

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	All data used in this study are publicly available. Weekly influenza-like illness (ILI) surveillance data were obtained from the CDC FluView Interactive dashboard (https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html). Respiratory syncytial virus (RSV) emergency department visit data were obtained from the CDC National Syndromic Surveillance Program (https://data.cdc.gov). Influenza hospitalization data were obtained from the CDC FluSight challenge (https://data.cdc.gov). Google Trends digital behavioral signal data were obtained through the Google Trends API (https://trends.google.com). Code and sample data to reproduce the results presented in this article are available at https://github.com/MIGHTE-lab/Respiratory-EWS-Sample . Source data are provided with this paper.
Data analysis	A sample of code and data to reproduce results is available at github.com/MIGHTE-lab/Respiratory-EWS-Sample and archived on Zenodo at https://doi.org/10.5281/zenodo.19371482 . Source data for all main figures are provided with this paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data used in this study are publicly available. Weekly influenza-like illness (ILI) surveillance data were obtained from the CDC FluView Interactive dashboard (<https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>). Respiratory syncytial virus (RSV) emergency department visit data were obtained from the CDC National Syndromic Surveillance Program (<https://data.cdc.gov>). Influenza hospitalization data were obtained from the CDC FluSight challenge (<https://data.cdc.gov>). Google Trends digital behavioral signal data were obtained through the Google Trends API (<https://trends.google.com>). Code and sample data to reproduce the results presented in this article are available at <https://github.com/MIGHTE-lab/Respiratory-EWS-Sample>. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender N/A

Reporting on race, ethnicity, or other socially relevant groupings N/A

Population characteristics N/A

Recruitment N/A

Ethics oversight N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative, observational study using a machine learning early warning system. The study integrates time-series data from digital behavioral signals (Google Trends) and traditional epidemiological surveillance to prospectively predict respiratory disease outbreak onsets and peaks across all 50 U.S. states.
Research sample	Existing population-level surveillance datasets from multiple sources: CDC ILI reports, CDC influenza hospitalization data, National Syndromic Surveillance Program (NSSP) RSV emergency department visits, and COVID-19 activity data (used for transfer learning validation), covering all 50 U.S. states at the weekly level. The sample is anonymized, aggregated, and publicly available public health surveillance data; no demographic breakdown of individuals is applicable.
Sampling strategy	N/A
Data collection	Data were collected from three existing institutional sources: CDC FluView (ILI), CDC FLUsight (influenza hospitalizations), and NSSP (RSV). Google Trends data were accessed via the Google Trends API. No human participants or instruments were involved. Researchers were not blind to hypotheses, as this is a computational modeling study.
Timing	Timing ILI data: October 2010 – April 2025 Influenza hospitalization data: July 2021 – May 2025 RSV data: October 2022 – March 202

COVID-19 data: early 2020 (first outbreak onset March 2020, used for transfer learning validation)
Google Trends data: October 2010 – May 2025

Data exclusions

Data Exclusions

Florida was excluded from ILI analyses due to unavailable ILINet data. Missouri was excluded from RSV analyses due to unavailable NSSP data. These exclusions were data-driven, not pre-established experimental criteria. ILI data from 2020-06-29 to 2022-06-26 were excluded due to reporting anomalies associated with early stages of the COVID-19 pandemic.

Non-participation

N/A

Randomization

N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A