pagepress

Advancements in Health Research

Advancements in Health Research 2026; 3:217
doi:10.4081/ahr.2026.217

©Copyright: the Author(s), 2026
Licensee PAGEPress, Italy

Temporal and seasonal patterns of online interest in ibuprofen across fourteen European countries: a five-year infodemiological study with robust time-series analysis and chronological predictive validation

Matteo Luigi Giuseppe Leoni,[1,2] Marco Mercieri,[2] Y Van Tran,[1,3] Phong Van Pham,[1,3] Giacomo Farì,[1,4] Piercarlo Sarzi Puttini,[1,5,6] Jessica Nicolas,[1,7,8] Giustino Varrassi[1,9]

[1]Visionary International Board for Research and Analgesia (VIBRA), Fondazione Paolo Procacci, Rome, Italy; [2]Department of Medical and Surgical Sciences and Translational Medicine, Sapienza University of Rome, Italy; [3]Department of Anesthesiology, Tam Anh General Hospital, Ho Chi Minh City, Vietnam; [4]Department of Experimental Medicine (Di.Me.S.), University of Salento, Lecce, Italy; [5]Department of Biomedical and Clinical Sciences, University of Milan, Italy; [6]Rheumatology Unit, IRCCS Ospedale Galeazzi-Sant'Ambrogio, Milan, Italy; [7]Chronic Pain Center, Paris Saint-Joseph Hospital, Paris, France; [8]Department of Rheumatology, Luarasi University, Tirana, Albania; [9]College of Medicine, University of Baghdad, Iraq

Abstract

Ibuprofen is widely used as a non-prescription analgesic and antipyretic, yet online information-seeking remains incompletely characterized. We assessed trends, seasonality, adjusted cross-country co-movement and out-of-sample predictability of Google search attention across 14 European countries. Monthly Google Trends relative search volume (RSV; 0-100) for the harmonized generic search term "ibuprofen" was extracted on 20 July using the Web Search property and all categories, each country being queried separately so that every series is normalized within country (July 2021 to June 2026; 60 months per country). Trend inference used calendar-month-adjusted linear regression with Newey-West standard errors and Benjamini-Hochberg false-discovery-rate correction. Seasonality was evaluated using robust STL decomposition and winter-versus-summer contrasts. Cross-country co-movement was assessed on STL remainders. Seasonal ARIMA forecasts were validated by rolling origins against seasonal naive. Random Forest and Gradient Boosting were tested on a final 12-month chronological holdout and compared with harmonic regression and seasonal naive. Across 840 country-month observations, eight countries showed significant increases: Hungary, Poland, Turkey, Sweden, Belgium, the United Kingdom, the Czech Republic and the Netherlands; none showed a significant decline. Seasonal peaks fell between December and February in all 14 countries (December in 10) and troughs between April and July (June in 7); seasonal strength ranged from 0.20 (Norway) to 0.72 (Poland). After removing trend and seasonality, the strongest residual correlations were Germany-Austria (r=0.79) and Poland-Czech Republic (r=0.76). ARIMA outperformed seasonal naive only at 1 month (RMSE 10.56 *vs* 11.18), but not at 3, 6, or 12 months. On the chronological holdout, harmonic regression performed best (R^2=0.242; RMSE=1.221); Random Forest (R^2=-0.026; RMSE=1.420) and Gradient Boosting (R^2=-0.030; RMSE=1.423) did not outperform seasonal naive (R^2=-0.005; RMSE=1.406). Ibuprofen search attention increased in a subset of countries and showed recurrent winter seasonality. Predictive performance was horizon-dependent and modest. Google Trends normalization means findings reflect within-country search attention rather than absolute cross-country interest or medication use.

Key words: ibuprofen; non-steroidal anti-inflammatory drugs; infodemiology; Google Trends; digital epidemiology; machine learning.

Correspondence to: Matteo Luigi Giuseppe Leoni, Department of Medical and Surgical Sciences and Translational Medicine, Sapienza University of Rome, 00189 Rome, Italy.
E-mail: matteolg.leoni@gmail.com

Introduction

Ibuprofen is one of the most widely used non-steroidal anti-inflammatory drugs (NSAIDs) worldwide.[1] Available over the counter in many countries for the relief of pain, inflammation and fever, it is frequently obtained without a prescription and used through self-medication.[2,3] Prescription databases therefore provide an incomplete view of total exposure. Internet-search data cannot directly quantify this unrecorded use, but they can characterize how population-level information-seeking about ibuprofen changes over time.

NSAID use is clinically relevant because ibuprofen and related agents are associated with dose- and patient-dependent gastrointestinal and cardiovascular adverse effects and with clinically important drug interactions.[4,5] These safety considerations provide a rationale for studying public attention to commonly self-medicated analgesics. Search activity, however, should be interpreted

OPEN ACCESS

as an information-seeking signal rather than as a surrogate for medication consumption, purchase, adherence, inappropriate use, or adverse events unless those relationships are validated against independent data.

Aggregated internet search activity has emerged as a complementary source for infodemiology and infoveillance, where health-related queries are used to study population attention and information-seeking.[6,7] Search-based indicators have been evaluated across health domains, including influenza activity, cancer screening and medication-related interest,[8-11] but their interpretation is limited by algorithm changes, sampling variability, normalization and instability.[12] Google Trends is widely used because it is publicly available, temporally granular and frequently updated, while its methodological limitations are increasingly well characterized.[13,14]

Despite ibuprofen's ubiquity, the multi-country temporal structure of online search attention to this specific drug has not been well characterized. Important questions include whether within-country attention is changing over time, whether seasonal patterns recur, whether countries share residual short-term fluctuations after trend and seasonality are removed, and whether future normalized search attention can be predicted out of sample. These questions are particularly relevant given the continued growth of online health-information seeking across Europe.[15,16]

We therefore analyzed five years of monthly Google Trends data for the search term "ibuprofen" across fourteen selected European countries. Our objectives were to quantify within-country secular trends with correction for seasonality, serial dependence and multiplicity; characterize the strength and recurrence of seasonality; assess cross-country co-movement after detrending and deseasonalizing; validate seasonal autoregressive integrated moving average (ARIMA) forecasts against a simple benchmark; and evaluate whether nonlinear machine-learning models added predictive value beyond structured temporal benchmarks.

Because reproducibility has been identified as a recurrent weakness of search-based health research, we additionally report the complete Google Trends extraction protocol, an explicit rationale for the choice of search term and for the composition of the country panel, and an explicit account of the sampling variability inherent in the platform.[17,18]

Methods

Study design and reporting

This was a retrospective, observational infodemiological study of aggregated, de-identified and publicly available internet search data; no individual-level data were accessed and no human participants were involved. Reporting follows the methodology framework proposed for the use of Google Trends in infodemiology and infoveillance,[13,18] and the full set of query parameters required for exact replication is given in Supplementary Table S1.

Data source and extraction protocol

Google Trends (https://trends.google.com) reports relative search volume (RSV) on a 0-100 scale, where 100 denotes the peak within the specified query, geography and time window.[13] We analyzed monthly RSV for the search term "ibuprofen" in France, Germany, Italy, Spain, Turkey, the United Kingdom, the Netherlands, Poland, the Czech Republic, Belgium, Norway, Sweden, Hungary and Austria from July 2021 through June 2026 (60 months per country; 840 country-month observations).

All series were downloaded from the public Google Trends interface (https://trends.google.com/trends/explore) on 20 July, within a single session and by one investigator, using query settings that were held constant for every country: search property, Web Search (Image, News, Shopping and YouTube search were not used); category, All categories (category identifier 0), so that no algorithmic topic filter was imposed; query type, the free-text search term "ibuprofen" rather than a Google Trends Topic; geography, one country per query, specified by ISO 3166-1 alpha-2 code; time range, 1 July 2021 to 30 June 2026; and interface language English (hl=en-GB), which governs the interface only and does not restrict the queries retrieved. No sub-national, city, age or device filters were applied, because these are not available for the Web Search property at country level. Each country was retrieved as a separate single-geography query rather than as a multi-region comparison, so that every series is normalized to its own within-country maximum and no cross-country rescaling is imposed by the platform. Data were exported using the platform's native CSV download function and were not otherwise transformed. Every query is therefore reproducible from the template https://trends.google.com/trends/explore?date=2021-07-01%202026-06-30&geo=XX&q=ibuprofen&cat=0&hl=en-GB, in which XX denotes the two-letter country code; the fourteen country-specific query strings are listed in Supplementary Table S1. The series were assembled as a balanced panel, excluding a partial trailing month so all countries shared identical start and end dates. Because Google Trends normalizes each country series independently, raw RSV levels, country means and forecast levels were interpreted only within country and were not ranked across countries.

Search term selection and linguistic considerations

A single harmonized generic query, the International Nonproprietary Name "ibuprofen", was used in every country rather than country-specific term sets, for three reasons. First, construct comparability: because Google Trends normalizes each query independently, a different term set in each country would confound genuine temporal differences with differences in query construction, whereas one harmonized term keeps the measured construct identical across all fourteen series. Orthographic stability: "ibuprofen" is the exact form used in ten of the fourteen countries (Germany, Austria, the United Kingdom, the Netherlands, Belgium, Poland, the Czech Republic, Norway, Sweden and Turkey), and the four remaining national forms (French ibuprofène, Italian ibuprofene, Spanish ibuprofeno and Hungarian ibuprofen) share the same stem, Google Trends term matching being case-insensitive and normalizing diacritics. Transparency: a search term corresponds to a defined and reproducible query set, whereas a Google Trends Topic is an algorithmically curated entity whose composition is undocumented and may change over time, which would preclude exact replication.

Two consequences of this choice are acknowledged explicitly. Term-based matching may capture inflected national forms incompletely, which could attenuate measured volume in the Romance-language countries and in Hungary. More importantly, lay searching for over-the-counter analgesics is frequently brand-led, and widely used national ibuprofen brands, among them Nurofen and Brufen in several markets, Moment (Italy), Advil (France, Belgium, the Netherlands), Dolormin and Ibuflam (Germany, Austria), Espidifen and Neobrufen (Spain), Ibuprom (Poland), Ibalgin (the Czech Republic), Ipren and Ibumetin (Sweden), Ibux (Norway) and Algoflex (Hungary), are not captured by the generic term. Brand penetration differs between markets, so the generic term captures a

OPEN ACCESS

different share of total ibuprofen-related searching in each country.

Two features of the design limit the consequences of this for the present inferences. Every analysis is conducted strictly within country, trends are estimated on each country's own normalized series, STL components are derived within country, remainders are correlated only after each country's own trend and seasonality have been removed, and holdout outcomes are standardized using each country's own training-period mean and standard deviation, so a constant proportional difference in generic-versus-brand search share alters the scale of a national series but not its estimated trend direction, seasonal timing or residual structure. In addition, generic and brand terms were deliberately not summed, because Google Trends renormalizes any multi-term comparison to the joint maximum; adding brand terms would change the meaning of the 0-100 index rather than produce an additive measure of volume. Estimates would be biased only if the generic-to-brand ratio itself drifted systematically within a country over the study period; this cannot be excluded and is stated as a limitation.

Country selection

The fourteen countries constitute a purposive rather than exhaustive. Countries were included when they were among the more populous markets of their European sub-region, so that query volume was sufficient for a single generic drug term to yield a continuous, non-zero monthly series across all 60 study months; had high national internet penetration, defined as internet use by at least 85% of individuals aged 16-74 in recent European statistics;[16] had Google as the dominant search engine, with a national market share above 90%, so that RSV reflects the majority of national search behavior rather than a minority platform; and together provided coverage of the main European sub-regions: Northern (Norway, Sweden), Western (France, Germany, the United Kingdom, the Netherlands, Belgium, Austria), Southern and South-Eastern (Italy, Spain, Turkey) and Central-Eastern (Poland, the Czech Republic, Hungary), with variation in health-system organization and in the regulatory and retail availability of over-the-counter ibuprofen.

Countries were not analyzed when any of the following applied: a small population or otherwise insufficient query volume, which for a single generic drug term yields months with RSV = 0 or truncated series unsuitable for STL decomposition and seasonal ARIMA modelling (for example Luxembourg, Malta, Cyprus, Iceland and the European microstates); a national search market in which Google is not dominant, or interruption of normal internet use during the study window (the Russian Federation, Ukraine and Belarus); or the absence of a complete balanced 60-month series. Several further eligible mid-sized countries, among them Portugal, Greece, Romania, Switzerland, Denmark, Finland, Ireland and Slovakia, were not included, in order to keep the panel at a size at which each country could be modelled individually, diagnostically checked and legibly displayed, and at which the multiplicity burden of simultaneous country-level tests remained tractable. This was a design decision rather than a data-driven exclusion; it limits generalizability to non-included European countries and extension to a full Europe-wide panel is identified as the natural next step. The criteria and their application are summarized in Supplementary Table S2.

Sampling variability

Google Trends reports a sample of the underlying query stream rather than a full census of queries, so values regenerated on a different day may differ slightly even when the query parameters are unchanged; published assessments indicate that this variability is smallest for high-volume queries over long observation windows and increases for low-volume queries and short windows.[17,19] Two features of the present design limit its influence. The analyzed series are monthly aggregates spanning five years for a high-volume consumer-medicine term, which is the regime in which sampling noise is smallest and every inference rests on structural features of the series, a secular slope estimated across 60 months, a seasonal profile averaged over five annual cycles, and correlations between remainder components, rather than on individual monthly values. Replicate extractions were not performed, so the magnitude of this variability was not quantified in the present dataset and exact numerical reproduction of the series at a later date cannot be guaranteed; this is stated as a limitation. To allow readers to judge the issue directly, the extraction date and the complete query specification are reported in Supplementary Table S1.

Statistical analysis

Temporal trends

For secular trend inference, each country was modeled using linear regression of RSV on continuous time with calendar month as a fixed effect. Newey-West heteroskedasticity- and autocorrelation-consistent standard errors with a 12-month lag were used for the primary slope, confidence interval and p value and Benjamini-Hochberg FDR correction was applied across the fourteen country tests.[20] Seasonal Mann-Kendall testing[21] and generalized least squares with an AR(1) residual structure served as sensitivity analyses; residual autocorrelation was also summarized. Both sensitivity analyses are reported country by country, alongside the primary estimate, in Supplementary Table S4. To visualize trajectory shape without implying comparability of absolute levels, descriptive trend displays were interpreted within country.

Seasonality

Seasonality was evaluated using robust periodic seasonal-trend decomposition by LOESS (STL),[22,23] from which seasonal strength, peak month, trough month and seasonal amplitude were derived. Seasonal strength was quantified as $F_S=\max(0, 1-\mathrm{Var}(R_t)/\mathrm{Var}(S_t+R_t))$, where S_t and R_t are the STL seasonal and remainder components, so that 0 indicates no detectable seasonality and values approaching 1 indicate a dominant seasonal signal;[24] seasonal amplitude was defined as the difference between the largest and smallest mean monthly seasonal component, and the peak and trough months as the calendar months at which those extremes occurred. Calendar-month effects were also tested jointly using HAC-adjusted inference with FDR correction, and recurrence was examined across complete seasonal cycles by comparing winter (December-February) with summer (May-July). Per-country values for all of these quantities are given in Supplementary Table S3.

Cross-country co-movement

Cross-country co-movement was assessed primarily using Pearson correlations between STL remainder components, thereby removing the common trend and seasonal structure; correlations between first differences of RSV were used as a sensitivity analysis and are reported in Supplementary Table S5.

Forecasting and out-of-sample validation

For forecasting, each series was transformed with a bounded logit transformation, preserving the monthly frequency, and seasonal ARIMA models were selected by AICc with KPSS regular-differencing and OCSB seasonal-differencing tests.[25] Twelve-month forecasts and 95% intervals were back-transformed to the 0-100 scale. Residual adequacy was assessed with Ljung-Box testing. Out-of-sample performance was evaluated by rolling-origin refitting from a minimum 36-month training history at horizons of 1, 3, 6 and 12 months and was compared with a seasonal-naive forecast.

Chronological predictive modelling

For predictive modelling, the final 12 months (July 2025-June 2026) formed a strict chronological holdout; the preceding 48 months formed the training period. The outcome was standardized within country using training-period means and standard deviations only. Random Forest[26] and Gradient Boosting[27] hyperparameters were selected using rolling temporal validation within the training data. Performance on the untouched holdout was compared with a country-specific harmonic regression and a seasonal-naive benchmark using R^2, RMSE, MAE and MBE.[28] Random Forest *vs* Gradient Boosting differences were evaluated with a paired country-block bootstrap (10,000 resamples) and holdout permutation importance was interpreted strictly as predictive importance, not causation or variance explained. The complete hyperparameter grids, the rolling temporal validation scheme and the bootstrap procedure are specified in the Supplementary Methods.

Software

Analyses were performed in R (version 4.5.2; R Foundation for Statistical Computing, Vienna, Austria). The packages used, with versions, are listed in the Supplementary Methods.

Results

A total of 840 monthly observations, 60 for each of the 14 countries, spanning July 2021 to June 2026, were analyzed. The raw series (Figure 1) showed substantial month-to-month variation and recurrent peaks. Because each country series is independently normalized, the figure is descriptive within country and does not permit comparison of absolute search intensity between countries.

Temporal trends

Using the prespecified primary model (calendar-month-adjusted regression with Newey-West HAC standard errors and FDR correction), eight countries showed significant increasing trends: Hungary (+6.20 RSV/year; 95% CI 3.18 to 9.22), Poland (+5.85; 4.11 to 7.59), Turkey (+5.58; 3.84 to 7.32), Sweden (+5.53; 3.68 to

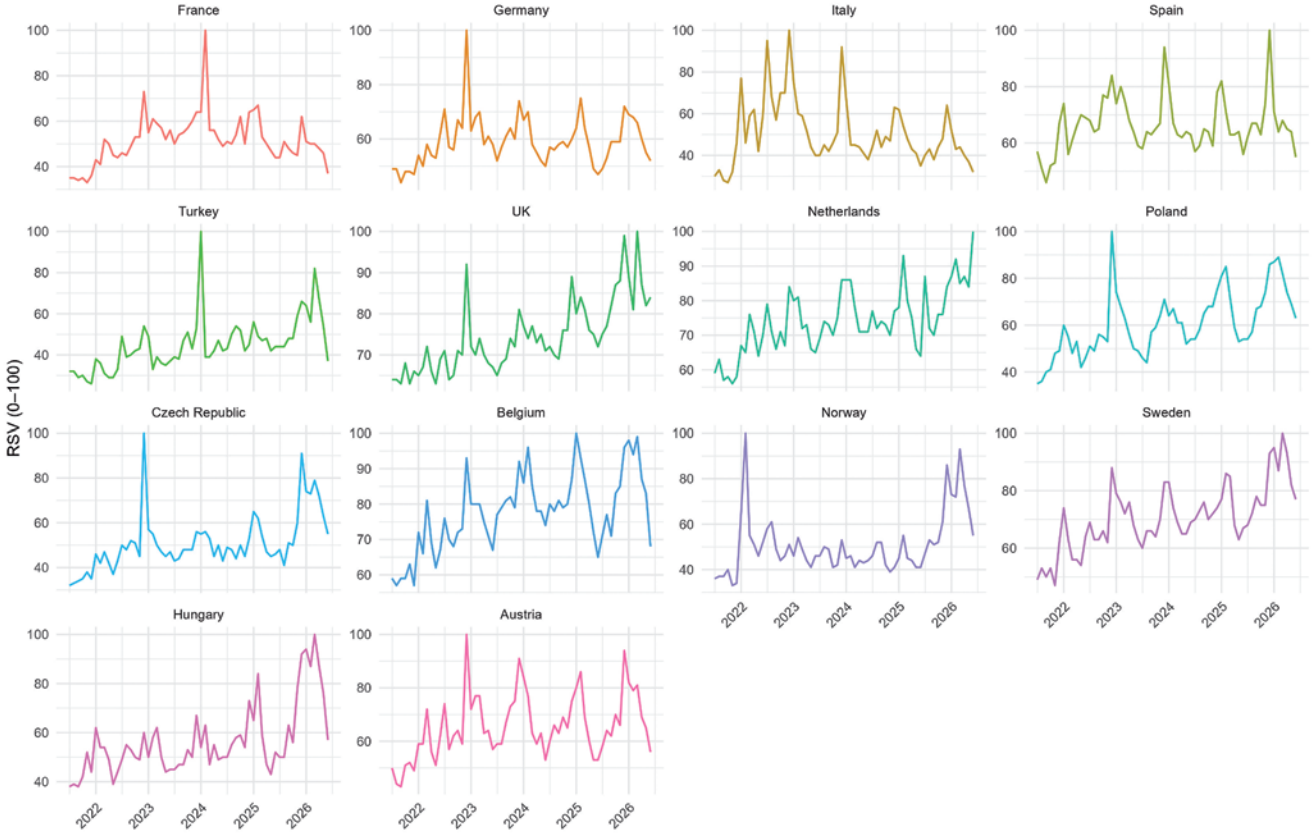


Figure 1. Monthly RSV for "ibuprofen" across fourteen European countries, July 2021 to June 2026. Each panel is normalized to that country's own peak (100); vertical levels must not be compared between countries.

OPEN ACCESS

7.38), Belgium (+4.66; 2.50 to 6.82), the United Kingdom (+4.57; 3.48 to 5.66), the Czech Republic (+4.56; 1.77 to 7.34), and the Netherlands (+4.13; 2.37 to 5.89) (Figure 2; Table 1). Austria had a positive point estimate (+3.07; 0.23 to 5.92; raw p=0.035) but did not remain significant after FDR correction. Norway (+2.97), France (+1.57), Spain (+1.02), Germany (+1.00), and Italy (-2.00) also had no FDR-significant trend. No country had an FDR-significant decrease. These slopes describe change within each independently normalized country series rather than differences in absolute national search intensity.

Table 1. Primary temporal-trend analysis. Slopes are from regression adjusted for calendar month with Newey-West HAC standard errors. Raw p-values are shown; the FDR column reflects Benjamini-Hochberg correction across 14 countries. Slope magnitudes are interpretable within country only.

Country	HAC slope/year (95% CI)	Raw p-value	BH-FDR	Primary trend
Hungary	+6.20 (3.18 to 9.22)	<0.001	Significant	Increasing
Poland	+5.85 (4.11 to 7.59)	<0.001	Significant	Increasing
Turkey	+5.58 (3.84 to 7.32)	<0.001	Significant	Increasing
Sweden	+5.53 (3.68 to 7.38)	<0.001	Significant	Increasing
Belgium	+4.66 (2.50 to 6.82)	<0.001	Significant	Increasing
United Kingdom	+4.57 (3.48 to 5.66)	<0.001	Significant	Increasing
Czech Republic	+4.56 (1.77 to 7.34)	0.002	Significant	Increasing
Netherlands	+4.13 (2.37 to 5.89)	<0.001	Significant	Increasing
Austria	+3.07 (0.23 to 5.92)	0.035	Not significant	No FDR-significant trend
Norway	+2.97 (-1.80 to 7.73)	0.217	Not significant	No FDR-significant trend
France	+1.57 (-2.76 to 5.89)	0.470	Not significant	No FDR-significant trend
Spain	+1.02 (-1.49 to 3.54)	0.417	Not significant	No FDR-significant trend
Germany	+1.00 (-1.94 to 3.94)	0.498	Not significant	No FDR-significant trend
Italy	-2.00 (-7.20 to 3.20)	0.443	Not significant	No FDR-significant trend

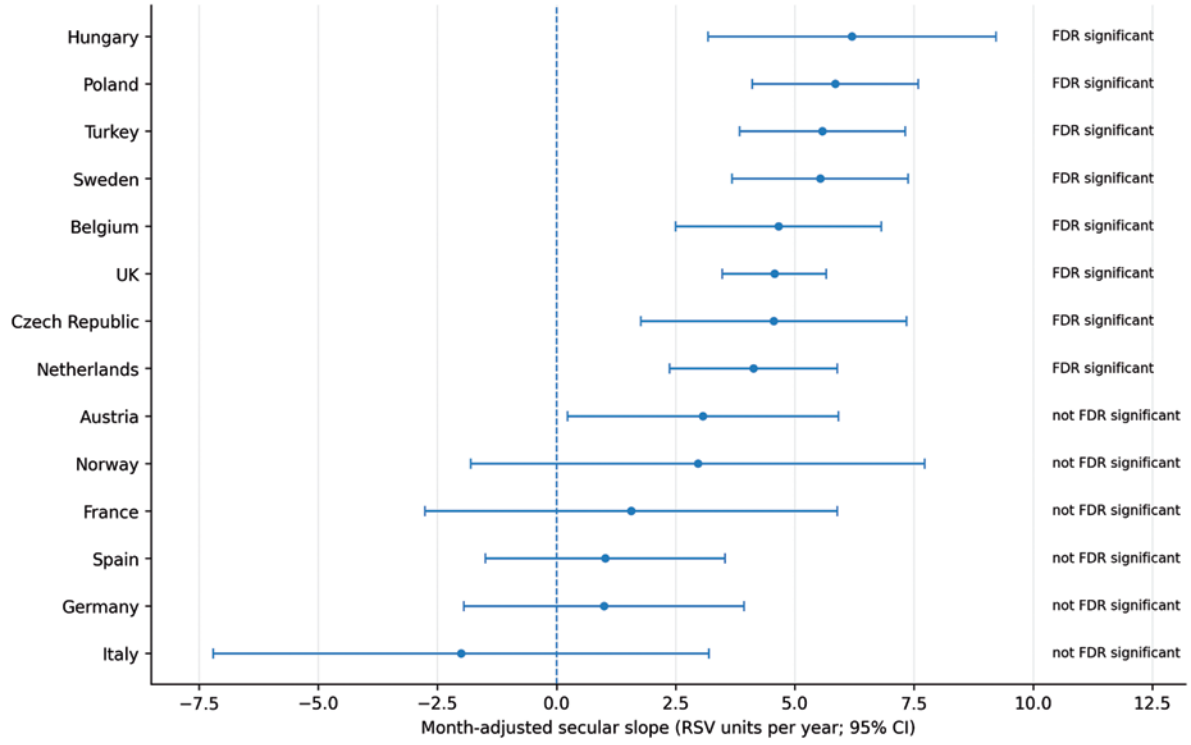


Figure 2. Within-country secular trend estimates. Points show the month-adjusted annual slope and horizontal bars the 95% confidence interval from the primary Newey-West HAC regression. FDR significance was determined across the 14 country tests using the Benjamini-Hochberg procedure. The zero line denotes no secular change.

Prespecified sensitivity analyses supported the primary trend inference (Supplementary Table S4). The seasonal Mann-Kendall test and generalized least squares with an AR(1) error structure agreed with the primary HAC model on the direction of change in all fourteen countries, and on FDR significance status in twelve of fourteen. The two exceptions were Austria, whose primary estimate lay immediately above the significance threshold (HAC q=0.054) and which reached FDR significance under both sensitivity models (seasonal Mann-Kendall q=0.002; GLS q=0.010), and Italy, for which the seasonal Mann-Kendall test indicated a monotonic decline (τ=-0.40; q=0.002) that was supported neither by the primary model (-2.00 RSV/year; q=0.498) nor by GLS (-1.10 RSV/year; q=0.629). Italy had the widest primary confidence interval in the panel, consistent with the high month-to-month volatility of its series. We therefore retain the primary finding that no country showed an FDR-significant decrease, while reporting this discordance explicitly. First-order residual autocorrelation of the primary models ranged from 0.24 (Turkey) to 0.63 (Norway), supporting the use of HAC-adjusted inference.

Seasonality

Robust periodic STL decomposition retained a recurrent winter-centered seasonal pattern in many country series, with seasonal components commonly higher in winter and lower in late spring or summer (Figure 3; Supplementary Figure S1). Peak timing, amplitude and seasonal strength varied by country, and Norway showed a comparatively weaker and less regular profile. The analysis also evaluated the joint effect of calendar month with HAC-adjusted inference and examined recurrence through winter-versus-summer contrasts across complete seasonal cycles. Given that only five annual cycles were available, seasonality is interpreted as recurrent rather than universal, and the figures should not be taken as evidence that winter respiratory illness is the causal driver.

Per-country seasonality metrics are reported in Supplementary Table S3. Seasonal peaks fell between December and February in all fourteen countries, in December in ten countries, in January in Poland and the Czech Republic, and in February in the Netherlands and Norway and troughs fell between April and July, most commonly in June (seven countries) or May (five). Seasonal strength ranged from 0.20 in Norway to 0.72 in Poland (median 0.47) and seasonal amplitude from 10.5 RSV points in Norway to 37.3 in Austria (median 18.3), confirming that Norway had both the weakest and the flattest seasonal profile. The joint calendar-month effect remained significant after FDR correction in all fourteen countries (all q<0.001). Across the four complete winter–summer cycles, the mean winter (December-February) minus summer (May-July) difference was positive in every country, ranging from 6.2 RSV points in the United Kingdom to 20.3 in Poland; winter exceeded summer in at least three of the four cycles in every country, and in all four cycles in Poland, Norway, Sweden and Hungary.

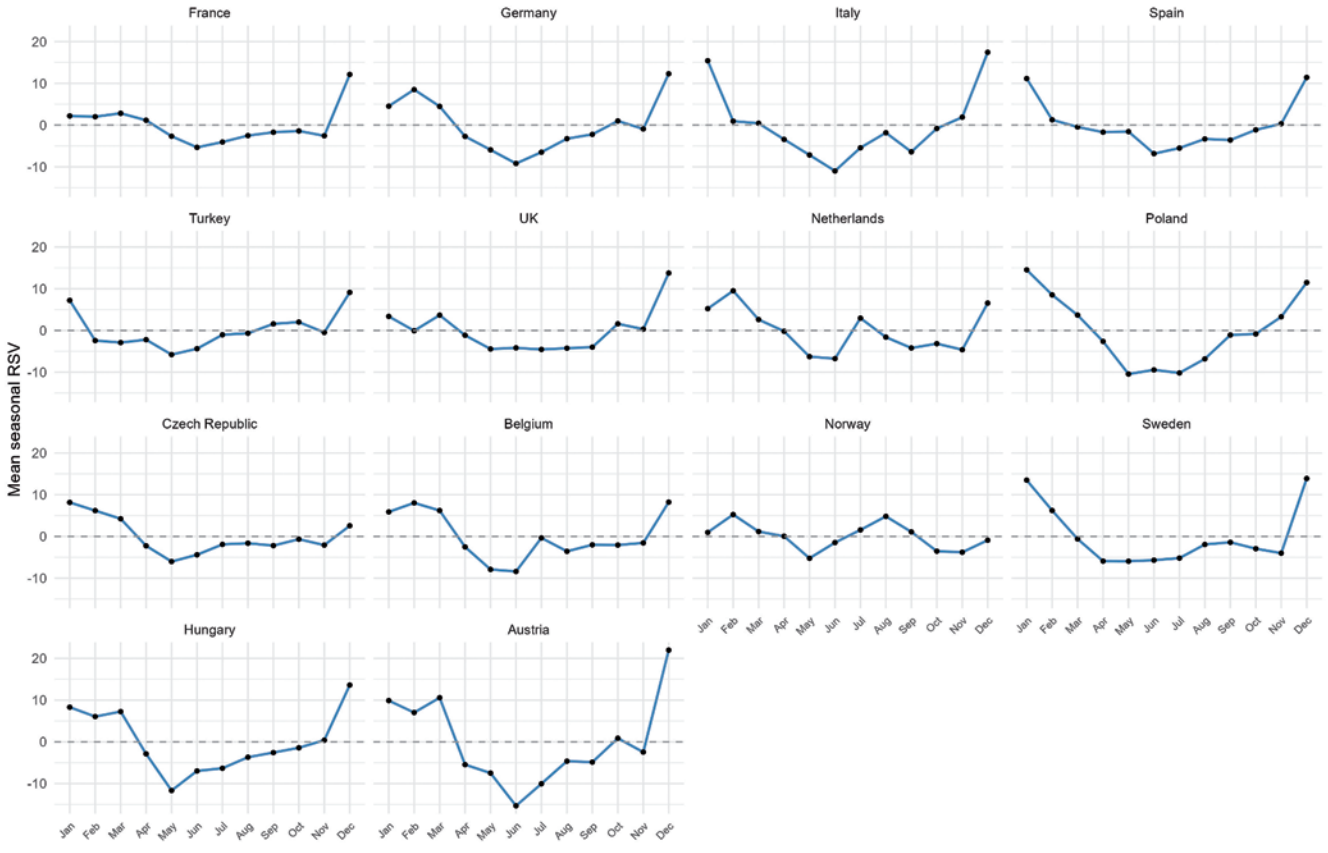


Figure 3. Seasonal component of ibuprofen RSV by calendar month for each country, derived from robust periodic STL decomposition of 60 monthly observations per country. Profiles illustrate within-country seasonal timing and amplitude; absolute RSV levels are not comparable across countries. RSV, relative search volume; STL, seasonal-trend decomposition by LOESS.

OPEN ACCESS

Adjusted cross-country co-movement

Cross-country co-movement was evaluated after removal of trend and seasonality rather than on the raw monthly series. Pearson correlations between robust STL remainder components were generally weaker and more heterogeneous than the raw-series correlations from the initial analysis (Figure 4). The largest positive residual correlations were observed for Germany-Austria (r=0.79), Poland-Czech Republic (r=0.76), Germany-Poland (r=0.66), Germany-United Kingdom (r=0.65), and Germany-Italy (r=0.61). In contrast, several pairs showed little residual synchrony, including France-Sweden (r=0.01), France-Norway (r=-0.05), and Germany-Norway (r=-0.04), while the Netherlands-Norway pair showed the most negative correlation (r=-0.36). These coefficients describe synchronous departures from each country's own trend and seasonal pattern; they do not indicate comparable absolute search levels, medication use, or common causation. Correlations between first differences were used as a sensitivity analysis and are reported in Supplementary Table S5.

Forecasting and out-of-sample validation

Seasonal ARIMA models produced 12-month within-country projections on the bounded 0-100 RSV scale (Figure 5).

Their predictive performance was mixed when tested by rolling-origin validation against a seasonal-naive benchmark (Table 2). At a 1-month horizon, ARIMA had a slightly lower RMSE than seasonal naive (10.56 *vs* 11.18), although its MAE was slightly higher (7.71 *vs* 7.60). At 3, 6 and 12 months, seasonal

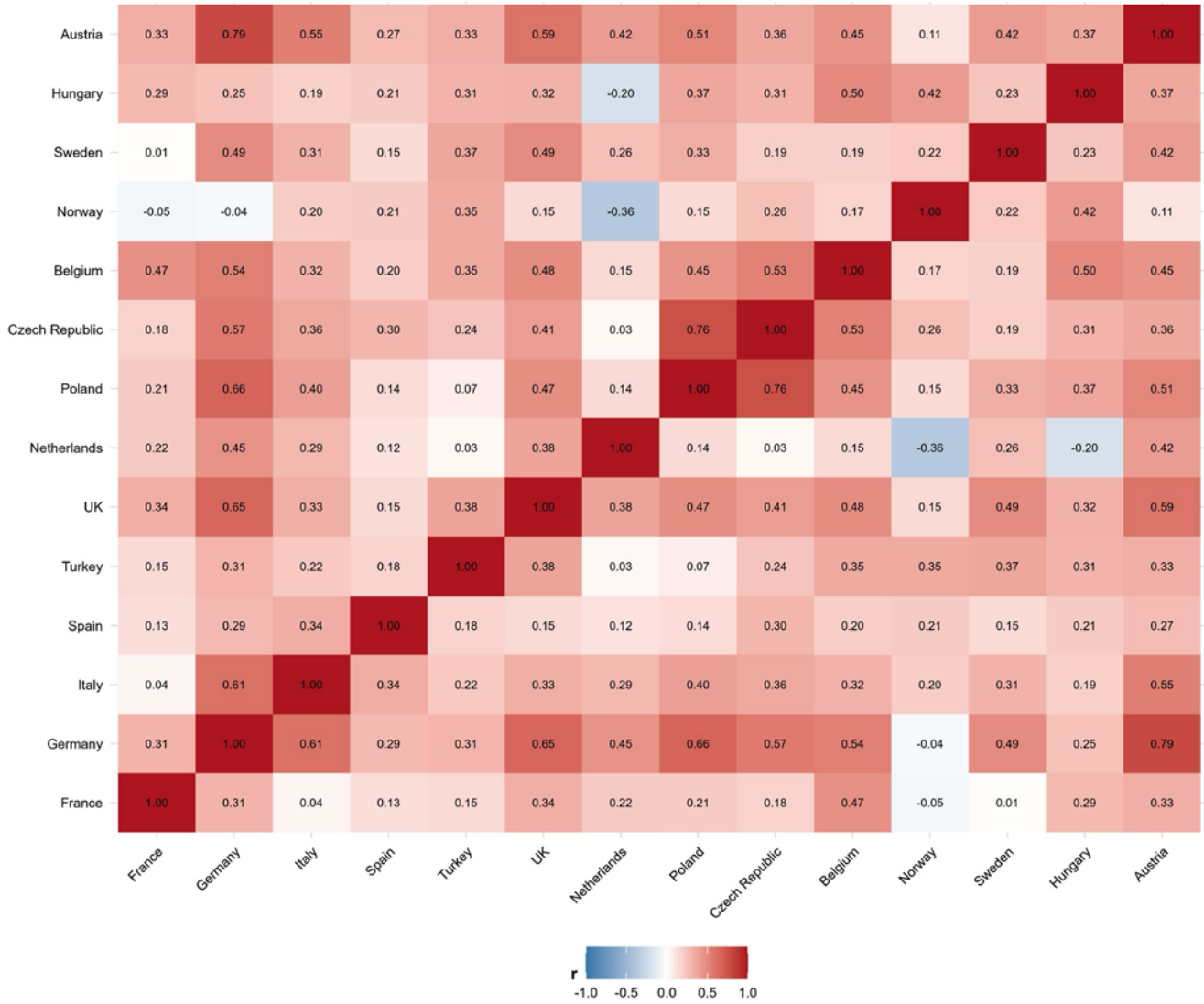


Figure 4. Adjusted cross-country co-movement in ibuprofen search attention. Cells show Pearson correlations between robust STL remainder components after removal of country-specific trend and seasonality (n=60 monthly observations per country). Positive values indicate synchronous residual deviations and negative values indicate opposing residual deviations. Because Google Trends normalizes each country series independently, the matrix must not be interpreted as a comparison of absolute search interest between countries.

naive had lower RMSE (11.53, 12.25 and 13.11) than ARIMA (12.86, 13.79 and 13.98), and lower MAE at each horizon. ARIMA bias became increasingly negative with horizon (MBE -0.13, -1.39, -2.81 and -3.84 at 1, 3, 6 and 12 months).

Residual Ljung-Box diagnostics were satisfactory for 13 of 14 final country models; Belgium retained residual autocorrelation (p=0.003) and its forecast should therefore be interpreted with additional caution (Table 3). These findings support presentation of the forecasts as exploratory model-based projections rather than evidence of uniformly strong predictability.

Table 2. Overall rolling-origin forecast validation for seasonal ARIMA and the seasonal-naive benchmark.

Model	Horizon (mo)	n	RMSE	MAE	MBE
ARIMA	1	336	10.56	7.71	-0.13
ARIMA	3	308	12.86	9.50	-1.39
ARIMA	6	266	13.79	10.48	-2.81
ARIMA	12	182	13.98	10.66	-3.84
Seasonal naive	1	336	11.18	7.60	-3.53
Seasonal naive	3	308	11.53	7.84	-3.49
Seasonal naive	6	266	12.25	8.45	-3.95
Seasonal naive	12	182	13.11	9.10	-5.78

RMSE, root mean square error; MAE, mean absolute error; MBE, mean bias error (predicted minus observed). Errors are expressed in within-country RSV units.

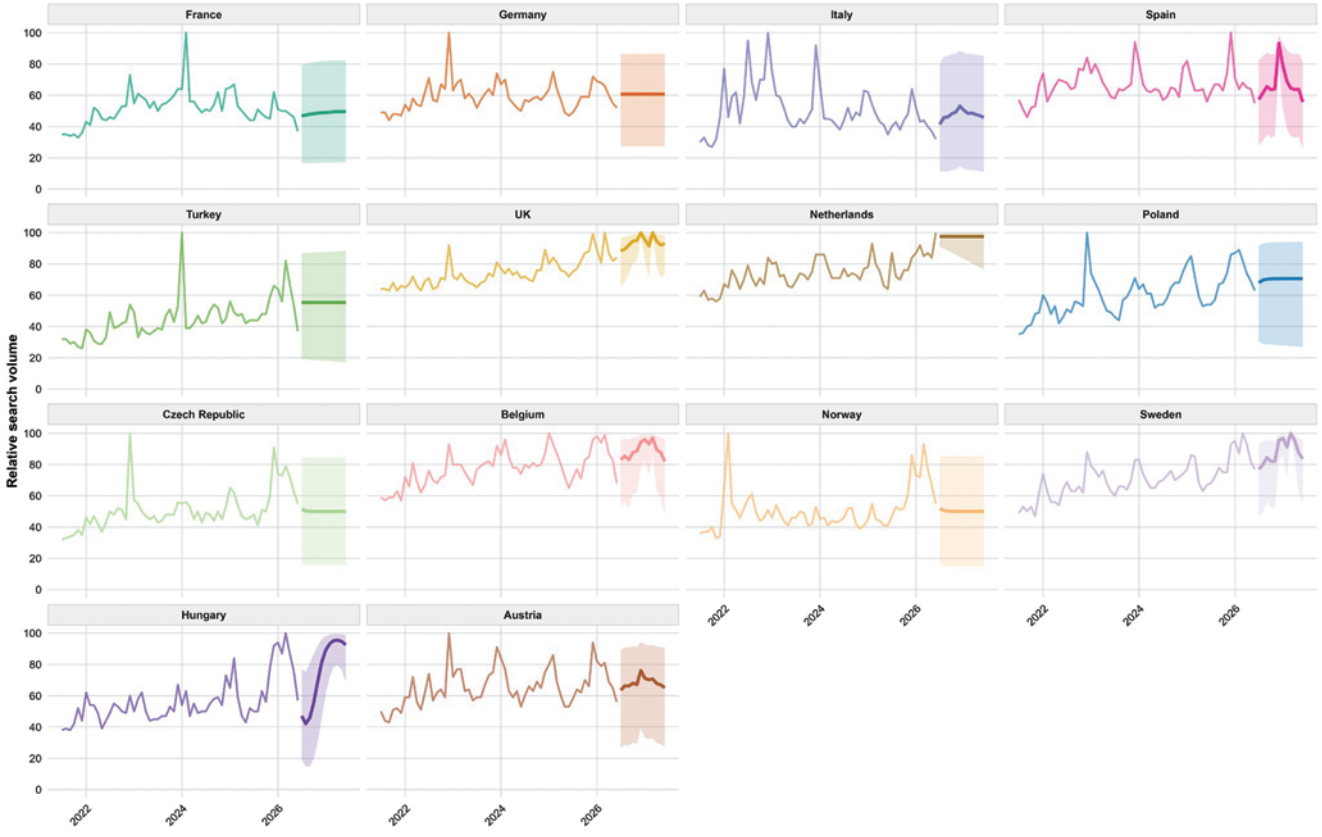


Figure 5. Validated 12-month seasonal ARIMA forecasts of ibuprofen RSV by country. Historical observations, forecast means and 95% prediction intervals are shown. Forecast levels are interpretable only within each country because Google Trends normalizes each country series independently; panels must not be ranked by forecast level. ARIMA, autoregressive integrated moving average. Selected model orders and residual diagnostics are given in Table 3.

OPEN ACCESS

Chronological predictive modelling

The predictive analysis used the final 12 months (July 2025-June 2026; n=168 country-month observations) as an untouched chronological holdout, with RSV standardized within country using training-period means and standard deviations only. Rolling temporal validation within the training period selected Random Forest with mtry=2 and nodesize=5 (mean CV RMSE 0.750, SD 0.223) and Gradient Boosting with interaction depth=3, shrinkage=0.05, 500 trees and minimum node size=10 (mean CV RMSE 0.796, SD 0.202). On the final holdout, harmonic regression provided the best overall performance (R^2=0.242, RMSE=1.221, MAE=0.865, MBE=-0.237), followed by the seasonal-naive benchmark (R^2=-0.005, RMSE=1.406, MAE=0.972, MBE=-0.700), Random Forest (R^2=-0.026, RMSE=1.420, MAE=1.002, MBE=-0.787), and Gradient Boosting (R^2=-0.030, RMSE=1.423, MAE=1.003, MBE=-0.839) (Table 4; Figure 6). Thus, the nonlinear machine-learning models did not demonstrate incremental future-prediction value over the simpler structured temporal benchmarks.

Random Forest and Gradient Boosting were statistically indistinguishable on the chronological holdout in the paired country-block bootstrap (10,000 resamples): R^2 difference 0.0046 (95% bootstrap interval -0.0686 to 0.0965; *p*=0.964), RMSE difference -0.0032 (-0.0577 to 0.0487; p=0.964), MAE difference -0.0019 (-0.0519 to 0.0469; *p*=0.946), and absolute-bias difference -0.0526

Table 3. Selected final seasonal ARIMA models and residual diagnostics. AICc was used for within-country model selection. The Ljung-Box test evaluated residual autocorrelation; Belgium was the only final model with p<0.05 (p=0.003). AICc values are not used to rank countries.

Country	Selected model	AICc	Ljung-Box lag	Ljung-Box p-value
France	ARIMA(1,0,1)(0,0,0)[12]	138.84	15	0.908
Germany	ARIMA(0,0,0)(0,0,0)[12]	132.62	15	0.899
Italy	ARIMA(1,0,0)(1,0,0)[12]	159.04	15	0.667
Spain	ARIMA(1,0,0)(1,1,0)[12]	102.58	15	0.956
Turkey	ARIMA(0,1,1)(0,0,0)[12]	150.39	15	0.990
UK	ARIMA(0,1,1)(0,1,0)[12]	101.14	15	0.899
Netherlands	ARIMA(0,1,1)(0,0,0)[12]	115.13	15	0.379
Poland	ARIMA(1,1,1)(0,0,0)[12]	146.68	15	0.607
Czech Republic	ARIMA(1,0,0)(0,0,0)[12]	150.43	15	0.999
Belgium	ARIMA(0,1,1)(1,0,0)[12]	142.57	15	0.003
Norway	ARIMA(1,0,0)(0,0,0)[12]	150.82	15	1.000
Sweden	ARIMA(1,0,0)(0,1,0)[12]	101.13	15	0.944
Hungary	ARIMA(3,1,2)(0,0,0)[12]	131.55	15	0.500
Austria	ARIMA(1,0,0)(1,0,0)[12]	147.50	15	0.805

Table 4. Predictive performance of machine-learning models and temporal benchmarks. The final 12 months were held out chronologically. Training and holdout outcomes were standardized within country using training data only.

Model	Set	n	R^2	RMSE	MAE	MBE
Random Forest	Training	672	0.882	0.340	0.228	0.000
Gradient Boosting	Training	672	0.821	0.418	0.284	-0.002
Harmonic regression	Training	672	0.410	0.760	0.530	-0.000
Random Forest	Chronological holdout	168	-0.026	1.420	1.002	-0.787
Gradient Boosting	Chronological holdout	168	-0.030	1.423	1.003	-0.839
Harmonic regression	Chronological holdout	168	0.242	1.221	0.865	-0.237
Seasonal naive	Chronological holdout	168	-0.005	1.406	0.972	-0.700

RMSE, root mean square error; MAE, mean absolute error; MBE, mean bias error (predicted minus observed).

Table 5. Paired country-block bootstrap comparison of Random Forest and Gradient Boosting on the chronological holdout (10,000 resamples). Differences are Random Forest minus Gradient Boosting.

Metric	Observed difference	95% bootstrap interval	Two-sided p-value
R2	0.0046	-0.0686 to 0.0965	0.9636
RMSE	-0.0032	-0.0577 to 0.0487	0.9636
MAE	-0.0019	-0.0519 to 0.0469	0.9464
AbsMBE	-0.0526	-0.1113 to 0.0072	0.0810

AbsMBE, absolute mean bias error.

Table 6. Random Forest permutation importance on the chronological holdout. Importance is the increase in holdout RMSE after feature permutation over 200 repetitions. Values are predictive and should not be interpreted causally.

Feature	Baseline RMSE	Mean RMSE increase	Median increase	95% interval	Repetitions
Calendar month	1.420	0.254	0.254	0.172 to 0.328	200
Country	1.420	0.084	0.085	0.042 to 0.124	200
Scaled time	1.420	0.000	0.000	0.000 to 0.000	200

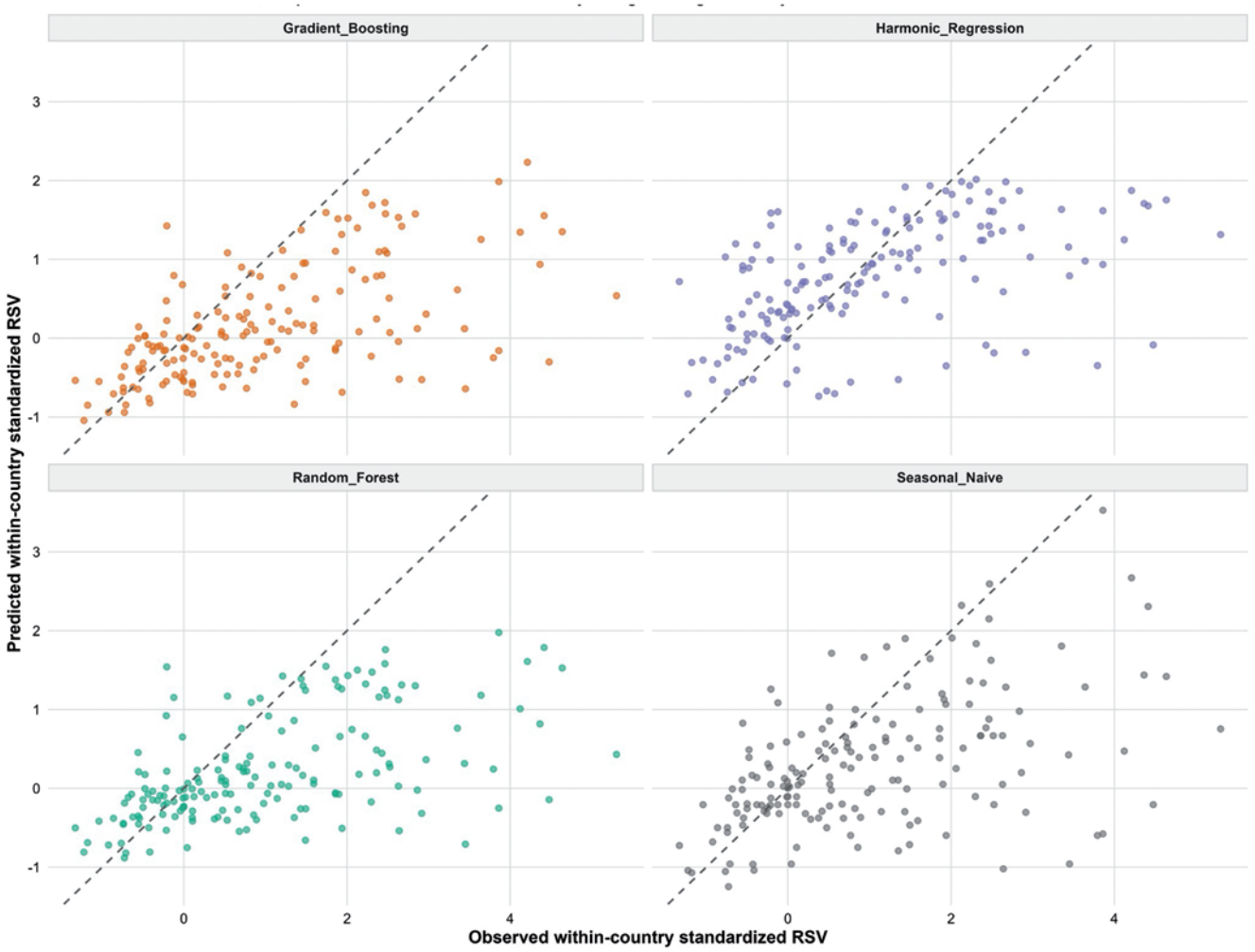


Figure 6. Chronological holdout: observed vs predicted within-country standardized RSV. The final 12 months were held out and outcome standardization used training data only. Panels show Gradient Boosting, harmonic regression, Random Forest and the seasonal-naive benchmark; the dashed diagonal is the line of identity. Closer concentration around the identity line indicates better prediction. n=168 country-month observations (14 countries × 12 months).

(-0.1113 to 0.0072; p=0.081) (Table 5). Random Forest holdout permutation importance identified calendar month as the most influential predictive feature (mean RMSE increase 0.254; 95% interval 0.172 to 0.328), followed by country (0.084; 0.04 to 0.124); permutation of the scaled secular-time variable produced no RMSE increase (0.000) (Tables 5 and 6; Figure 7). These values quantify predictive sensitivity within the fitted model and are not causal effects or proportions of variance explained.

Discussion

In this five-year analysis, after adjustment for calendar month, serial dependence, and multiple testing, online search attention for ibuprofen increased significantly in eight of fourteen countries and declined significantly in none. However, cross-country co-movement was evaluated only after removing trend and seasonality, ARIMA forecasts showed mixed out-of-sample performance rela-

OPEN ACCESS

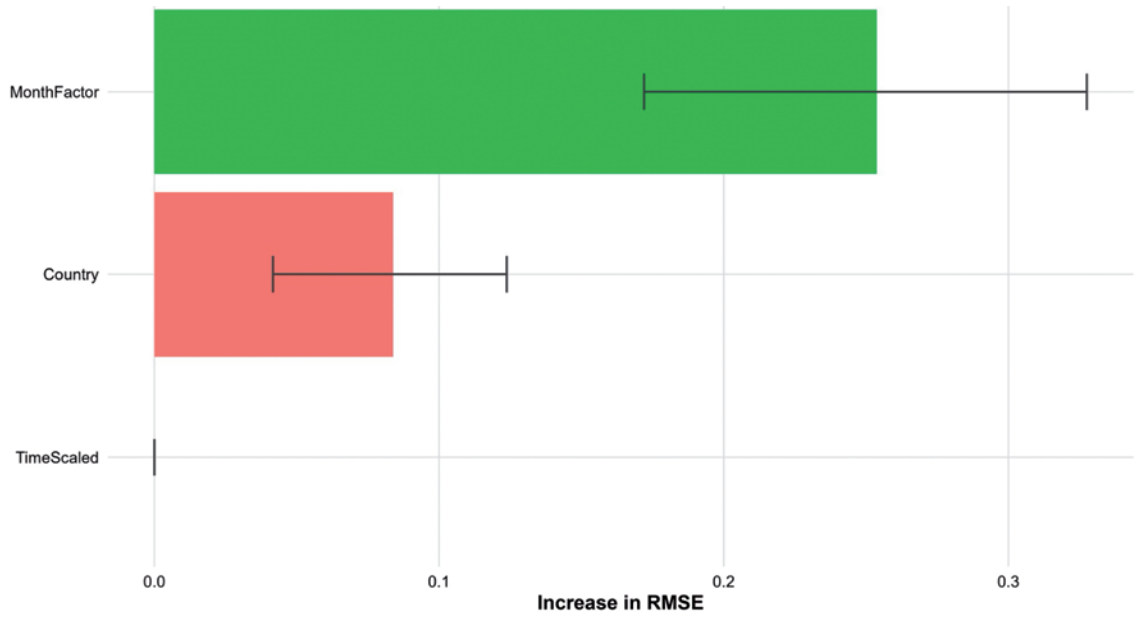


Figure 7. Random Forest predictive importance on the chronological holdout. Bars show the increase in holdout RMSE after permutation of each feature; error bars show the empirical 95% interval across 200 permutations. Importance is predictive, not causal, and does not represent variance explained. RMSE, root mean square error.

tive to a seasonal-naive benchmark, and harmonic regression outperformed the more complex machine-learning models on the strict chronological holdout. Together, these findings describe structured online information-seeking, not direct evidence of drug consumption, national demand, or validated pharmacovigilance performance.

The upward trends identified in Hungary, Poland, Turkey, Sweden, Belgium, the United Kingdom, the Czech Republic and the Netherlands are compatible with increasing online health-information seeking over the study period,[15,16] but the mechanism cannot be established from Google Trends alone. Importantly, Austria illustrates why the inferential framework matters: its positive slope was nominally significant before multiplicity adjustment but did not remain significant after FDR correction. Because every national series is normalized to its own peak, the slopes quantify change within each country's normalized series; they should not be used to infer that one country had a larger absolute increase in search activity than another.

The sensitivity analyses reported in Supplementary Table S4 refine this picture in two ways. Austria, which fell just above the FDR threshold in the primary model, was significant under both the seasonal Mann-Kendall and the GLS AR(1) specifications, so its classification is better regarded as borderline than as negative. Conversely, Italy showed a significant monotonic decline under the seasonal Mann-Kendall test alone; because that test responds to monotonic ordering rather than to the magnitude of annual change, and because Italy had the most volatile series in the panel, we treat this as a signal warranting replication rather than as evidence of a national decline in search attention.

The winter-centered pattern is clinically plausible because ibuprofen is used for both pain and fever and respiratory and febrile illnesses are more common during winter. Nevertheless, temporal coincidence is not causal validation. The analysis therefore treats respiratory-virus circulation, temperature, media attention and seasonal purchasing behavior as possible explanations rather than demonstrated drivers. Confirmation would require linkage with independent influenza-like-illness or respiratory-virus surveillance, temperature, dispensing or retail-sales data. The limited number of complete annual cycles also argues for describing the observed pattern as recurrent rather than near-universal.

The adjusted co-movement analysis also changes the interpretation of geographic structure. High correlations between raw country series can be generated by a common upward trend or a shared seasonal cycle even when short-term deviations are unrelated. After these components were removed, residual correlations were substantially more heterogeneous: Germany-Austria (r=0.79) and Poland-Czech Republic (r=0.76) remained relatively strong, but many pairs were weak or near zero and the Netherlands-Norway correlation was negative (r=-0.36). This narrower analysis asks whether countries experience synchronous departures from their own expected temporal pattern; it does not establish common causal drivers, similar consumption, or comparable national search intensity.

The forecasting and predictive-modelling results further argue for restraint. Seasonal ARIMA was competitive with the seasonal-naive benchmark only at the shortest horizon and did not consistently improve longer-horizon errors; one final model (Belgium) also retained significant residual autocorrelation. On the strict chronological holdout, harmonic regression achieved the best overall RMSE (1.221) and positive R^2 (0.242), whereas Random Forest and Gradient Boosting both had negative holdout R^2 values and RMSE values around 1.42. The paired country-block bootstrap found no meaningful difference between the two machine-learning algorithms. Moreover, Random Forest permutation importance was greatest for calendar month rather than country, while scaled secular time added no measurable holdout RMSE penalty when permuted. These results indicate that the most transferable predictive structure in this dataset was primarily regular temporal/seasonal structure, and that greater algorithmic complexity did not confer clear out-of-sample advantage. Google Trends remains attractive as a low-cost, fre-

quently updated source of population-level search attention, and the broader infodemiological literature supports its use for descriptive and hypothesis-generating surveillance.[7,8,13] However, the history of search-based surveillance[12] and the present validation results support a complementary role only: any application to medication safety or public-health monitoring should be externally validated against independent behavioral, epidemiological, dispensing, sales or adverse-event data before operational use.

Three reporting decisions in this analysis bear directly on how the findings should be read. First, the complete extraction protocol is specified, so the analysis can be repeated exactly; this matters because Google Trends output is sample-based and query-parameter dependent, and small differences in category, search property or geography specification can materially change the retrieved series.[17-19] The country panel is a purposive sample of large, high-penetration, Google-dominant markets, so the results describe those settings rather than Europe as a whole, and the eight countries with rising attention should not be read as the only European countries in which attention is rising. The harmonized generic term measures generic-name searching only; because brand penetration differs markedly between European markets, the proportion of ibuprofen-related searching that is captured is not constant across countries, which reinforces the decision to confine every comparison to within-country structure.

Limitations

Several limitations define the boundaries of interpretation. First, Google Trends provides a sampled, relative index rather than absolute query counts, and each country series is independently normalized to its own peak. Absolute RSV levels, means, forecast levels and model coefficients identifying countries therefore cannot be ranked across countries. A future peak may also rescale the historical series when the query is rerun, so forecasts predict the current normalized index rather than a permanently fixed future scale. Google Trends sampling can produce run-to-run variability. The extraction protocol is now fully specified (Supplementary Table S1) so that the query can be repeated exactly, but replicate downloads were not performed, the magnitude of this variability was therefore not quantified, and a later re-extraction may not reproduce the values exactly. The analysis used only the generic search term “ibuprofen”; national brand names and Google Trends Topics are not captured, and because brand share differs between markets and could drift over time within a market, a systematic change in the generic-to-brand ratio would bias the estimated trends. Multi-term or Topic-based extractions address a different question and would require separate downloads. The fourteen countries were selected purposively using the criteria set out in the Methods and Supplementary Table S2 rather than through an exhaustive Europe-wide eligibility framework, so the findings generalize to large, high-penetration, Google-dominant European markets rather than to Europe as a whole. Only five annual cycles were available, limiting inference about long-term stability of seasonality and making structural changes during and after the COVID-19 period difficult to separate. Residual correlations can still reflect shared unmeasured shocks, platform changes, media events, shortages or safety communications[23] and should not be interpreted causally. Forecast accuracy was horizon-dependent and not consistently better than a seasonal-naive benchmark; Belgium also showed residual autocorrelation in the final ARIMA fit. The predictive models used only country, time and calendar-month information and therefore test interpolation/extrapolation of deterministic temporal structure rather than clinical, behavioral or market-level prediction. Google Trends also gives no information on who is searching: queries may originate from patients, carers, students, health professionals or automated traffic, and the demographic composition of the searching population is unknown and may differ between countries and over time.[19] Finally, the study has no external validation against dispensing, retail sales, respiratory-disease activity, prescribing, adverse-event reporting or clinical outcomes. Consequently, the observed signal represents online information-seeking and cannot be equated with ibuprofen use, demand, morbidity, inappropriate self-medication or pharmacovigilance performance.

Conclusions

Between July 2021 and June 2026, Google search attention for ibuprofen showed FDR-significant increases in eight of fourteen selected European countries and recurrent winter-centered seasonality in many series. The analyses that account for seasonality, autocorrelation and multiplicity provide stronger inference for within-country trends, while correlations of STL remainders show that residual cross-country synchrony is more heterogeneous than suggested by raw-series correlations. Forecasting performance was mixed: seasonal ARIMA did not consistently outperform a seasonal-naive benchmark beyond the shortest horizon, and in the strict chronological predictive analysis harmonic regression outperformed Random Forest and Gradient Boosting. Random Forest permutation importance further indicated that calendar month carried more hold-out predictive information than country or secular time. These findings support Google Trends as a descriptive source of population-level online information-seeking, but not as a direct measure of medication consumption or a validated pharmacovigilance instrument. External validation against independent clinical, epidemiological, dispensing, sales and safety data is required before surveillance applications are inferred.

References

1. Ainyanbhor IE, Edo GI, Ali ABM, et al. Ibuprofen: A review on its synthesis, mechanism of action, pharmacological properties, and environmental impact. Pharmacol Res - Rep 2025;4: 100066.
2. Conaghan PG. A turbulent decade for NSAIDs: update on current concepts of classification, epidemiology, comparative efficacy, and toxicity. Rheumatol Int 2012;32:1491-502.
3. Mehuys E, Crombez G, Paemeleire K, et al. Self-medication with over-the-counter analgesics: a survey of patient characteristics and concerns about pain medication. J Pain 2019;20:215-23.
4. Kim A, Kim SC, Kim J, et al. Risk of cardiovascular diseases and gastrointestinal bleeding is comparable between celecoxib and non-selective non-steroidal anti-inflammatory drugs in patients with ankylosing spondylitis: a nationwide retrospective cohort study. Scand J Rheumatol 2025;54:204-12.
5. Grosser T, Ricciotti E, FitzGerald GA. The cardiovascular pharmacology of nonsteroidal anti-inflammatory drugs. Trends Pharmacol Sci 2017;38:733-48.
6. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11:e11.
7. Nuti SV, Wayda B, Ranasinghe I, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9: e109583.

OPEN ACCESS

8. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2009;457: 1012-4.
9. Lippi G, Mattiuzzi C, Cervellin G, Favaloro EJ. Direct oral anticoagulants: analysis of worldwide use and popularity using Google Trends. Ann Transl Med 2017;5:322.
10. Domnich A, Panatto D, Signori A, et al. Age-related differences in the accuracy of web query-based predictions of influenza-like illness. PLoS One 2015;10:e0127754.
11. Di Gennaro G, Licata F, Greco F, et al. Interest in mammography across European countries: a retrospective "Google Trends" comparative study. Quant Imaging Med Surg 2023;13:7523-9.
12. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014; 343:1203-5.
13. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019;5:e13439.
14. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017;7:185-9.
15. Di Novi C KM, Orso CE. Online health information seeking behavior, healthcare access, and health status during exceptional times. J Econ Behav Organ 2024:220:675-90.
16. Eurostat [Internet]. Digitalisation in Europe – 2024 edition. Avalable from: https://ec.europa.eu/eurostat/web/interactive-publications/digitalisation-2024
17. Rovetta A. Reliability of Google Trends: analysis of the limits and potential of web infoveillance during COVID-19 pandemic and for future research. Front Res Metr Anal 2021;6:670226.
18. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J Med Internet Res 2018;20:e270.
19. Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. Health Policy 2019;123:338-41.
20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57:289-300.
21. Hirsch RM, Slack JR. A nonparametric trend test for seasonal data with serial dependence. Water Resour Res 2010;20:727-32.
22. Hamedin MAA, Musa KI, Sulong MR. Time-series decomposition and modeling of dengue cases in Malaysia, 2022-2024: a nationwide observational study. Osong Public Health Res Perspect 2026;17:50-60.
23. Sanchez-Vazquez MJ, Nielen M, Gunn GJ, Lewis FI. Using seasonal-trend decomposition based on loess (STL) to explore temporal patterns of pneumonic lesions in finishing pigs slaughtered in England, 2005-2011. Prev Vet Med 2012;104:65-73.
24. Wang X, Smith K, Hyndman R. Characteristic-based clustering for time series data. Data Min Knowl Discov 2006;13:335-64.
25. Alabdulrazzaq H, Alenezi MN, Rawajfih Y, et al. On the accuracy of ARIMA based prediction of COVID-19 spread. Results Phys 2021;27:104509.
26. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. Brief Bioinform 2023;24:bbad002.
27. Boldini D, Grisoni F, Kuhn D, et al. Practical guidelines for the use of gradient boosting for molecular property prediction. J Cheminform 2023;15:73.
28. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7:e623.

Online supplementary material:
Supplementary Methods.
Table S1. Google Trends extraction protocol.
Table S2. Application of the country-selection criteria.
Table S3. Seasonality metrics by country.
Table S4. Sensitivity analyses for the within-country secular trend.
Table S5. Cross-country co-movement.
Figure S1. STL decomposition of ibuprofen.

Received: 3 August 2026; Accepted: 28 August 2028.

Contributions: Matteo Luigi Giuseppe Leoni, conceptualization, methodology, writing – original draft. Marco Mercieri, writing – review & editing. Y Van Tran, literature review, evidence synthesis, writing – review & editing. Phong Van Pham, literature review, evidence synthesis, writing – review & editing. Giacomo Farì, writing – review & editing. Piercarlo Sarzi Puttini, writing – review & editing. Giustino Varrassi, conceptualization, supervision, writing – original draft.

All authors have read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of interest: the authors declare no competing interests and all authors confirm accuracy.

Funding: this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability statement: the source data were derived from the publicly accessible Google Trends platform. Because Google Trends uses sampling and dynamic normalization, values regenerated at a later date may differ slightly from those in the analyzed dataset. The complete query specification required to regenerate the raw series is provided in Supplementary Table S1. The processed country-month dataset and the full analysis code are available from the corresponding author on reasonable request.

Acknowledgements: the authors express their gratitude to VIBRA Group of the Fondazione Paolo Procacci for the scientific support received during the publication process.

Ethics approval: this study used only aggregated, anonymized and publicly available data and did not involve human participants or identifiable information; ethical approval was therefore not required.

Publisher's note: all claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).