing this potential requires laboratory data to be interpreted in relation to valid tested-sample denominators, epidemiological timing, territorial context and, when available, information from epidemiological surveillance systems.

In Brazil, respiratory virus surveillance operates through established national epidemiological systems. National guidelines include sentinel surveillance for influenza-like illness (ILI) and surveillance of severe acute respiratory infection (SARI), with notifications recorded in SIVEP-Gripe (Sistema de Informação da Vigilância Epidemiológica da Gripe), the national epidemiological information system for influenza and other respiratory viruses[10,11].

Laboratory surveillance is organized within Brazil's National Public Health Laboratory System (SISLAB; *Sistema Nacional de Laboratórios de Saúde Pública*), which includes 27 Central Public Health Laboratories (Lacens), one in each state and the Federal District. Within the respiratory virus laboratory network, clinical specimens are routinely tested for SARS-CoV-2, influenza viruses, RSV and other respiratory pathogens. Laboratory test requests, associated metadata and results are recorded in the Laboratory Environment Management System (GAL; Gerenciador de Ambiente Laboratorial), the national laboratory information system used across the Lacen network to manage and report public health laboratory testing[12,13].

Although laboratory results generated within SISLAB already contribute to respiratory virus surveillance, routine laboratory workflows necessarily prioritize specimen processing, result generation, reporting and established monitoring functions. However, the temporal, territorial, demographic and operational dimensions of data routinely generated by Lacens are not necessarily integrated into a standardized analytical workflow. These data are generated within a heterogeneous testing ecosystem in which specimens reach public health laboratories through surveillance, diagnostic, clinical and referral pathways, including sentinel respiratory virus surveillance, SARI surveillance and other respiratory virus

testing programmes. Consequently, using these records for surveillance requires more than counting positive results. This requires accounting for testing volume and sample-stream composition, defining valid tested-sample denominators, assigning epidemiologically meaningful reference dates, incorporating territorial context and, when possible, linking laboratory records to epidemiological surveillance data. Without such an analytical structure, raw detection counts cannot be interpreted as reflecting pathogen activity alone because they are also shaped by testing demand, eligibility criteria, referral pathways, clinical priorities and changes in testing practices[14–16].

To address this analytical gap, we developed a denominator-aware surveillance intelligence (DASI) approach that structures routinely generated public health laboratory data around valid tested-sample denominators, epidemiological timing, territorial context, linkage to epidemiological surveillance data when available and uncertainty-aware analyses. DASI was implemented through GALAXIS (GAL Analytics and eXtensible Integrated Surveillance), a modular data-integration and analytical framework developed to combine heterogeneous laboratory, epidemiological, territorial and population data sources and construct auditable analytical datasets. In this implementation, DASI provided a systematic analytical layer for laboratory surveillance that complemented, rather than replaced, established epidemiological surveillance systems.

Minas Gerais provided a suitable setting for this implementation because of its large population[15] and the diversity of referral pathways represented in routine laboratory records. Data generated across the state capture geographic and operational heterogeneity in surveillance streams, tested population groups, referral pathways and diagnostic demand. This setting therefore supported a large-scale subnational implementation within Brazil's national public health laboratory infrastructure, without implying that the epidemiological patterns observed in Minas Gerais are representative of other Brazilian states.

In this study, we applied the DASI approach to routine RSV testing data generated by Funed/Lacen-MG between 2021 and 2025. Funed/Lacen-MG is the Central Public Health Laboratory of Minas Gerais and part of Brazil's National Public Health Laboratory System (SISLAB). We examined testing volume and positivity, post-pandemic RSV seasonality, territorial heterogeneity in testing intensity and laboratory-detected activity, and the distribution of samples across surveillance-stream categories. We further used Bayesian modelling to estimate age-specific RSV positivity trajectories and quantify uncertainty in annual peak timing across age groups. Through this large-scale subnational implementation, we assessed whether structuring routine laboratory data around tested-sample denominators, epidemiological timing, territorial context and linkage to epidemiological surveillance records could extend the analytical contribution of public health laboratories beyond test generation and reporting.

## Materials and methods

# Raw operational data sources, specimens and RSV testing

We retrospectively analysed routine operational data generated through public health RSV testing in Minas Gerais, Brazil, between 2021 and 2025. Laboratory records were obtained from GAL, the Brazilian laboratory information management system used for public health laboratory test requests, associated metadata and results. These records were enriched with SIVEP-Gripe data from influenza-like illness (ILI) and severe acute respiratory infection (SARI) surveillance, territorial reference files defining municipalities and health macroregions in Minas Gerais, and 2024 municipal population estimates from the Brazilian Institute of Geography and Statistics (IBGE) (Fig. 1). The corresponding clinical respiratory specimens were processed by Funed/Lacen-MG through routine respiratory virus surveillance and other laboratory testing pathways.

Records were eligible for JAC construction if they represented a clinical respiratory specimen tested for RSV by Funed/Lacen-MG, were assigned to Minas Gerais and fell within the prespecified source-data window. After sample-level deduplication, records with an epidemiological reference date assigned to ISO epidemiological years 2021–2025 formed the primary analytical panel. Detectable, non-detectable and inconclusive RSV outcomes were retained for sample-flow and testing-volume analyses; however, only definitive detectable and non-detectable outcomes were included in RSV positivity denominators. Clinical specimens included nasopharyngeal swabs, aspirates, lavage specimens, respiratory tissue specimens and respiratory secretions. Nucleic acids were extracted using the Extracta 96 automated platform with the Extracta Viral DNA/RNA kit (Loccus, São Paulo, Brazil). RSV testing was performed by real-time reverse-transcription PCR (RT-qPCR) using either the 1copy COVID-19/FluA/FluB/RSV qPCR Kit (1drop Inc., Seongnam, Republic of Korea) or the VR1/VR2 assay (Bio-Manguinhos/Fiocruz, Rio de Janeiro, Brazil) on a 7500 Real-Time PCR System (Applied Biosystems).

Because two diagnostic assays were used during the study period, RSV outcomes were classified according to the assay-specific interpretation criteria applied in the routine Funed/Lacen-MG workflow rather than by imposing a single cross-assay cycle-threshold cutoff. RSV-detectable samples were defined as those with an assay-generated detectable *Amp Status* and characteristic RSV amplification, supported by inspection of the amplification curve and fluorescence-intensity profile in the raw data plot. RSV-non-detectable samples showed no RSV-specific amplification and met assay validity criteria. Samples were classified as inconclusive when the assay generated an inconclusive *Amp Status* or when abnormalities identified during analytical interpretation precluded definitive classification as either detectable or non-detectable. Inconclusive results were retained in the analytical audit trail but excluded from RSV positivity denominators.

## GALAXIS data integration and analytical dataset construction

Data processing, harmonization, integration and analytical dataset construction were coordinated through GALAXIS, an in-house modular data-integration and analytical framework used to operationalize the DASI approach across laboratory, epidemiological, territorial and population data sources (Fig. 1).

GAL-derived laboratory exports were initially integrated using a Power BI/Power Query workflow to generate a consolidated laboratory dataset. This laboratory data layer was subsequently processed using Python-based GALAXIS modules to create standardized analytical variables while retaining the source laboratory information required for auditability. Laboratory fields were harmonized for sample identification, RSV result classification, biological material, demographic characteristics, requester information and laboratory dates. The epidemiological reference date was defined as the symptom-onset date when available, with the specimen collection date used as a fallback. Registration, receipt and result-release dates were not used as epidemiological anchors. ISO epidemiological year, ISO week and ISO year-week were derived from this reference date. A standardized sample key was generated from the laboratory sample code. Records sharing the same standardized key were consolidated deterministically, prioritizing records with a valid epidemiological reference date, symptom-onset over collection-date anchoring, greater field completeness and the earliest valid date, with stable source order used as the final tie-breaker. The resulting MASTER LAB dataset contained one record per distinct laboratory sample.

Clinical, epidemiological and territorial enrichment was subsequently performed by JANUS, a GALAXIS module that used the MASTER LAB as its laboratory data backbone while preserving the sample-level denominator. Health macroregion was assigned using the requesting municipality code. Accordingly, territorial indicators represent the geographic distribution of laboratory requests according to the requesting municipality and not necessarily the municipality of patient residence. Laboratory records were linked to SIVEP-Gripe ILI and SARI records when the information required for linkage was available. JANUS incorporated linkage indicators and selected clinical and epidemiological information without redefining the primary laboratory result, epidemiological reference date, or sample-level denominator.

JANUS generated the JANUS Analytical Core (JAC), an auditable, analysis-ready sample-level dataset used for downstream analyses (Fig. 1). The analytical unit was an individual laboratory sample tested for RSV; therefore, distinct samples did not necessarily correspond to distinct individuals or clinical episodes.

The JAC provided the integrated analytical dataset used in subsequent analyses, including standardized RSV outcome categories, valid RSV-tested denominators, demographic characteristics, epidemiological timing based on ISO weeks, municipality and health macroregion identifiers, and clinical and epidemiological information incorporated through SIVEP-Gripe enrichment when available. For population-standardized territorial analyses, JAC-derived macroregional data were combined with 2024 municipal population estimates from IBGE within the GALAXIS analytical workflow. Figure 1 summarizes the complete analytical workflow, from heterogeneous operational data sources through laboratory data integration, GALAXIS-mediated data processing and analysis, JANUS enrichment and JAC generation to the analyses reported in this study.

## SIVEP-Gripe enrichment and linkage classification

Laboratory samples were deterministically linked to SIVEP-Gripe ILI and SARI notification records using the corresponding laboratory-request identifier after prespecified character normalization. The laboratory dataset was preserved through left joins. When multiple notification candidates shared the same normalized identifier, a single candidate was selected using the following prespecified hierarchy: greater clinical completeness, availability of clinical outcome information, availability of severity information, shorter temporal distance to the laboratory dates, earlier notification date and stable notification-identifier order. Laboratory samples were initially classified as SARI-linked, ILI-linked, linked to both surveillance streams, or without identified ILI/SARI linkage. These classifications were treated as surveillance-stream indicators rather than as mutually exclusive epidemiological strata. Samples without identified SIVEP-Gripe linkage were retained in the analytical dataset as records with no identified ILI/SARI linkage. This classification indicates only that no corresponding ILI or SARI notification was identified and does not establish the original testing or referral pathway.

For analyses requiring mutually exclusive surveillance-stream categories, samples linked to both SARI and ILI were assigned to the SARI-linked group to avoid double-counting. The resulting categories were therefore SARI-linked, ILI-linked only, and no identified ILI/SARI linkage.

## Descriptive and territorial analyses

Descriptive analyses used the JAC to summarize RSV testing volume, RSV-detectable samples, weekly RSV positivity, seasonality, surveillance-stream composition and linkage, comparison with SIVEP-Gripe SARI notifications, territorial heterogeneity, and age and sex distributions. RSV positivity was calculated as the number of RSV-detectable samples divided by the number of valid RSV-tested samples within each analytical stratum. Weekly positivity was summarized using centred 3-week moving averages. For the surveillance-stream-specific peak visualization in Fig. 4, annual peak positivity was identified only among weeks with at least 10 valid RSV-tested samples within the corresponding surveillance-stream category. For descriptive demographic summaries, age was grouped as younger than 2 years, 2–5 years, 6–18 years, 19–59 years and 60 years or older. For the observed age-specific weekly positivity trajectories shown in Fig. 5, age groups were aligned with those used in the Bayesian peak-timing analysis: younger than 5 years, 5–17 years, 18–59 years, 60–79 years and 80 years or older.

Territorial analyses used municipalities as the lowest spatial unit and health macroregions as the primary aggregation level. Municipality codes and health macroregion classifications were obtained from territorial reference files. Municipal population estimates for 2024 from IBGE were aggregated to the health macroregion level and used as fixed population denominators across all study years. RSV testing intensity was defined as the number of distinct laboratory samples tested for RSV within the ISO-week analytical panel per 100,000 residents, by health macroregion and year. This testing-volume indicator included samples with inconclusive outcomes. Laboratory-detected RSV activity was defined as the number of distinct RSV-detectable laboratory samples per 100,000 residents by health macroregion and year. Global empirical Bayes smoothing was applied separately within each ISO epidemiological year to reduce instability associated with sparse detections and smaller population

denominators. For each year, the pooled detection rate across macroregions was used as the shrinkage mean, and the between-macroregion variance was estimated by subtracting the mean Poisson sampling variance from the observed variance of the crude rates, truncated at zero. Each crude macroregional rate was then weighted according to its estimated sampling variance and shrunk towards the corresponding annual pooled rate. These indicators were explicitly interpreted as measures of laboratory testing activity and laboratory-detected RSV activity.

## Bayesian age-specific peak-timing analysis

Bayesian modelling was used to estimate age-specific weekly RSV positivity trajectories and annual peak timing. Weekly counts of RSV-detectable samples and valid RSV-tested samples were aggregated from the JAC by ISO epidemiological week and age group. A beta-binomial observation model with a logit link was used to account for extra-binomial variation. Spline-based temporal components represented smooth changes in RSV positivity over time, while annual sine and cosine terms and age-specific intercept and spline deviations allowed seasonal trajectories to vary among age groups. Age groups for Bayesian modelling were selected to provide stable weekly trajectory estimates rather than to reproduce the categories used in descriptive demographic summaries. Children younger than 5 years were grouped together to capture the main paediatric RSV surveillance signal while reducing sparse weekly strata. Older adults were separated into 60–79 years and those aged 80 years or older to preserve a clinically relevant distinction while retaining sufficient weekly information for estimation. The Bayesian age groups were therefore younger than 5 years, 5–17 years, 18–59 years, 60–79 years and 80 years or older.

For each posterior draw, year and age group, the annual peak week was defined as the epidemiological week with the highest estimated RSV positivity. Pairwise posterior comparisons were then used to estimate the probability that one age group peaked earlier than another and to quantify corresponding differences in peak timing. These comparisons were interpreted as uncertainty-aware temporal surveillance indicators and not as evidence of causal transmission between age groups. Full model specification, priors, sampling settings, diagnostic results and posterior predictive checks are reported in the **Supplementary Information**.

# Statistics and Reproducibility

This retrospective study included all eligible routine RSV laboratory sample records available during the study period after application of the predefined data-processing, deduplication and outcome-classification criteria. No a priori sample-size calculation was performed because the study analysed routinely generated surveillance records rather than a prospectively sampled study population. The analytical unit was a distinct laboratory sample, identified by its unique laboratory sample identifier; therefore, samples did not necessarily correspond to unique individuals or independent clinical episodes.

Descriptive analyses were based on counts, proportions, RSV positivity estimates, and population-standardized laboratory surveillance indicators, as defined above. Bayesian estimates were summarized using posterior medians, 50% and 95% credible intervals, and posterior probabilities. Model convergence and computational stability were assessed using divergent transitions, R-hat statistics, and effective sample sizes. Model-adequacy was further evaluated using posterior predictive checks and Pareto-smoothed importance sampling leave-one-out cross-validation diagnostics. Full model specification, priors, sampling settings and diagnostic results are reported in the **Supplementary Information.**

Because this study used secondary, routinely collected surveillance records rather than experimental observation generated specifically for the study, biological and technical replicates were not defined for the statistical analyses. Data-processing procedures followed predefined analytical rules and generated auditable intermediate datasets and outputs to support traceability and reproducibility throughout the analytical workflow.

Data processing and analysis used Microsoft Power BI Desktop 2.155.756.0, Python 3.11.13 and R 4.3.3; Bayesian models were fitted using PyMC 5.27.0.

## Ethics approval and consent

This study was approved by the Research Ethics Committee of Fundação Ezequiel Dias (CEP-FUNED), Minas Gerais, Brazil (approval No. 8.605.939; CAAE 00138726.5.0000.9507). The requirement for individual informed consent was waived by the Research Ethics Committee because the study used secondary data generated through routine public health surveillance. Direct personal identifiers and variables that could enable patient identification were excluded from the analytical dataset. Analyses were conducted using the laboratory sample code as the sample identifier, and results are reported only in aggregate.

## Code availability

Custom code was used for study-specific data integration, analytical derivations, statistical analyses and figure generation. A de-identified study-level analytical dataset, data dictionary, dataset-construction specification, supporting aggregate and territorial inputs, and frozen article-specific R and Python scripts required to reproduce the reported aggregate results, tables and figures from the analysis-ready dataset will be made publicly available in the RSV-DASI-Minas-Gerais repository upon publication. The complete GALAXIS/JANUS platform—including reusable orchestration, source-system integration, deployment components and functionality extending beyond this study—is not part of the public reproducibility package. Raw GAL and SIVEP-Gripe data, credentials, private mappings and other restricted operational materials will not be released. Additional study-specific materials required for editorial or peer-review assessment may be provided subject to applicable ethical, confidentiality, institutional and data-governance restrictions.

# Use of Generative Artificial Intelligence

During manuscript preparation and the development of study-specific analytical code, the first author used ChatGPT (OpenAI; GPT-5.6 Thinking) to support language editing, code review, debugging and documentation. All AI-assisted suggestions were critically reviewed, tested and modified as appropriate by the authors. Generative artificial intelligence was not used as an autonomous analytical or decision-making system. The authors performed and validated the analyses, interpreted the findings, determined the conclusions and take full responsibility for the final manuscript and associated code.

## Results

# Laboratory surveillance dataset

Eight normalized laboratory exports contributed 152,633 source rows. After restriction to records assigned to Minas Gerais, 152,609 rows remained, of which 140,207 fell within the prespecified source-data window. Deterministic sample-level consolidation removed 5,236 duplicate source rows across 5,077 duplicate groups, yielding 134,971 unique sample-level records in the JAC. After restriction to ISO epidemiological years 2021–2025, 134,730 records were retained for the primary analytical panel, with 241 boundary records excluded. Of these, 134,235 had a valid RSV outcome classification; 495 inconclusive outcomes were retained for testing-volume analyses but excluded from the positivity denominator. Among valid RSV-tested samples, 11,713 were RSV-detectable and 122,522 were RSV-non-detectable, corresponding to an overall observed positivity of 8.7%. The JAC included samples from 441 municipalities and all 14 health macroregions of Minas Gerais. The sex and descriptive age-group distributions of RSV-tested and RSV-detectable samples are provided in Supplementary Tables S1A and S1B.

Symptom-onset date was used as the epidemiological reference date for 123,806 records (91.9%), while specimen collection date was used as a fallback for 10,924 records (8.1%). No statistical imputation was performed. Variables required for the main analyses had no missing values after standardization; sex was classified as unknown or other for 85 records (0.06%).

SIVEP-Gripe enrichment provided additional clinical and epidemiological information for laboratory records with identified ILI or SARI linkage. Overall, 81,818 of 134,730 records (60.7%) had an identified linkage to at least one SIVEP-Gripe surveillance stream. This total comprised 34,975 SARI-linked records and 48,658 ILI-linked records, including 1,815 records linked to both streams. The remaining 52,912 records (39.3%) had no identified ILI/SARI linkage and were retained in the analytical dataset; this category indicates only that no corresponding notification was identified under the applied linkage rules.

# RSV testing volume, positivity and seasonality

Annual RSV testing volume increased substantially over the study period, from 19,051 samples in 2021 to 42,139 in 2025, with intermediate totals of 20,842 in 2022, 22,647 in 2023 and 30,051 in 2024. Weekly

testing volume was also higher during the later years of the study period, particularly in 2025 (Fig. 2). Weekly RSV detection counts and positivity showed marked temporal variation. The annual detection-count maximum occurred near the ISO epidemiological-year boundary in 2021 and 2022, in week 51 of 2021 and week 1 of 2022, respectively. From 2023 onward, the annual detection-count maximum occurred progressively later in the first half of the year, in ISO week 11 in 2023, week 15 in 2024 and week 21 in 2025.

Across the study period, RSV detections were concentrated predominantly between ISO weeks 10 and 25, with the highest aggregated detection counts around weeks 18–21. Lower detection counts were observed during much of the second half of the year, particularly between ISO weeks 32 and 45 (Fig. 2).

The centred 3-week moving average of RSV positivity reached annual maxima of approximately 20.6% in 2021, 19.8% in 2022, 27.7% in 2023, 20.1% in 2024 and 36.0% in 2025, with the highest annual moving-average positivity observed in 2025 (Fig. 2). Complementary analyses of testing volume, RSV detections and temporal structure are provided in Supplementary Figures S1–S3. A robustness analysis comparing three retrospective methods for defining laboratory activity periods is provided in Supplementary **Figure S4** and **Supplementary Method S3A**.

# Territorial heterogeneity in laboratory surveillance

RSV testing intensity varied across health macroregions and years (Fig. 3A). Overall, testing intensity per 100,000 residents was highest in 2025. The macroregion with the greatest testing intensity varied by year, with Central-South ranking highest in 2021–2022, Central in 2023 and 2025, and East in 2024, while other macroregions also showed year-to-year increases of varying magnitude.

Laboratory-detected RSV activity per 100,000 residents likewise varied across macroregions and years (Fig. 3B). Empirical Bayes-smoothed estimates were highest overall in 2025. The East macroregion showed the highest smoothed laboratory-detected RSV activity in that year, with elevated values also observed in selected central and northern macroregions. These indicators describe geographic heterogeneity in laboratory testing activity and laboratory-detected RSV activity and should not be interpreted as population incidence, transmission intensity, disease burden, population coverage or individual risk.

# SARI notifications and RSV laboratory testing by surveillance-stream linkage

Annual SARI notifications in Minas Gerais declined from 210,659 in 2021 to 71,822 in 2022 and 26,371 in 2023 before increasing to 27,473 in 2024 and 45,280 in 2025 (Fig. 4). Over the same period, RSV testing performed by Funed/Lacen-MG increased from 19,051 samples in 2021 to 42,139 in 2025.

Using the mutually exclusive surveillance-stream classification shown in Fig. 4, the number of RSV-tested samples assigned to the SARI-linked group increased from 4,690 in 2021 to 14,412 in 2025, while

the number classified as ILI-linked only increased from 2,358 to 15,607. Samples with no identified ILI/SARI linkage ranged from 8,011 to 12,120 annually over the study period. Weekly RSV positivity among valid RSV-tested samples showed clear seasonal variation across the three surveillance-stream categories, with differences in the timing and magnitude of annual peaks (Fig. 4).

# Observed age-specific RSV positivity

Weekly RSV positivity was first examined descriptively by age group before the model-based analysis of peak timing. Positivity was calculated as the number of RSV-detectable samples divided by the number of valid RSV-tested samples within each age stratum and summarized using centred 3-week moving averages. Children younger than 5 years showed higher laboratory-observed RSV positivity than older age groups, with recurrent seasonal peaks across the study period. Positivity was lower overall among adult and older-adult groups, although clear seasonal increases were also observed during periods of greater RSV activity. These age-specific positivity trajectories are shown in Fig. 5.

# Bayesian age-specific peak-timing analysis

To formally compare annual RSV positivity peak timing across age groups while accounting for uncertainty, Bayesian modelling was applied to the same five age groups shown in Fig. 5. The analysis included 134,235 valid RSV-tested samples and 11,713 RSV-detectable samples from the JAC.

Estimated peak timing varied across years. Estimates for 2021 were sensitive to the epidemiological-year boundary, with some age groups reaching their estimated peaks near the end of the ISO epidemiological year. In 2022, children younger than 5 years peaked approximately 4 weeks earlier than adults aged 18–59 and 60–79 years, with posterior probabilities of earlier peak timing of 1.000 and 0.998, respectively. From 2023 onward, the posterior median peak among children younger than 5 years occurred on 20 February in 2023, 15 April in 2024 and 5 May in 2025, preceding the estimated peaks of adult and older-adult groups by approximately 1–3 weeks. Posterior probabilities that children younger than 5 years peaked earlier than these groups ranged from 0.974 to 1.000 in 2023, 0.989 to 0.999 in 2024 and 0.935 to 0.988 in 2025. Differences relative to the 5–17-year age group were smaller, with median lead times of 0–1 week and posterior probabilities of earlier peak timing increasing from 0.481 in 2022 to 0.877 in 2025.

## Discussion

In this large subnational analysis, routine RSV testing records generated within the public health laboratory network provided complementary surveillance evidence when interpreted using valid tested-sample denominators, epidemiological timing, territorial context and SIVEP-Gripe clinical and epidemiological enrichment. Across 2021–2025 in Minas Gerais, this approach characterized post-pandemic changes in laboratory-observed RSV seasonality, substantial heterogeneity in testing intensity

across health macroregions, and earlier estimated positivity peaks among children younger than 5 years than among adult and older-adult groups from 2023 onward.

The principal contribution of this study is the operational demonstration that routinely generated laboratory records within SISLAB can support a complementary analytical layer for surveillance when they are standardized, integrated and interpreted through a modular and auditable workflow. The value of the approach lies in the complete analytical process—from laboratory data engineering and construction of valid tested-sample denominators to clinical and epidemiological enrichment, territorial contextualization and downstream analyses—rather than in any single indicator considered in isolation. By preserving the laboratory sample as the analytical backbone while incorporating information from complementary surveillance sources, this implementation illustrates how existing public health laboratory infrastructure can support a more systematic and extensible use of routinely generated data for epidemiological interpretation and public health surveillance.

The analytical architecture implemented here is consistent with international recommendations for integrated respiratory virus surveillance, which emphasize virological testing, valid testing denominators, epidemiological timing, clinical context, standardized laboratory–epidemiological indicators and linkage across complementary surveillance systems[14,17]. In this study, these principles were operationalized by combining laboratory data engineering, territorial contextualization and SIVEP-Gripe enrichment, when linkage was available, within a single auditable workflow. GALAXIS therefore provided the computational framework through which the DASI approach was implemented at subnational scale while remaining complementary to established epidemiological surveillance systems.

The temporal patterns observed in Minas Gerais are consistent with the disruption and subsequent reorganization of RSV seasonality described following the COVID-19 pandemic[8,9]. Annual detection count peaks occurred near the ISO epidemiological-year boundary in 2021–2022 and shifted progressively into the first half of the year thereafter, occurring in weeks 11, 15 and 21 in 2023, 2024 and 2025, respectively. Together with the corresponding seasonal variation in positivity among tested samples, these findings suggest an evolving post-pandemic seasonal pattern rather than an immediate return to a fixed calendar of RSV activity. These trends should nevertheless be interpreted within the context of laboratory surveillance, because changes in testing volume, referral pathways, and surveillance practices may also have influenced the observed temporal patterns. Substantial territorial heterogeneity was also evident across health macroregions. Using 2024 municipal population estimates from IBGE as fixed denominators across the study period, we observed marked variation in both RSV testing intensity and laboratory-detected RSV activity per 100,000 residents[15]. These geographic differences likely reflect a combination of RSV activity among tested samples and variation in referral pathways, surveillance participation, access to public health laboratory testing, network organization and clinical testing practices. Accordingly, these indicators should not be interpreted as measures of population incidence, transmission intensity or individual-level risk. Rather, they provide a standardized view of where the public health laboratory system captured relatively greater or lower testing activity and

RSV detections in relation to population size, allowing geographic differences in laboratory surveillance activity to be identified and compared.

The contrasting trajectories of SARI notifications and RSV laboratory testing illustrate the value of incorporating the laboratory-tested denominator. SARI notifications declined markedly after 2021, whereas the number of samples tested for RSV by Funed/Lacen-MG increased across the study period. This divergence indicates that the volume and composition of public health laboratory testing cannot be inferred from SARI notifications alone. The integrated dataset contained samples linked to SARI surveillance, samples linked to ILI surveillance, and samples for which no corresponding ILI/SARI linkage was identified, highlighting the heterogeneous surveillance and referral pathways through which specimens entered the public health laboratory system. Integrating laboratory records with SIVEP-Gripe therefore allows testing patterns to be interpreted alongside, rather than as a substitute for, established syndromic surveillance[14,17].

Age-specific analyses added an important temporal dimension to the laboratory surveillance data. Children younger than 5 years showed higher observed RSV positivity than older age groups and, from 2023 onward, their estimated positivity peaks preceded those of adult and older-adult groups by approximately 1–3 weeks. Previous studies using laboratory-confirmed respiratory virus data have similarly identified age-specific differences in the timing of influenza and RSV activity[18]. In the present study, Bayesian modelling allowed these temporal differences to be quantified together with their associated uncertainty rather than relying solely on visual comparison of epidemic curves. The observed ordering is also epidemiologically plausible given the high burden of RSV infection in young children[3,4], but the analysis does not establish transmission direction or identify the mechanisms underlying differences in peak timing. Nevertheless, age-stratified laboratory positivity trajectories may provide useful temporal information for surveillance, particularly if their potential value as an early indicator of broader RSV activity is evaluated prospectively.

The modular architecture of GALAXIS used in this study also provides a basis for progressively incorporating additional data layers at both the sample and contextual levels. By preserving laboratory sample identity while enriching records with available clinical, epidemiological, temporal and territorial information, the same structure could support future integration of pathogen genomic data, allowing viral genetic diversity and lineage dynamics to be examined alongside the epidemiological context of the corresponding samples. The framework could also incorporate population, socioeconomic and environmental data linked by geographic area and epidemiological period, including live-birth indicators and meteorological variables such as temperature, precipitation and humidity. Such extensions could support future analyses of how demographic, social and environmental conditions are associated with temporal and territorial patterns of laboratory-observed respiratory virus activity, while allowing the analytical workflow to evolve as new surveillance questions and data sources emerge.

The introduction of maternal RSV vaccination into Brazil's public health system further increases the relevance of age-specific RSV surveillance[19]. Because the present study covers 2021–2025, the

laboratory data provide a pre-implementation baseline for Minas Gerais against which subsequent changes in RSV positivity, age-specific patterns, seasonal timing and laboratory testing demand can be evaluated. Continued application of the analytical framework may therefore support monitoring of RSV epidemiology as preventive interventions are incorporated into routine public health practice.

Several limitations should be considered. This study used routinely generated public health laboratory data rather than a population-based sampling design; therefore, RSV positivity and population-standardized laboratory indicators describe activity among tested samples and should not be interpreted as population incidence, transmission intensity, or disease burden. Temporal and territorial comparisons may also have been influenced by changes in testing practices, referral pathways, surveillance priorities, and access to laboratory testing. Health macroregion assignment was based on the requesting municipality rather than the patient's municipality of residence. No corresponding ILI/SARI notification was identified for 52,912 of 134,730 records (39.3%) under the applied linkage rules, limiting clinical and epidemiological enrichment for these records. In addition, symptom-onset date was unavailable for 10,924 records (8.1%), for which specimen collection date was used as the epidemiological reference-date fallback, potentially introducing temporal misclassification. Furthermore, the analytical unit was the laboratory sample rather than the individual, so multiple samples could represent the same person or clinical episode. Finally, Bayesian peak-timing estimates describe differences in laboratory-observed positivity across age groups but do not establish transmission direction, and estimates near epidemiological-year boundaries, particularly in 2021, require cautious interpretation.

# Conclusion

Routine RSV testing data generated by Funed/Lacen-MG provided complementary surveillance evidence when standardized and integrated with SIVEP-Gripe clinical-epidemiological records and territorial and population data, using valid tested-sample denominators and epidemiological timing. Across Minas Gerais, this approach characterized post-pandemic shifts in the timing of annual RSV detection-count peaks, geographic heterogeneity in laboratory testing and laboratory-detected RSV activity, and higher positivity and earlier estimated peaks among children younger than 5 years than among adult and older-adult groups. The contrasting trajectories of SARI notifications and RSV testing, together with the substantial proportion of laboratory records without identified ILI/SARI linkage, showed that public health laboratory activity encompassed heterogeneous surveillance, diagnostic, hospital, and referral pathways not fully represented by syndromic surveillance records alone. From 2023 onward, estimated positivity peaks among children younger than 5 years preceded those among adult and older-adult groups by 1–3 weeks; the potential value of this pattern for seasonal preparedness requires prospective evaluation against reporting delays, hospital demand, and intervention timelines. More broadly, this large subnational implementation demonstrates how integrating routinely generated data from complementary sources can support a systematic analytical function within laboratory surveillance while strengthening, rather than replacing, established epidemiological surveillance systems. The same analytical principles may be adaptable to other pathogens and geographic settings, subject to local data availability, prospective validation, and workflow-specific requirements.

**Addendum:** GRIPE-MG Team

The GRIPE-MG Team is a collaborative laboratory surveillance group within the Minas Gerais public health laboratory network. Cássia de Miranda Santos, Lívia Gomes do Nascimento, Mauro Andrade de Freitas Guimarães, Eduarda Emanuela Maia Alves, Paulo Eduardo de Souza da Silva, Antônio Carlos Gomes Júnior, and Ludmila Oliveira Lamounier contributed to respiratory virus diagnostic activities, laboratory data generation, and data validation that supported this study.

# Declarations


### Acknowledgements

The authors acknowledge FAPEMIG, Fundação Ezequiel Dias (Funed), Secretaria de Estado de Saúde de Minas Gerais (SES-MG), the Brazilian Ministry of Health, CGLAB and Fundação Oswaldo Cruz (Fiocruz) for institutional and technical support. The authors also thank Glauco C. Pereira, the Biological Samples Service (SGAB), the sentinel ILI units and all healthcare and surveillance services that routinely submitted respiratory specimens to Funed/Lacen-MG. J.L. acknowledges institutional support from the Fundação para a Ciência e a Tecnologia (FCT) to the Católica Biomedical Research Centre (CBR) (DOI: 10.54499/UID/06497/2025; UID/06497/2025).


### Author Contributions

A.F.L.B.: Conceptualization; Methodology; Software; Data curation; Formal analysis; Investigation; Project administration; Visualization; Writing – original draft; Writing – review and editing.

B.L.O.L.A.: Data curation; Validation; Writing – review and editing.

G.R.: Validation; Writing – review and editing.

J.F.A.: Validation; Writing – review and editing.

J.B.P.: Validation; Writing – review and editing.

R.S.C.: Validation; Writing – review and editing.

E.L.A.M.: Validation; Writing – review and editing.

P.A.A.: Writing – review and editing.

M.F.C.G.: Writing – review and editing.

R.G.A.: Writing – review and editing.

W.A.F.A.: Writing – review and editing.

M.T.F.P.L.: Writing – review and editing.

F.C.M.I.: Writing – review and editing.

M.G.: Validation, Writing – review and editing.

L.C.J.A.: Supervision; Project administration; Writing – review and editing.

J.L.: Conceptualization; Methodology; Validation; Supervision; Writing – review and editing.

All authors reviewed and approved the final version of the manuscript.

**Funding**

This study used routine institutional laboratory surveillance infrastructure from Funed/Lacen-MG. André Felipe Leal Bernardes was supported by a doctoral scholarship from FAPEMIG (grant BPG-00063-24). No specific external funding was received exclusively for the conduct of this study. FAPEMIG had no role in the study design, data collection, data analysis, data interpretation, preparation of the manuscript, or decision to submit the manuscript for publication.

**Competing interests**: The authors declare no competing interests.

**Data Availability**: A de-identified, analysis-ready version of the JANUS Analytical Core, containing the variables required to reproduce the aggregate analyses reported in this study, is publicly available in the RSV-DASI-Minas-Gerais GitHub repository, together with a data dictionary and the aggregated analytical outputs underlying the reported figures and results. Direct identifiers, operational laboratory identifiers, exact dates, and variables presenting a potential risk of re-identification are not included in the shared dataset. The underlying individual-level GAL and SIVEP-Gripe records will not be publicly available because they originate from routine public health surveillance and are subject to ethical, legal, institutional, and data-governance restrictions. The 2024 municipal population estimates used for population-standardized territorial analyses are publicly available from the Brazilian Institute of Geography and Statistics (IBGE).

**Permission to Reproduce Material from Other Sources:** No material from other sources was reproduced in this manuscript.

# References

1. Mazur, N. I., Caballero, M. T. & Nunes, M. C. Severe respiratory syncytial virus infection in children: burden, management, and emerging therapies. *The Lancet* **404**, 1143–1156 (2024).
2. Ciapponi, A. *et al.* Respiratory syncytial virus disease burden in children and adults from Latin America: a systematic review and meta-analysis. *Front. Public Health* **12**, 1377968 (2024).

3. Li, Y. *et al.* Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis. *The Lancet* **399**, 2047–2064 (2022).
4. Kenmoe, S. & Nair, H. The disease burden of respiratory syncytial virus in older adults. *Current Opinion in Infectious Diseases* **37**, 129–136 (2024).
5. Pinhata, J. M. W. *et al.* Rapid response of a public health reference laboratory to the COVID-19 pandemic. *Journal of Medical Microbiology* **72**, 001757 (2023).
6. Barberia, L. G. *et al.* An assessment of the public health surveillance strategy based on molecular testing during three major pandemic waves of COVID-19 in Brazil. *PLOS Glob Public Health* **3**, e0002164 (2023).
7. Pabbaraju, K. *et al.* A Public Health Laboratory Response to the Pandemic. *J Clin Microbiol* **58**, e01110-20 (2020).
8. Lastrucci, V. *et al.* Seasonality and severity of respiratory syncytial virus during the COVID-19 pandemic: a dynamic cohort study. *International Journal of Infectious Diseases* **148**, 107231 (2024).
9. Hamid, S. *et al.* Seasonality of Respiratory Syncytial Virus — United States, 2017–2023. *MMWR Morb. Mortal. Wkly. Rep.* **72**, 355–361 (2023).
10. Ministério da Saúde. *Guia de vigilância integrada da covid-19, influenza e outros vírus respiratórios de importância em saúde pública*. (Ministério da Saúde, Brasília, DF, 2024).
11. Brasil. Ministério da Saúde. Portaria de Consolidação nº 4, de 28 de setembro de 2017: Consolidação das normas sobre os sistemas e os subsistemas do Sistema Único de Saúde. (2017).
12. Paula, F. J. D. *et al.* Sistema Gerenciador de Ambiente Laboratorial - GAL: Avaliação de uma ferramenta para a vigilância sentinela de síndrome gripal, Brasil, 2011-2012*. *Epidemiologia e Serviços de Saúde* **26**, 339–348 (2017).
13. Brasil. Ministério da Saúde. Sistema Nacional de Laboratórios de Saúde Pública. https://www.gov.br/saude/pt-br/composicao/svsa/sislab/sistema-nacional-de-laboratorios-de-saude-publica-sislab/.
14. World Health Organization. Implementing the integrated sentinel surveillance of influenza and other respiratory viruses of epidemic and pandemic potential by the Global Influenza Surveillance and Response System. (2024).
15. Instituto Brasileiro de Geografia e Estatística. Estimativas da população residente para os municípios e para as unidades da Federação com data de referência em 1º de julho de 2024. (2024).
16. Fundação Ezequiel Dias. MANUAL DE ORIENTAÇÕES PARA O ENVIO DE AMOSTRAS BIOLÓGICAS PARA A FUNED. (2025).
17. Teirlinck, A. C. *et al.* Recommendations for respiratory syncytial virus surveillance at the national level. *Eur Respir J* **58**, 2003766 (2021).

18. Yang, L. *et al.* Age-specific epidemic waves of influenza and respiratory syncytial virus in a subtropical city. *Sci Rep* **5**, 10390 (2015).
19. Ministério da Saúde. Estratégia de vacinação contra o vírus sincicial respiratório em gestantes. (2026).

# Figures

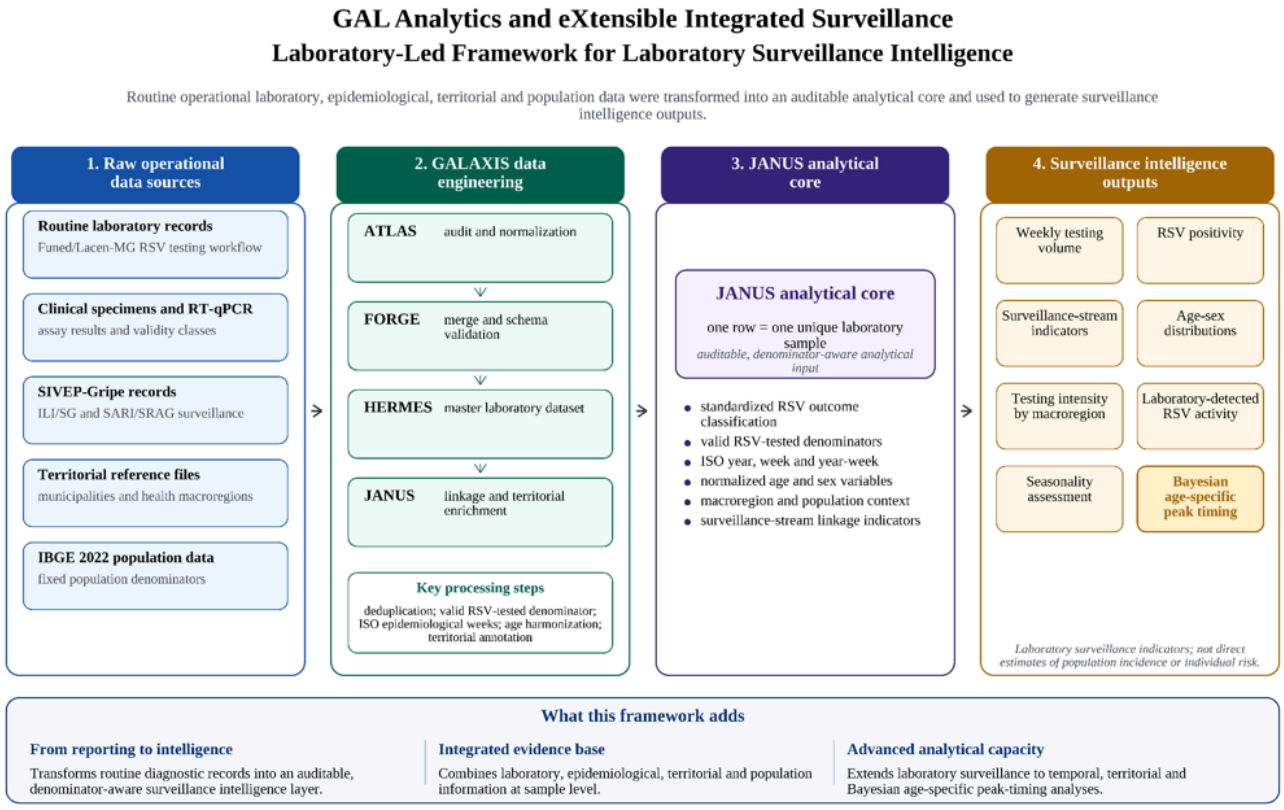
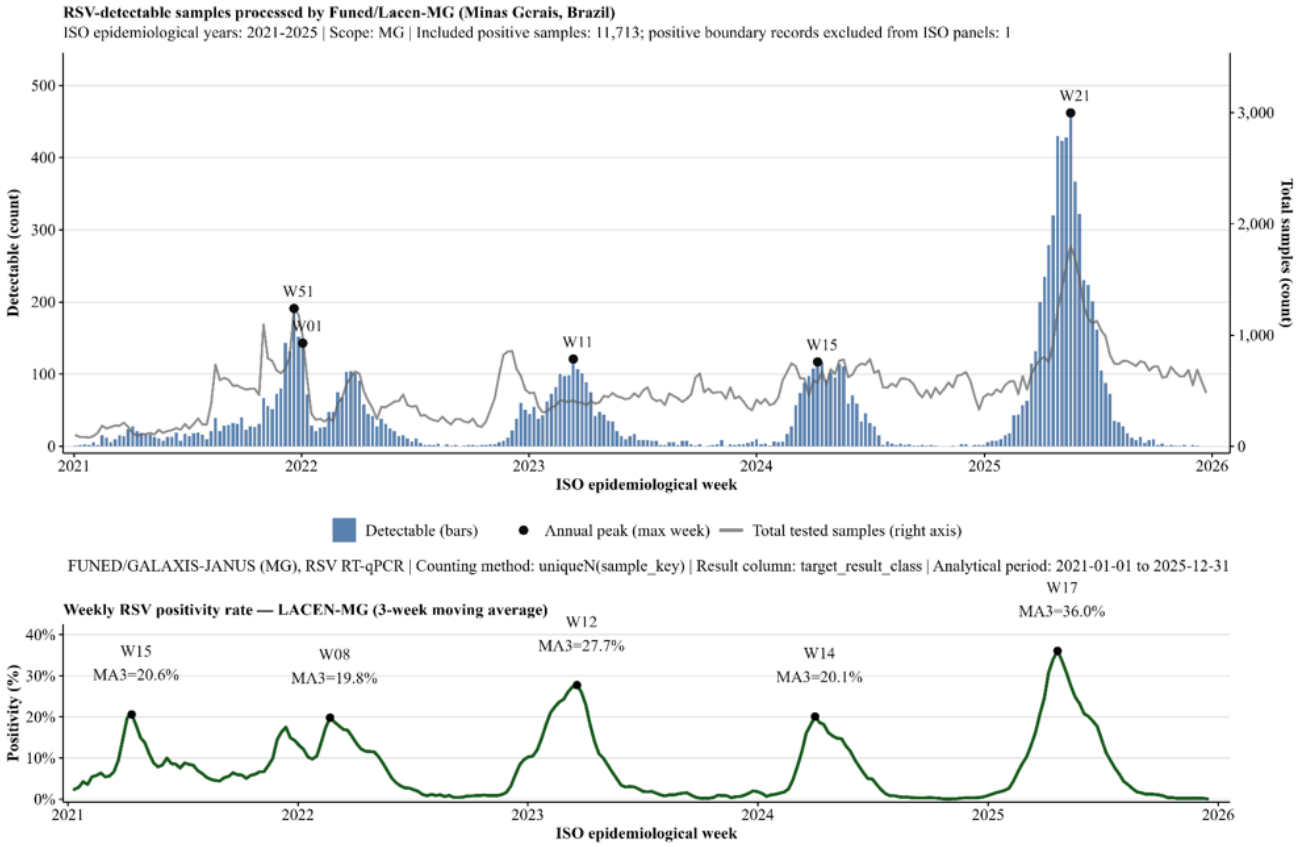



**Figure 1**

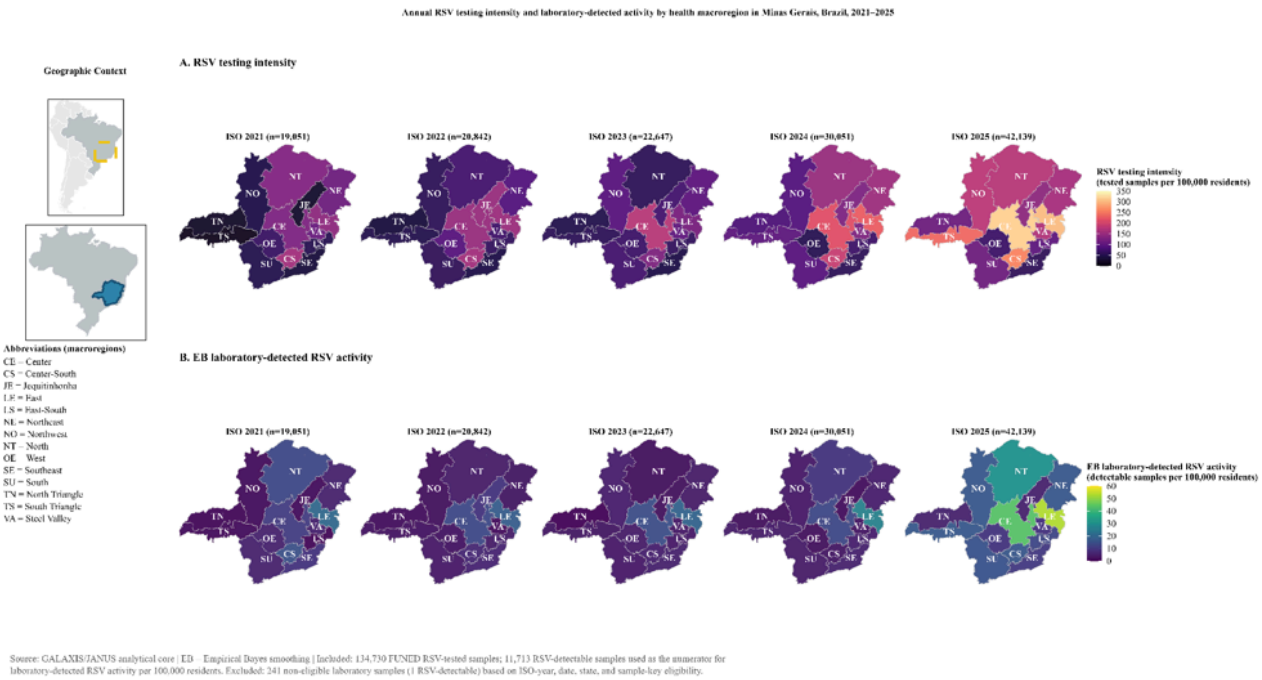
**GALAXIS data-integration and engineering framework for RSV laboratory surveillance intelligence.** Routine laboratory, epidemiological, territorial and population data from Minas Gerais, Brazil, were processed through GALAXIS under the DASI approach. Within GALAXIS, JANUS integrated the harmonized laboratory records with SIVEP-Gripe clinical and epidemiological information when linkage was available and generated the JAC, an auditable sample-level analytical dataset. The JAC contained standardized RSV outcomes, valid testing denominators, ISO epidemiological timing, demographic and territorial variables, and SIVEP-Gripe linkage indicators when available. For population-standardized territorial analyses, JAC-derived macroregional data were combined with 2024 municipal population estimates from IBGE within the GALAXIS analytical workflow. These integrated data supported the denominator-aware surveillance intelligence (DASI) layer, comprising descriptive, territorial, seasonality

and Bayesian peak-timing analyses interpreted as laboratory surveillance indicators rather than measures of incidence, transmission or individual risk.



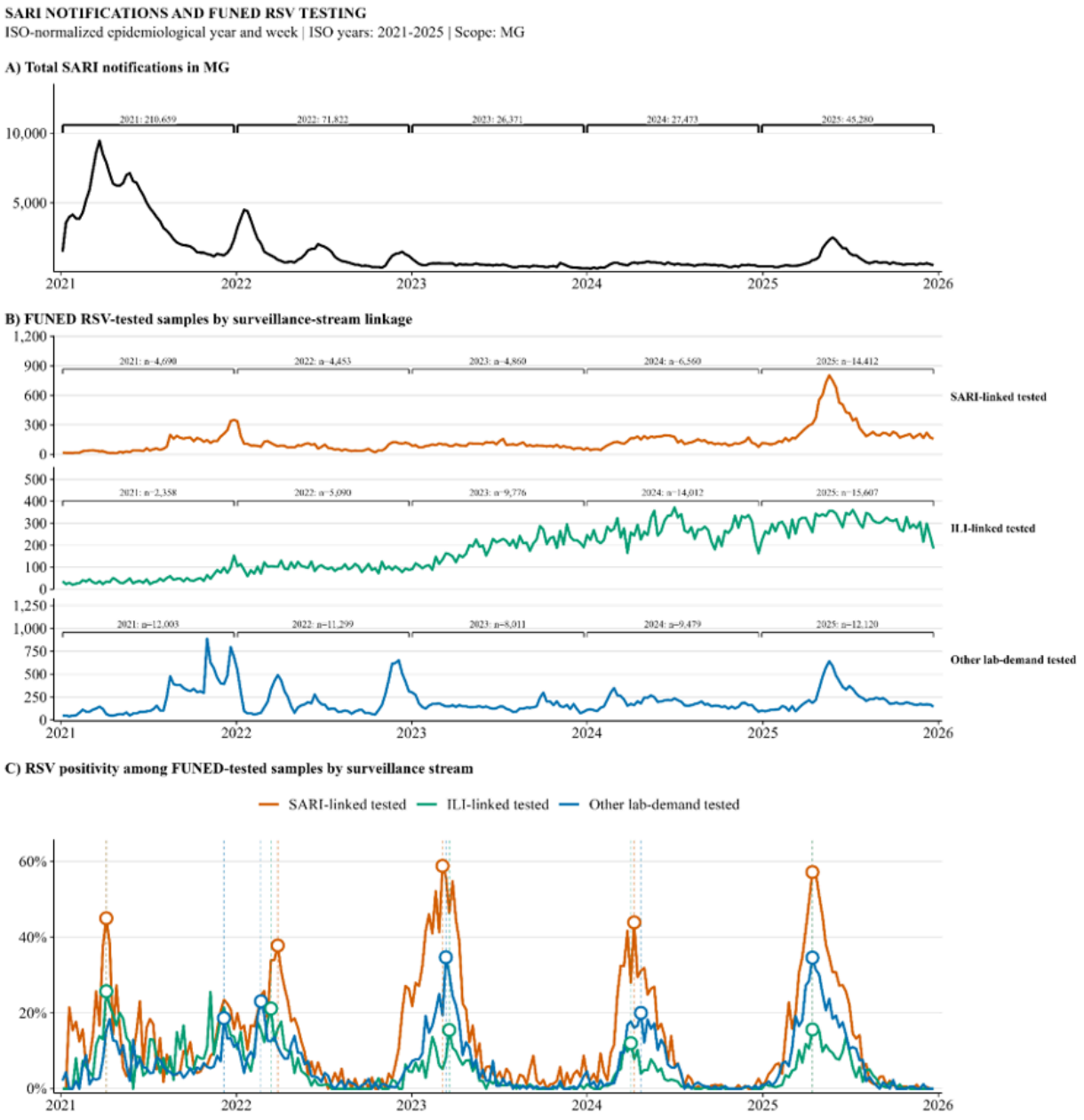

## Figure 2

**Weekly RSV testing, RSV detections and RSV positivity in Minas Gerais, Brazil, 2021–2025.** The upper panel shows unique RSV-detectable samples as bars and total laboratory samples tested for RSV within the ISO-week analytical panel as a grey line on the right axis; black points indicate annual peak detection weeks. The lower panel shows the centred 3-week moving average of RSV positivity among valid tested samples, with the annual maxima of the moving-average positivity annotated. The figure shows annual detection-count peaks near the ISO epidemiological-year boundary in 2021–2022 and in the first half of the year from 2023 onward. Counts and positivity represent laboratory-observed activity among tested samples, not population incidence, disease burden or individual-level risk.



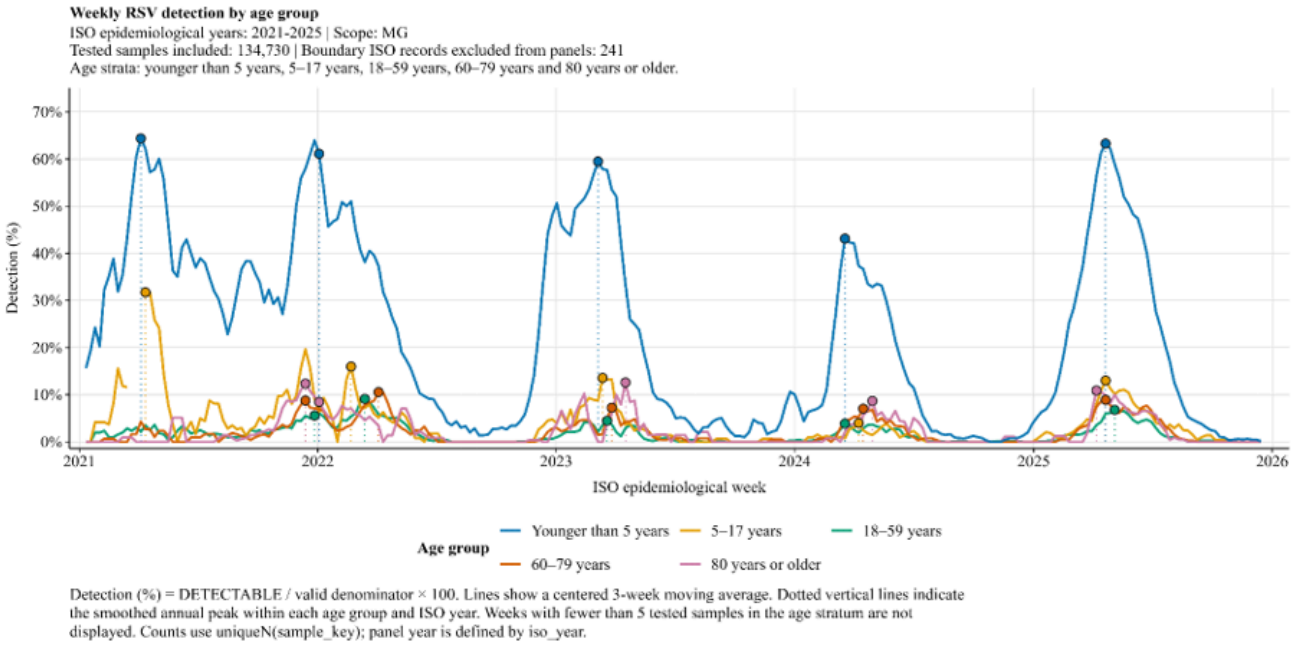

## Figure 3

**Annual RSV testing intensity and laboratory-detected activity by health macroregion in Minas Gerais, Brazil, 2021–2025.** (A) Distinct laboratory samples tested for RSV within the ISO-week analytical panel per 100,000 residents by health macroregion and ISO epidemiological year. (B) Empirical Bayes-smoothed RSV-detectable samples per 100,000 residents by health macroregion and ISO epidemiological year. Annual sample totals are indicated for each year. These indicators describe Funed/Lacen-MG laboratory surveillance activity and should not be interpreted as direct measures of RSV incidence, transmission intensity or disease burden. EB, empirical Bayes; RSV, respiratory syncytial virus**.**



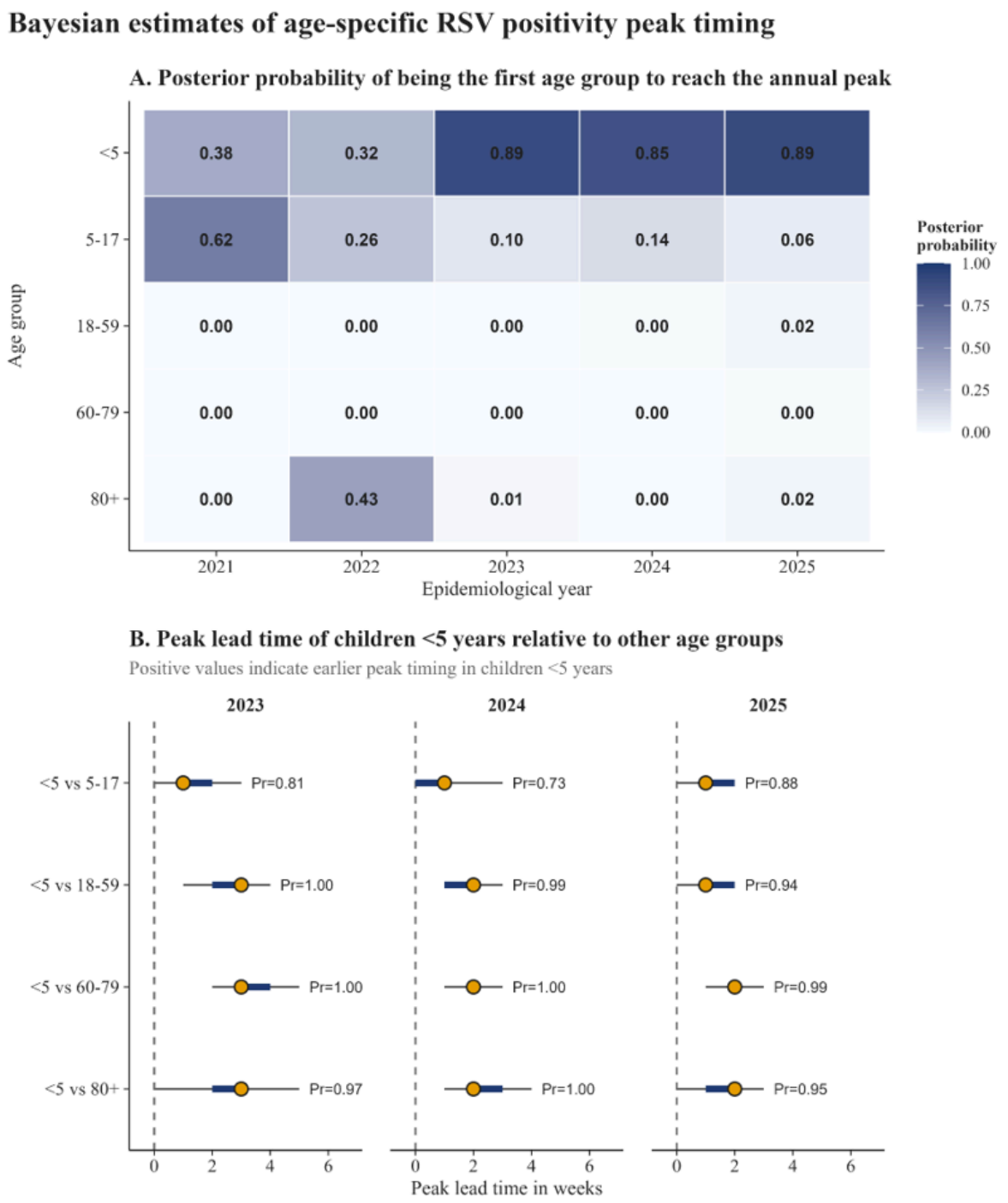

## Figure 4

**SARI notifications and Funed/Lacen-MG RSV testing by surveillance stream, Minas Gerais, Brazil, 2021–2025.** (A) Weekly SARI notifications in Minas Gerais, with annual totals. (B) Weekly RSV-tested samples processed by Funed/Lacen-MG according to mutually exclusive surveillance-stream categories: SARI-linked, ILI-linked only, and no identified ILI/SARI linkage. Samples linked to both SARI and ILI were assigned to the SARI-linked category to avoid double-counting. Annual stream-specific totals are shown above each panel. (C) Weekly RSV positivity among valid RSV-tested samples by surveillance-stream category. Positivity was calculated as RSV-detectable samples divided by valid RSV-tested samples, excluding inconclusive outcomes. Points indicate annual peak positivity among weeks with at least 10 valid RSV-tested samples within each stream.




## Figure 5

**Observed weekly RSV positivity by age group in Minas Gerais, Brazil, 2021–2025.** Weekly RSV positivity was calculated as the number of RSV-detectable samples divided by the number of valid RSV-tested samples within each age group. Lines show centred 3-week moving averages by ISO epidemiological week; weeks with fewer than five valid tested samples within an age stratum were not displayed. These trajectories describe laboratory-observed positivity among tested samples and do not represent population incidence.




## Figure 6

**Bayesian estimates of age-specific RSV positivity peak timing in Minas Gerais, Brazil, 2021–2025.** (A) Posterior probability that each age group reached the annual RSV positivity peak first among tested samples. (B) Posterior peak lead time for children younger than 5 years compared with other age groups in 2023–2025. Points indicate posterior medians; thick and thin intervals indicate 50% and 95% credible intervals. Estimates represent temporal differences in laboratory-observed RSV positivity, not causal transmission between age groups.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- JACPUBLIC.csv
- SupplementaryInformationMG.docx