RESEARCH

BMC Public Health



Exploring different modelling approaches to forecast the community acute respiratory infections burden in children: an Italian epidemiological time series study



Riccardo Boracchini^{1,2*}, Benedetta Canova¹, Pietro Ferrara^{3,4}, Elisa Barbieri⁵, Pietro Giorgio Lovaglio⁶, Antonio Scamarcia⁷, Giovanni Corrao¹, Daniele Donà⁵, Carlo Giaquinto⁵, Costanza Di Chiara^{5†} and Anna Cantarutti^{1†}

Abstract

Background Acute respiratory infections (ARIs) in young children pose a significant global health challenge, leading to high rates of illness and death. They are estimated to be the fourth leading cause of mortality worldwide, particularly impacting children under five. This study aimed to identify the most effective time series model(s) for forecasting the epidemiological season burden of ARIs for the current 2023/2024 period in Italy.

Methods Data on the burden of ARIs' in children aged 0–14 years were retrieved from Pedianet, an Italian paediatric primary care database which includes over 200 family paediatricians. We analysed monthly incidence rates of ARIs from September 2010 to September 2023, following the typical seasonal pattern of these infections. Several forecasting models were compared to predict the future burden of ARI: Error, Trend, Seasonality (ETS); Seasonal Auto-Regressive Integrated Moving Average (SARIMA); Unobserved Component Model (UCM); and Trigonometric, Box Cox, ARMA errors, Trend, Seasonal (TBATS). We evaluated each model's accuracy by examining the residuals and the Mean Absolute Percentage Error (MAPE). The period between March 2020 and February 2022 was forecasted to represent the normal trend without COVID-19. Model parameters were estimated using the in-sample and out-of-sample approach.

Results The analysis included data from over 1.4 million cases of ARIs retrieved in children aged 0–14 years. The ETS model was implemented to predict the pandemic period. Overall, our findings suggest that exponential smoothing models as ETS (MAPE = 6.85) and TBATS (MAPE = 6.87) were most effective in predicting future trends in monthly ARIs' burden compared to other methods (i.e., UCM MAPE = 11.08, and SARIMA MAPE = 25.33).

Conclusions These findings suggest that exponential smoothing models are preferable for forecasting pediatric ARIs' burden trends in Italy. However, epidemiological data from the ongoing season are crucial for understanding whether residual pandemic effects continue affecting respiratory infection patterns.

Keywords Acute Respiratory Infections, Burden, Time series models comparison

[†]Anna Cantarutti and Costanza Di Chiara contributed equally as co-last authors.

*Correspondence: Riccardo Boracchini riccardo.boracchini@unimib.it Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Introduction

Acute Respiratory Infections (ARIs) represent one of the leading causes of morbidity and mortality in the paediatric population, especially in children aged 0 to 5 years, who are more susceptible due to factors such as immature immune systems, close contact with daycare or school settings, and other environmental exposures [1, 2]. The World Health Organization (WHO) has identified ARIs, particularly lower respiratory infections, as a significant global health burden, ranking them as the fourth leading cause of death and disability-adjusted life years (DALYs) [3]. These infections significantly strain healthcare systems, constituting a major reason for paediatric outpatient and inpatient care [4].

ARIs are predictable and seasonal despite generating outbreaks that significantly impact global health and healthcare systems. Understanding the impact of ARIs in real-world scenarios is crucial for effective public health interventions, including vaccination programs and healthcare systems preparedness, to mitigate this public health challenge [5]. Moreover, outlining the epidemiology of ARIs over time, including peaks, trends, and seasonality, can provide valuable insights for health policymakers in implementing targeted surveillance systems and preventing future ARIs outbreaks [6].

Time series models are the widest-used methods to analyze temporal variations in various phenomena. They are particularly well-suited for studying the dynamic nature of infectious disease outbreaks and their fluctuations over time. Some models focus on explaining the natural history of infections in the population, delineating states such as Susceptible, Infected and Recovered (SIR) models, and providing a foundational understanding of transmission and immune dynamics [7]. Others are more oriented to using past values and errors to predict future values, accounting for trends and seasonality (i.e., autoregressive integrated moving average models (ARIMA)) aiding in understanding overall patterns and cyclical variations (Exponential Smoothing and Unobserved Component Models) [8–11]. Lastly, new techniques combine the best-fitting parts of these models to build a new category defined as hybrid models [12].

In the past three years since the COVID-19 pandemic, an increasing number of studies have reported time series approaches to forecasting new cases, hospital admissions and deaths [13, 14]. However, many studies focus on a single type of time series approach, resulting in uncertain accuracy measures regarding forecasting [15]. Assuming that each historical series has inherent characteristics that distinguish them from others; conversely, each family model is able to capture only specific patterns. Testing different methodologies on the same time series is critical to providing the best fit and forecast. This study conducted a comparative analysis of various time series forecasting methods to determine the most accurate predictor of monthly fluctuations in pediatric acute respiratory illnesses at the community level. The objective was to mitigate the burden on pediatric public health resources during the 2023–2024 season using retrospective data up to 2010 and establish the optimal method for predicting future outbreaks of pediatric acute respiratory illnesses.

Materials and methods

Data source

Pedianet (http://www.pedianet.it) is a self-sufficient network of over 200 family paediatricians (FPs) who integrate the Junior Bit software into their clinical practices, forming an established paediatric primary care database. The network encompasses approximately 3% of the Italian paediatric population and was previously comprehensively detailed [16, 17]. Pedianet captures patientlevel information, including demographics, health status, clinical symptoms, drug prescriptions, and outpatient diagnoses, which is anonymized and securely stored in a protected cloud environment, identified only by a unique numerical identifier.

Study design and study population

This epidemiological time series study conducted a comparative analysis of various models to determine the optimal model for explaining the trend and seasonality within the monthly incidence rate of pediatric ARIs. The study cohort consists of all ARIs recorded by the FPs during a 13-year study period spanning September 2010 and September 2023. Each year, the observation period was divided into epidemiological seasons from September 1st to August 31st.

Cohort selection

The study included all children aged 0–14 enrolled in Pedianet between September 2010 and September 2023. Under the Italian National Healthcare Service, FPs are healthcare professionals responsible for conducting routine check-up examinations for preventive and medical purposes at specific stages of the child's life (i.e. 1st month, 2nd-3rd months, 6th month, 8-9th months, 12th month, 18th month, 24th month, 36th month, 5-6th years, 8-10th years, 10-12th years, and 12-14th years of age). To minimize misclassification, children without scheduled well-child visits with their FPs, consistent with previous literature, were excluded [18].

Definition of exposure

ARIs were identified using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes and a freely-entered text field designed and validated by an expert clinical data manager and divided into Lower Respiratory Tract Infections (LRTIs) and Upper Respiratory Tract Infections (URTIs) (Supplemental Materials, Table S1). All ARI cases from September 2010 to September 2023 were collected. Thereafter, cases were summarized into calendar monthly incidence rates per 1000 person-day.

Time series models

Several time series models were employed to identify the optimal model for explaining pediatric monthly ARI incidence rate patterns and forecasting future trends. The Seasonal Autoregressive Integrated Moving Average (SARIMA), Error Trend and Seasonality (ETS), Trigonometric seasonality, Box-Cox transformation, ARMA errors and Trend and Seasonal components models (TBATS), and Unobserved Component Models (UCM) models were selected based on their ability to capture different aspects of time series data and forecast accuracy. Each model's features are described below.

(i) SARIMA is the classic time series model used in several time series studies to capture both non-seasonal and seasonal patterns. It can be summarized as [8]:

$$ARIMA (p, d, q)(P, D, Q)_m$$
(1)

where p is the autoregressive order (AR) necessary to make forecasting based on past values, d is the degree of differencing involved, and q is the order of the moving average part (MA) to forecast starting from the previous forecast error; while P, D and Q indicate the corresponding seasonal parameters for a m seasonal period [8]. The selection of the order of the ARIMA model was based on the Akaike information criterion (AIC) and Maximum Likelihood Estimation (MLE) from *auto.arima* () function in the package *forecast* [19] using R software. For ARIMA models, a common technique to stabilize the variance in the data (achieve homoscedasticity) is a logarithmic transformation. We applied this transformation to our data before analyzing it with the ARIMA model [20].

(ii) ETS, an innovative state space model for exponential smoothing, is used without a clear trend or seasonal pattern. The core idea behind ETS models is to decompose the time series data into these three components (i.e., error, trend, and seasonality) using exponential smoothing techniques. This allows for more accurate forecasts compared to more straightforward methods like naïve forecasting. Exponential smoothing acts as the foundation for this intricate combination. It allows the model to assign different weights to past observations based on their relevance to the current prediction, giving more weight to recent data points while gradually fading out the influence of older ones. It can be expressed as [20]:

The error component may exhibit an additive (A) or multiplicative nature (M); the trend component can manifest as non-existent (N), additive, or additive with damping characteristics (A_d); and the seasonal component may be non-existent, additive, or multiplicative. The nature-component-model selection is made on AIC [9, 20]. The analysis was performed in R using the function *ets()* in the package *forecast* [19].

(iii) TBATS are designed to handle intricate seasonal patterns in time series data using a combination of the specified techniques. This approach is particularly effective in capturing dynamic seasonal patterns that may evolve over time. As above, exponential smoothing allows the model to assign different weights to past observations based on their relevance to the current prediction, giving more weight to recent data points while gradually fading out the influence of older ones. Overall, TBATS is a powerful tool for forecasting time series data with complex seasonal behaviour. It's particularly useful in scenarios where traditional methods may struggle due to non-linearity, diverse seasonality, or evolving trends. The general structure is as follows [20, 21]:

$$TBATS(\omega, p, q, \varphi, \{m_1, k_1\}, \{m_2, k_2\}, \dots, \{m_T, k_T\})$$
(3)

Where ω is the smoothing parameter for the Box-Cox transformation to stabilize the variance of the series, p and q are the parameters of the ARMA process, φ is the trend smoothing parameter, and the pairs $\{m_t, k_t\}$ identify the seasonality and the corresponding terms of the Fourier series [22]. This model is one of the most versatile, considering nonlinear patterns and autocorrelations in the residuals [20]. Auto model selection is based on the AIC criterion by adding and removing transformations, considering or not the trend and its damping version, with or without ARMA (p, q) residual modelling and using several harmonics to determine the seasonal effects if present. The analysis was

performed in R using the function *tbats()* in the package *forecast* [19, 21].

(iv) UCM are powerful tools for decomposing time series data into various underlying components. They combine the adaptability of ARIMA with the interpretability characteristic of smoothing models to capture the complexity of temporal variations through the decomposition into clearly defined components. Its terminology per se highlights the focus on capturing the underlying "structure" of the data rather than just fitting a model to the observed values. They assume that the observed data combines several unobserved components, hence the "unobserved" part of the name. Overall, UCMs are valuable tools for time series analysis when the goal is to gain insights into the underlying structure and driving forces behind the data. They excel at decomposing complex patterns into interpretable components, aiding in understanding past trends and making informed predictions. They can be expressed as [22]:

$$Y_t = \mu_t + \gamma_t + \sigma_t + \varepsilon_t \tag{4}$$

With *Y* the observed series, μ the trend to capture the overall direction, γ the cycle to identify periodic variations, σ , the seasonal component reflects regular and repetitive variations at specific intervals, ε , and the residual error represents fluctuations not explained by the previous combine factors [22]. The model was selected by adding/removing/modifying the components and comparing them with the AIC. The analyses were performed using PROC UCM in SAS software version 9.4 (SAS Institute, Inc., Cary, NC, USA).

Forecast assessment

The historical series was divided into in-sample (60%) and out-of-sample (40%) data for parameter estimation and model validation. The parameters were estimated on the training set, and the selected models were performed on the test set. The more efficient model was used to predict the whole series.

To compare the accuracy of the four models in predicting future values, we calculated the Mean Absolute Percentage Error (MAPE) metric as:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| * 100$$

where *y* is the observed and \hat{y} the fitted series from periods t=1 to N (total number of observations). The metric is independent of scale since it relies on percentage

errors, making it unit-free. Using a percentage measure of accuracy is always preferable when comparing forecasts on different time series [23]. Some disadvantages are infinite or undefined results with zero data points and a bias towards negative errors [24].

Statistical analysis

The socio-demographic and clinical characteristics of the cohort were described through frequencies and percentages for categorical variables and median and interquartile range (IQR) for continuous ones. Descriptive statistics were used to check if the data aligns with the existing literature.

The time series was examined by applying the Seasonal and Trend decomposition using the LOESS (STL) method. This robust approach employs locally fitted regression models with LOESS smoothing to decompose the time series into trend, seasonal, and remainder components. The Augmented Dickey-Fuller and the Seasonal Kendall Test assessed stationarity and monotonic trends, respectively.

We used a counterfactual approach to model the hypothetical epidemiological situation in Italy from March 2020 to February 2022, assuming no COVID-19 outbreak, constructing a model that represents the "normal" trends would have been without the pandemic.

For each of the four-time series models previously mentioned, (i) first, we predicted the months heavily impacted by the COVID-19 outbreak (March 2020 to February 2022) using data on previous epidemiological seasons considering in-sample (September 2010—February 2016) and out-of-sample (March 2016—February 2020) period; and (ii) second, we forecasted the upcoming epidemiological season (October 2023 to September 2024) based on data spanning September 2010 to September 2023 (in-sample: September 2010—August 2018 and out-of-sample: September 2018 – September 2023), which incorporated the previously forecasted COVID-19 period (March 2020 to February 2022).

The underlying assumptions of each model were evaluated through the model's residuals to assess whether (i) the spread of the residuals is consistent (no heteroscedasticity), (ii) the residuals are normally distributed, and (iii) there is any correlation between the residuals at different points in time. The final performance of the different proposed approaches was compared using an appropriate error measure, the MAPE.

A <0.01 two-sided *p*-value was considered significant for all the tested hypotheses. The statistical analyses were performed using SAS software version 9.4 (SAS Institute, Inc., Cary, NC, USA) and R Statistical Software version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). We followed the strengthening of the reporting of observational studies (STROBE).

Results

Over two-thirds of children (236,642 / 345,877) included in the study from September 2010 to September 2023 had at least one ARI diagnosis. Of the 236,642 children, they reported a total of 1.4 million ARI cases. Among them, 769,637 (53.19%) were found in males, with higher prevalence in children aged 0 - 2 (34.09%) and 5 - 11 (35.75%) years at infection. LRTIs were less frequent compared to URTIs (16.75% vs 83.25%, respectively), also presenting a lower median age (3 vs 4 years, respectively) (Table S2).

Figure S1 displays the monthly incidence rate time series of ARIs spanning the study period from September 2010 to September 2023, along with the STL decomposition. Notably, the Augmented Dickey-Fuller test confirms the stationary nature of the time series (p = 0.01). A consistent trend is shown throughout the entire timeframe, featuring a gradual decline in the incidence rate over the years (p < 0.01). The time series revealed distinct seasonal variation, characterized by fluctuations and identifiable peaks in incidence during the winter to early spring months. In the remainder component, outliers emerge in the COVID-19 years, indicating a pronounced decrease in the incidence rate during this period.

Tables 1 shows the chosen models and their referred accuracy in terms of MAPE according to the two different forecast periods: (i) the COVID-19-related period and (ii) the overall period and the forecast assessment method.

The best models for fitting the in-sample period in forecasting the months impacted by the COVID-19 outbreak (i) were a SARIMA with drift, p=P=D=1 and the others parameters equal to zero; an ETS with

multiplicative error, none trend and multiplicative seasonality; a TBATS with a Box-Cox transformation based on value 0.485, no ARMA error, no trend, monthly seasonality and 5 terms of Fourier; and a UCM with irregular, level and slope component and a spline with fourth degree and 6 knots from position four to nine to model the seasonal component. The ETS has produced a lower MAPE (6.93), similar to the TBATS (7.11), compared with the SARIMA (14.60) and the UCM (12.27). By validating the models in the out-ofsample set of data (March 2016 - February 2020), all the models exhibited a lower MAPE except the SARIMA (19.98 vs 14.60, respectively, for in- and outof-sample). As the ETS maintained high performance in the test set (MAPE = 5.43), this model was used to forecast COVID-19 years with an overall MAPE of 6.65 (Table 1A, Fig. 1).

Furthermore, the best models for fitting the whole series (i.e., from September 2010 and September 2023) in forecasting the upcoming epidemiological season (ii) were characterized in the in-sample by a SARIMA with drift, p=D=1, P=2, and the others parameters equal to zero; an ETS with multiplicative error, no trend and multiplicative seasonality; a TBATS with a Box-Cox transformation based on value 0.123, no ARMA error, no trend, monthly seasonality and five terms of Fourier; a UCM with irregular, level, fixed slope and a fourth degree spline with six knots from December to May component. Similar to the previous forecast, this prediction again identifies the ETS and TBATS models as the most accurate, with MAPE values of 6.69 and 6.72, respectively. The UCM and SARIMA models follow, with MAPEs of 11.70 and 13.07, respectively. In the validation phase, the TBATS showed a lower MAPE (6.81), while in the

| Table 1 MAPEs for in-sample and out-of-sample approach in the two dif | fferent forecast period |
|---|-------------------------|
|---|-------------------------|

| Model | In-sample | Out-of-sample | Overall |
|--|-----------|---------------|---------|
| A) September 2010—February 2020 | | | |
| ETS(M,N,M) | 6.93 | 5.43 | 6.65 |
| TBATS(0.485,{0,0},-,{<12,5>} | 7.11 | 5.61 | 7.77 |
| ARIMA(1,0,0)(1,1,0) [12] with drift ^a | 14.60 | 19.98 | 17.62 |
| UCM with irregular, level, slope and seasonality modelled with a fourth degree spline and 6 knots from December to May | 12.27 | 10.40 | 10.97 |
| B) September 2010 – September 2023 with forecasted COVID-1 | 9 | | |
| ETS(M,N,M) | 6.69 | 7.46 | 6.85 |
| TBATS(0.123,{0,0},-,{<12,5>} | 6.72 | 6.81 | 6.87 |
| ARIMA(1,0,0)(2,1,0) [12] with drift ^a | 13.07 | 29.94 | 25.33 |
| UCM with irregular, level, slope and seasonality modelled with a fourth degree spline and 6 knots from December to May | 11.70 | 10.20 | 11.08 |

^a Applied on log-transformed series



Fig. 1 ETS forecast assessment (a) and estimation (b) of the ARI monthly-IRx1000 person-day, seasons 2010 – 2020. Series contains original data; Fitted Train and Fitted Test are referred to the parameters estimation and validation, respectively; Fitted overall are the results of the validated models application; Forecast and the correspondent 95% confidence interval explain the pandemic years prevision (March 2020 – February 2022). Abbreviations: ETS, Error, Trend and Seasonality; ARI, Acute Respiratory Infection; IR, Incidence Rate

overall period, no differences were found in using the TBATS or the ETS (MAPE 6.87 vs 6.85, respectively) (Table 1B, Fig. 2).

By evaluating the parameter selection disparities for the two forecasts utilizing the same modelling framework, we can see that the ETS and the UCM maintained the same structure, the TBATS decreased the Box-Cox transformation parameter, and the SARIMA varied in the autoregressive order.

Discussion

Different time series models were compared to predict the pediatric monthly incidence rate of acute respiratory illnesses (ARIs) in Italy. Initially, we predicted the period from March 2020 to February 2022, hypothesizing the scenario without the COVID-19 outbreak. The ETS (M,N,M) model outperformed others, with a MAPEs of 5.43. Subsequentially, we focused on predicting ARIs for the ongoing 2023/2024 season, factoring in the simulated



Fig. 2 Forecast assessment (**a-c**) and estimation (**b-d**) of the ARI monthly-IRx1000 person-day, seasons 2010–2023. *, Error Trend and Seasonality; **, Trigonometric seasonality, Box-Cox transformation, ARMA errors and Trend and Seasonal components; Series contains original data with the pandemic years forecasted with an ETS model; Fitted Train and Fitted Test are referred to the parameters estimation and validation, respectively; Fitted overall are the results of the validated models application; Forecast and the correspondent 95% confidence interval explain the current epidemiological season prevision. Abbreviations: ARI, Acute Respiratory Infection; IR, Incidence Rate

COVID-19 period. Both TBATS and ETS (M,N,M) emerged again as the most accurate models (MAPE 6.81 and 7.46, respectively), followed by UCM and SARIMA (MAPE 10.20 and 29.94, respectively). The in-sample and out-of-sample approach was adopted to ensure prediction appropriateness.

We selected models from two common forecasting families, ARIMA and Exponential Smoothing (ETS, TBATS, UCM), covering a wide range of forecasting tasks. While some Exponential Smoothing models (linear ones) can be represented by ARIMA models, the more complex non-linear Exponential Smoothing models lack direct ARIMA equivalents. Conversely, there are ARIMA models that cannot be replicated using Exponential Smoothing [20]. In order to achieve homoscedasticity before ARIMA implementation, a logarithmic transformation has been applied.

As we have proposed a model comparison study, using a monthly time series was an acceptable balance between the prediction's accuracy and the model's adaptability to the data. However, if the intention was to do surveillance with the need to make short-term predictions, a weekly time series would be recommended to guide healthcare providers in targeted and prompt interventions.

Our data were representative of the epidemiological burden associated with ARI. We observed that males are more susceptible to contracting ARI, as confirmed by Hasan et al. with 2.9% vs 2.5% and Chobe et al. with 20.5% vs 7.2% of males and females with confirmed ARIs, respectively. [25, 26] Furthermore, consistent with previous literature, children under five years of age were found to be more vulnerable to suffering from ARIs [1, 2].

Our evaluation using metrics like MAPE revealed that the SARIMA model had higher error rates than models like ETS and TBATS when fitted to the ARIs' monthly time series. This suggests that SARIMA may not be able to fully capture the intricate patterns within this specific data set. This aligns with previous studies on infectious

diseases. Wang et al. and Kuan et al. showed that ETS performs better than ARIMA models. [27, 28] In our case, the TBATS and ETS models, which inherently use exponential smoothing techniques, likely excel due to their capability to handle such complexities. This allows for more accurate forecasts compared to more straightforward methods like naive forecasting (e.g., simple average, last value, moving average methods). Exponential smoothing acts as the foundation for this intricate combination. It enables the model to assign different weights to past observations based on their relevance to the current prediction, giving more weight to recent data points while gradually fading out the influence of older ones. However, TBATS has a more complex structure, adding Fourier's term to model seasonality, allowing forecasting more effectively with the anomalies in the last year. Furthermore, the reason why exponential smoothing models outperformed the others can be seen in the decomposition in Fig. 1. The consistent and stable seasonal pattern over time observed in our data drive this superior performance, as the model's ability to effectively capture these patterns enhances its predictive accuracy. Sorokina et al. demonstrated the superiority of TBATS vs SARIMA in predicting severe acute respiratory infections. [29] While the UCM model performed reasonably well in our comparison, it has not received much attention in previous research and lacks essential features for directly modelling ARIs at the community level. Interestingly, all our decomposition-based models concurred in excluding the trend component, suggesting its stability over time and reinforcing the consistency of our model selection process.

A major strength of our study lies in the nature of the data. There is no evidence in the literature of research that analysed community-based ARIs in such a large population of almost three hundred thousand children with more than a million infections. As the data came from the FP ambulatorial practice, the representativeness of the ARI's burden is more accurate with respect to using data from different healthcare sources. This allowed us to train our models in a representative framework to guide future epidemiological burden predictions. Our monthly forecasting of ARI outbreaks is a crucial tool for the public health service to minimize the epidemiological burden of these illnesses and ensure a more efficient response to seasonal fluctuations. This enables them to optimize staffing, medication stockpiles, and other resources to manage the influx of patients effectively. Furthermore, compared to machine learning models, the tested models are easy to implement, still achieving more than satisfactory results.

Our study has potential limitations. First, emerging complex hybrid and machine learning models were developed to study time series data, potentially contributing to improving forecast accuracy. Additionally, the epidemiological season 2022–2023 may still present anomalous characteristics due to the COVID-19 pandemic, biasing new ARI burden estimates. Third, a misclassification is possible in the case in which the FPs do not report the ARI case. Moreover, our predictions are derived from forecasts made during the COVID-19 years. Nevertheless, these years do not influence model selection as they are not included in the in-sample data. Furthermore, the predictions for the new epidemiological season assign greater weight to recent observations (i.e., 18 months after the predicted years).

Conclusion

Although our models showed promising results, incorporating data from the ongoing epidemiological season remains crucial for accurately defining ARI trajectories in the post-COVID-19 era and guiding targeted interventions. Given the inherent variability of ARIs, we reiterate the recommendation of employing multiple time series models to identify the best fit for specific prediction tasks. Furthermore, the user-friendliness and adaptability of the presented models make them valuable tools for researchers and practitioners alike. Additionally, this study highlights the importance of ongoing monitoring and exploring new models to improve our understanding and prediction of pediatric ARI trends, representing a promising strategy for preparedness programs. The clinical implications of these findings should be further explored through larger multicenter studies, encompassing countries from various hemisphere, to better predict the ARIs epidemiology and promptly inform public health entities.

Abbreviations

| ARI | Acute Respiratory Infection |
|----------|--|
| WHO | World Health Organization |
| DALY | Disability Adjusted Life years |
| SIR | Susceptible Infected and Recovered |
| ARIMA | Autoregressive Integrated Moving Average |
| FP | Family Paediatricians |
| ICD-9-CM | International Classification of Diseases Ninth Revision Clinical |
| | Modification |
| LRTI | Lower Respiratory Tract Infection |
| URTI | Upper Respiratory Tract Infection |
| SARIMA | Seasonal Autoregressive Integrated Moving Average |
| ETS | Error Trend and Seasonality |
| TBATS | Trigonometric seasonality Box-Cox transformation ARMA errors |
| | and Trend and Seasonal components |
| UCM | Unobserved Component Models |
| AIC | Akaike Information Criterion |
| MLE | Maximum Likelihood Estimation |
| MAPE | Mean Absolute Percentage Error |
| IQR | Interquartile Range |
| STL | Seasonal and Trend decomposition using the Loess |
| STROBE | Strengthening of the Reporting of Observational studies |
| IR | Incidence Rate |
| | |

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12889-025-21984-1.

Supplementary Material 1.

Acknowledgements

The authors gratefully acknowledge the contributions of Luigi Cantarutti (Pedianet), and all Pedianet family paediatricians (Pedianet): Alongi Angelo, Andreola Barbara, Angelini Roberta, Apostolo Maria Grazia, Arcangeli Bruno, Argo Giovanna, Avarello Giovanni, Azzoni Lucia Barbazza, Maria Carolina, Barbieri Patrizia, Belluzzi Gabriele, Benetti Eleonora, Biasci Paolo, Boe Franca, Bollettini Stefano, Bonaiuto Francesco, Bontempelli Anna Maria, Brusaterra Ivana, Budassi Roberto, Caccini Massimo, Calì Laura, Cantalupi Laura, Cardarelli Chiara Carli Giovanna, Carnazza Sylvia, Castaldo Massimo, Castelli Stefano, Cavedagni Monica, Cera Giuseppe Egidio, Chillemi Chiara, Cichello Francesca, Cicione Giuseppe, Ciscato Carla, Clerici Schoeller Mariangela, Collacciani Giuseppe, Comaita Fabrizio, Conte Valeria, Conte Ugo Alfredo, Corro' Roberta, Costagliola Rosaria, Costanzo Nicola, Cozzani Sandra, Cuboni Giancarlo, Curia Giorgia, D'alia Caterina, D'Amanti Vito Francesco, D'Avino Antonio, De Clara Roberto, De Giovanni Lorenzo, De Marchi Annamaria, Del Ponte Gigliola, Di Giampietro Tiziana, Di Mauro Giuseppe, Di Santo Giuseppe, Di Saverio Piero, Dieli Mattea, Dolci Marco, Doria Mattia, El Mazloum Dania, Fadda Maria Carmen, Falco Pietro Fama Mario, Faraci Marco, Farina Maria Immacolata, Favilli Tania, Federico Mariagrazia, Felice Michele, Ferraiuolo Maurizio, Ferrara Enrico Ferretti Michele, Forcina Paolo, Foti Patrizia, Frison Ezio, Fusco Fabrizio, Gallo Giovanni, Galvagno Andrea, Genova Grazia, Gentilucci Pierfrancesco, Giampaolo Giuliana, Gianfredi Francesco, Girotto Silvia, Grelloni Mauro, Grugnetti Mirco, Isca Antonina, Lagrasta Urania, lisabetta Landi Massimo, Letta Maria Rosaria, Lietti Giuseppe, Lista Cinzia, Lucantonio Ricciardo, Luise Francesco, Macropodio Nadia, Marine Francesca Mariniello Lorenzo, Marostica Gabriella, Masotti Sergio, Meneghetti Stefano, Milani Massimo, Milone Stella Vittoria, Moggia Donatella, Monteleone Angela Maria, Mussinu Pierangela, Naccari Anna, Nicoloso Flavia, Olimpi Laura Maria, Ongaro Riccardo, Palma Maria Maddalena Pasinato Angela, Passarella Andrea, Pazzola Pasquale, Perin Monica, Perri Danilo, Pescosolido Silvana Rosa, Petrazzuoli Giovanni, Petrotto Giuseppe, Picco Patrizia, Pirola Ambrogina, Pisanello Lorena, Pittarello Daniele, Porro Elena, Porto Adolfo Francesco, Puma Antonino, Puocci Maria Paola, Righetti Andrea, Rizzari Rosaria, Rosas Paolo, Ruffato Bruno, Ruggieri Lucia, Russo Annarita, Salamone Pietro, Sambugaro Daniela, Saretta Luigi, Sarno Vittoria, Savio Valentina, Sciolla Nico Maria, Semenzato Rossella, Silvan Carla, Soldà Giorgia, Spanevello Valter Spedale Sabrina, Speranza Francesco, Stefani Sara, Storelli Francesco, Tamassia Gianni, Tambaro Paolo, Toffol Giacomo, Tonelli Gabriele Tulone Silvia, Tummarello Angelo Giuseppe, Venditti Sergio, Vitale Maria Grazia, Volpe Francescopaolo, Volpe Concetta, Vozzi Aldo Zanon Giulia, Zuccolo Mari

Author's contributions

Conceptualization, B.R., C.B., D.C.C., C.A.; Data Curation and Formal Analysis, B.R., C.B., A.S.; Writing – original draft preparation, B.R., C.B., D.C.C., C.A.; Writing – review and editing, F.P., B.E., L.P.G., C.G., D.D., G.C.. All authors read and approved the final manuscript.

Funding

This study was supported by grants from the Italian Ministry of Education, University and Research ('PRIN-PNRR' 2022, project P20224MZE4. Covid-19 and Acute Respiratory Infections: the Clinical and Epidemiological changes in the Pediatric Population: the CARICE project).

Data availability

Data may be obtained from a third part and are not publicly available.

Declarations

Ethics approval and consent to participate

This is an observational, retrospective, non-interventional study. According to a bylaw on the classification and implementation of observational drug-related research, as issued by the Italian National Drug Agency (an entity belonging to the Italian Ministry of Health), this study does not require approval by an ethics committee in Italy (Italian Drug Agency note on August 3, 2007). This study was conducted in accordance with the tenets of the Declaration of Helsinki and was compliant with the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance's Guide on Methodological Standards in Pharmacoepidemiology. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Consent for publication

The study and access to the database was approved by the Internal Scientific Committee of So.Se.Te. Srl, the legal owner of Pedianet. The corresponding author confirm that he had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Competing interests

Giovanni Corrao received research support from the European Community (EC), the Italian Agency of Drug (AIFA), and the Italian Ministry for University and Research (MIUR). He took part in various projects funded by pharmaceutical companies (i.e., Novartis, GSK, Roche, AMGEN, and BMS). He also received honoraria as a member of the Advisory Board from Roche. The remaining authors declare that the research was conducted without commercial or financial relationships that could be construed as potential conflicts of interest. The other authors have no conflict of interest to declare. We affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of any organization or company.

The other authors have no conflict of interest to declare. We affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of any organization or company.

Author details

¹Department of Statistics and Quantitative Methods, Division of Biostatistics, Epidemiology and Public Health, Laboratory of Healthcare Research and Pharmacoepidemiology, University of Milan-Bicocca, Via Bicocca Degli Arcimboldi, 8, Milan 20126, Italy. ²Department of Medicine and Surgery, University of Milan-Bicocca, Via Cadore 48, Monza 20900, Italy. ³Center for Public Health Research, University of Milan-Bicocca, Via Cadore 48, Monza 20900, Italy. ⁴Laboratory of Public Health, IRCCS Istituto Auxologico Italiano, Via Ariosto 13, Milan 20145, Italy. ⁵Department of Women's and Children's Health, University of Padova, Via Giustiniani, 3, Padua 35128, Italy. ⁶Department of Statistics and Quantitative Methods and CRISP, University of Milan-Bicocca, Via Bicocca Degli Arcimboldi, 8, Milan 20126, Italy. ⁷Società Servizi Telematici (SoSeTe), Pedianet Project, Via Giacomo Medici 9/A, Padua 35138, Italy.

Received: 9 May 2024 Accepted: 17 February 2025 Published online: 28 February 2025

References

- 1. Berry M, Gamieldien J, Fielding BC. Identification of new respiratory viruses in the new millennium. Viruses. 2015;7(3):996–1019.
- Shi C, Wang X, Ye S, Deng S, Cong B, Lu B, Li Y. Understanding the risk of transmission of respiratory viral infections in childcare centres: protocol for the DISeases TrANsmission in Childcare (DISTANCE) multicentre cohort study. BMJ Open Respir Res. 2023;10(1):e001617.
- World Health Organization. Mortality and global health estimates. Retrieved from https://www.who.int/data/gho/data/themes/mortalityand-global-health-estimates. Accessed on 22 March 2024.
- Mulholland K. Global burden of acute respiratory infections in children: implications for interventions. Pediatr Pulmonol. 2003;36(6):469–74.
- World Health Organization. Clinical care for severe acute respiratory infection: toolkit, update 2022. COVID-19 adaptation. https://www.who.

int/europe/publications/i/item/WHO-2019-nCoV-SARI-toolkit-2022-1. Accessed on 22 March 2024.

- World Health Organization. Basic epidemiology. Geneva: World Health Organization; 1993.
- 7. Perra N, Gonçalves B. Modeling and Predicting Human Infectious Diseases. Computational Social Sciences: Springer; 2015.
- Box G, Jenkins G. Time series analysis: forecasting and control. San Francisco: Holden-Day; 1970.
- Holt, C. C. Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. 1957.
- Winters PR. Forecasting sales by exponentially weighted moving averages. Manage Sci. 1960;6(3):324–42.
- Rahardja Dewi. Statistical methodological review for time-series data. J Statistics Manag Systems. 2020;23(8):1445–61.
- Panigrahi S, Behera HS. A hybrid ETS–ANN model for time series forecasting. Eng Appl Artif Intell. 2017;66:49–59.
- Panaggio MJ, Rainwater-Lovett K, Nicholas PJ, Fang M, Bang H, Freeman J, Peterson E, et al. A time-series model for COVID-19 hospital admission forecasting. Epidemics. 2022;39:100580.
- Barría-Sandoval C, Ferreira G, Benz-Parra K, López-Flores P. Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study. PLoS ONE. 2021;16(4):e0245414.
- Zeng Y, Zhang Z, Kusiak A, Tang F, Wei X. Optimizing wastewater pumping system with data-driven models and a greedy electromagnetism-like algorithm. Stoch Environ Res Risk Assess. 2016. https://doi.org/10.1007/ s00477-015-1115-4.
- Cantarutti A, Barbieri E, Didonè F, Scamarcia A, Giaquinto C, Corrao G. Influenza Vaccination Effectiveness in Paediatric "Healthy" Patients: A Population-Based Study in Italy. Vaccines (Basel). 2022. https://doi.org/10. 3390/vaccines10040582.
- Barbieri E, Cantarutti A, Cavagnis S, Cantarutti L, Baraldi E, Giaquinto C, et al. Impact of bronchiolitis guidelines publication on primary care prescriptions in the Italian paediatric population. NPJ Prim Care Respir Med. 2021. https://doi.org/10.1038/s41533-021-00228-w.
- Di Chiara C, Barbieri E, Chen YX, Visonà E, Cavagnis S, Sturniolo G, et al. Comparative study showed that children faced a 78% higher risk of newonset conditions after they had COVID-19. Acta Paediatr. 2023. https:// doi.org/10.1111/apa.16966.
- 19. Hyndman, R. J. forecast: Forecasting Functions for Time Series. R package version 2.09; 2010.
- Hyndman, R.J., & Athanasopoulos, G. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia; 2018. Accessed on 22 February 2024.
- De Livera AM, Hyndman RJ, Snyder RD. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. J Am Stat Assoc. 2010. https://doi.org/10.1198/jasa.2011.tm09771.
- 22. Harvey, A. Forecasting, Structural Time Series Models and the Kalman Filter, New York: Cambridge University Press. 1989. https://doi.org/10. 2307/2583225.
- Bowerman B. L., O'Connell R. T., Koehler A.B. Forecasting, Time Series, and Regression: An Applied Approach (4th ed.). Thomson Brooks/Cole. Duxbury applied series; 2005.
- 24. Perone G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. Eur J Health Econ. 2022. https://doi.org/10.1007/s10198-021-01347-4.
- Hasan MM, Saha KK, Yunus RM, Alam K. Prevalence of acute respiratory infections among children in India: Regional inequalities and risk factors. Matern Child Health J. 2022. https://doi.org/10.1007/s10995-022-03424-3.
- Choube A, Kumar B, Mahmood SE, Srivastava A. Potential risk factors contributing to acute respiratory infections in under five age group children. Int J Med Sci Public Health. 2014;3:1385–8. https://doi.org/10.5455/ ijmsph.2014.200820141.
- Wang Y, Xu C, Yao S, Zhao Y, Li Y, Wang L, et al. Estimating the Prevalence and Mortality of Coronavirus Disease 2019 (COVID-19) in the USA, the UK, Russia, and India. Infect Drug Resist. 2020. https://doi.org/10.2147/2FIDR. S265292.
- Kuan MM. Applying SARIMA, ETS, and hybrid models for prediction of tuberculosis incidence rate in Taiwan. PeerJ. 2022. https://doi.org/10. 7717/peerj.13117.

 Sorokina M., Korshukov I., Koichubekov B., Takuadina A. Comparison of forecast performance of severe acute respiratory infection incidents using ARIMA and TBATS models in pre-pandemic and COVID-19 pandemic period. 2023. https://doi.org/10.1063/5.0164979.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.