



Stochastic Modelling of Scalar Time Series of Varicella Incidence for a Period of 92 Years (1928-2019)

Ralitsa Raycheva¹, Ani Kevorkyan², Yordanka Stoilova²

¹ Department of Social Medicine and Public Health, Faculty of Medicine, Medical University of Plovdiv, Plovdiv, Bulgaria

² Department of Epidemiology and Disaster Medicine, Faculty of Medicine, Medical University of Plovdiv, Plovdiv, Bulgaria

Corresponding author: Ralitsa D. Raycheva, Department of Social Medicine and Public Health, Faculty of Medicine, Medical University of Plovdiv, 15A Vassil Aprilov Blvd., 4002 Plovdiv, Bulgaria; Email: dirdriem@gmail.com; Tel.: +359 899 965 439

Received: 15 Mar 2021 ♦ **Accepted:** 29 Apr 2021 ♦ **Published:** 31 Aug 2022

Citation: Raycheva RD, Kevorkyan AK, Stoilova YD. Stochastic modelling of scalar time series of varicella incidence for a period of 92 years (1928-2019). *Folia Med (Plovdiv)* 2022;64(4):624-632. doi: 10.3897/folmed.64.e65957.

Abstract

Introduction: Varicella is an acute, highly contagious disease, characterised by generalised vesicular exanthema caused by the initial infection with varicella zoster virus (VZV) which usually affects children aged 2 to 8 years.

Aim: To analyse the changes of varicella incidence in Bulgaria over the period of 1928–2019.

Materials and methods: The time series analysis is based on the official data for varicella incidence (per 100,000) in Bulgaria for ninety-two years (1928–2019), obtained from three major sources. We utilized the method to construct a time series model of overall incidence (1928–2019) using time series modeller in SPSS v. 25. We followed all three steps of the standard ARIMA methodology to establish the model – identification, parameter estimation, and diagnostic checking.

Results: Stochastic scalar time series modelling of the varicella incidence from 1928 to 2019 was performed. The stochastic ARIMA (0,1,1) was identified to be the most appropriate model. The decomposition of varicella incidence time series into a stochastic trend and a stationary component was reasoned based on the model defined. In addition, we assessed the importance of the long-term and immediate effect of one shock. The long-term forecast was also under discussion.

Conclusions: The ARIMA model (0,1,1) in our study is an adequate tool for presenting the varicella incidence trend and is suitable to forecast near future disease dynamics with acceptable error tolerance.

Keywords

ARIMA models, dynamics, epidemiological forecasting, stationary series, trends

INTRODUCTION

Varicella is an acute, highly contagious disease characterised by generalised vesicular exanthema caused by the initial infection with varicella zoster virus (VZV), which usually affects children aged 2-8 years.^[1] Although the infection is usually mild with self-limited evolution, all non-immune individuals are at risk of infection and com-

plications. Groups at higher risk for severe complications are pregnant women, preterm infants, and immunocompromised patients of all ages. In addition, WHO estimates that the annual global burden of varicella is approximately 140 million cases of which 4.2 million with severe complications and 4200 deaths.^[2] The almost universal natural infection, especially in the absence of vaccine prophylaxis^[3], determines the need for accurate data collection for assess-

ment of the actual size of the epidemic process. Robust epidemiological data is essential for setting the priorities while planning the immunization strategies on national level. A limited number of industrialised countries have introduced varicella vaccination into their childhood immunisation programs, which has led to substantial reductions in varicella-related incidence and mortality.^[2]

Varicella is not on the list of reportable infectious diseases and subject of registration in the European Union, thus the overall burden of disease is not well known and a country-level estimate is challenging.^[4]

Bulgaria is among the countries with national mandatory surveillance system covering the total country population since 1940^[5], but data on incidence dates back to 1928. The analysis of the key epidemiological indicators, including the prevalence of acute infectious diseases (excluding tuberculosis, AIDS and sexually transmitted infections) in Bulgaria, reveals that varicella usually takes the first place in their distribution. In the last eleven years (2008–2019), the proportion ranges from 39.18% (2008 – 22693/57916) with an incidence of 297.02 (per 100,000)^[6] to 47.73% (2017 – 25007/52393) with an incidence of 352.12 (per 100,000)^[7]. The notable exception is 2010, when severe measles outbreak^[8] displaced varicella from the first place of acute contagious diseases distribution^[9]. In Bulgaria, according to the current Ordinance for Registration of Infectious Diseases^[10], varicella cases are classified as possible, probable, and confirmed, predominantly possible.

Ninety-two years of national surveillance of varicella incidence as one of the major intensive epidemiological indicators enables us to outline the trends and make an epidemiological forecast for varicella infection in this country. It should be noted that the stochastic modelling of different infectious disease parameters, although not often, is still available.^[11-14]

AIM

In this context, our study is a continuation of the present trends for new experimental data accumulation with help of the modern achievements of statistics in the study of non-stationary time series.

MATERIALS AND METHODS

Incidence data collection

The time series analysis is based on official data for varicella incidence (per population of 100,000) in Bulgaria for ninety-two years from 1928 to 2019. The information was collected from three major sources: The National Center for Public Health and Analysis (NCPHA), the National Center for Infectious and Parasitic Diseases (NCIPD) and data from the chronicles of Bulgarian epidemiology in the 20th century.^[15]

Statistical analysis

We used the method of time series analysis to construct a time series model of overall incidence (1928–2019) using a Time Series Modeller in SPSS v. 25. We established autoregressive integrated moving average (ARIMA) model for prediction.^[16] A p -value <0.05 was considered statistically significant.

ARIMA model

The ARIMA model was based on the yearly incidence varicella rates (per 100,000 population) from 1928 to 2019 in Bulgaria. We followed all three steps of the standard ARIMA methodology to establish the model: identification, parameter estimation, and diagnostic checking. We analysed the graphs for the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify the possible values of non-seasonal autoregressive order (p), non-seasonal differencing (d), and non-seasonal moving average order (q). Three test procedures were used to determine whether the time series after differencing was stationary or not. In the second step, the ARIMA model parameters was estimated. In the last step, we diagnosed the residual error sequence for white-noise sequence using the Ljung-Box Q test. Finally, the SPSS Expert Modeller has identified and estimated the best fitting ARIMA.

Preliminary test procedures

Motivation ground

We visually inspected **Fig. 1** and concluded that time series of varicella incidence for the period from 1928 to 2019 is not stationary.

Fig. 2 demonstrates the autocorrelation and partial autocorrelation function of this series. The horizontal axis presents the phase response, more precisely the lag number of autocorrelation function (ACF) from 0 to 25 and partial autocorrelation function (PACF) from 1 to 25. The sampling interval is one year. The values of the autocorrelation function are displayed on the vertical axis of the top panel, while the values of the partial autocorrelation function are displayed on the vertical axis of the lower panel. The two horizontal lines, symmetric above and below zero, represent the confidence limits assuming the input signal is white noise. **Fig. 2** demonstrates that autocorrelation function fades with slowly increasing delay. Up to the 21st year, autocorrelations are above the confidence limit. As for the partial autocorrelation function, it has a high isolated peak at a delayed unit, then practically subsides below the confidence limits. As is well known, such behaviour of the lengthwise samples indicates non-stationary series.^[17]

Thus, there is enough evidence for a more in-depth analysis of non-stationarity of the surveyed time series. Appropriate stochastic modelling, on the other hand, necessitates a study of the nature of eventual non-stationar-

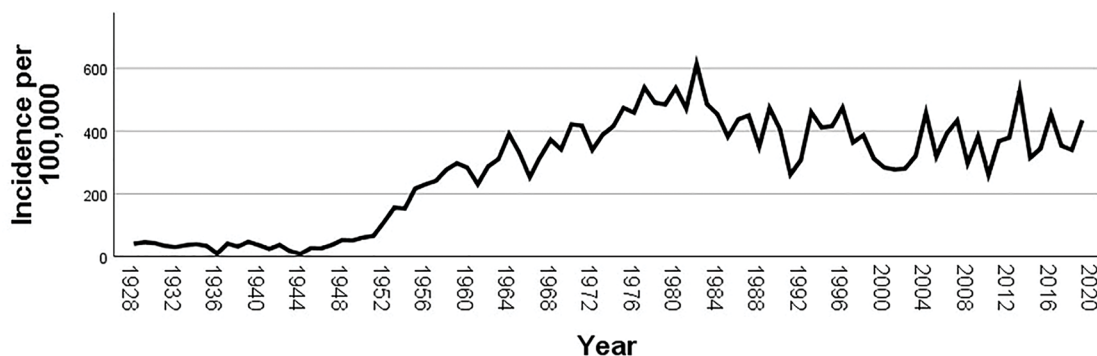


Figure 1. Time series of varicella incidence rate for the period of 1928-2019.

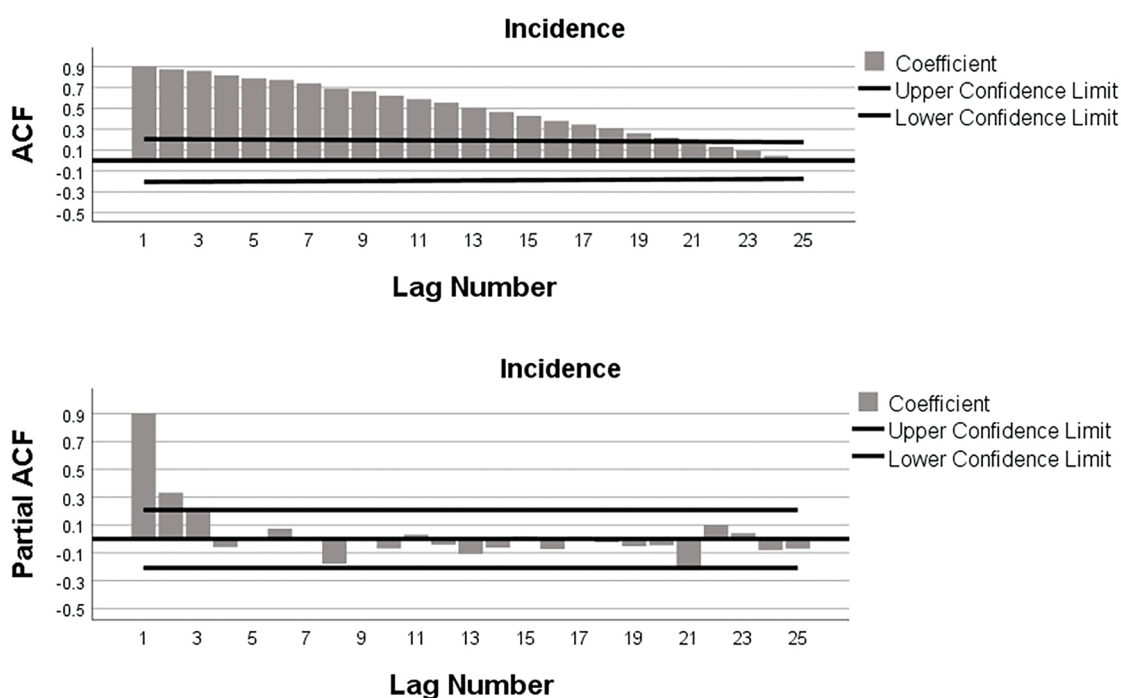


Figure 2. Autocorrelation function (top) and partial autocorrelation function (below) of the varicella incidence.

ity. The contemporary time series analysis elaborates four options: trend stationarity, a unit root (a difference stationary process), structural changes, and long memory. While in the trend stationarity, the original series becomes stationary after proper linear detrending, in the case of a unit root process, this happens after the differentiation of the series. Structural changes in an autoregressive model are of interest as the time series properties of the model, such as stationarity, may be different before and after the change.^[18] Structural changes mean stationarity around a linear trend, but with one or several points of structural changes in which the trend parameters undergo a jump. Long memory processes are positioned between stationarity and unit root. These are time series that become stationary after applying the fractional differentiation filter. When the differentiation indicator is a number between 0.5 and 1, the original series is non-stationary, because its dispersion increases over time.

Time series stationarity check

First, we conducted Dickey-Fuller test to test the null hypothesis that a unit root is present in a time series of varicella incidence.^[19] The testing methodology follows Pfaff.^[20] The test runs in three steps in its entirety. The procedure starts with the regression of differentiated base series on a linear trend, on the original series at a delayed one unit, and on a final number of endogenous variables. Endogenous variables are actually the differentiated base series at a delay of 1, 2, etc. They are included to provide non-correlation of the regression residues. In our case, two endogenous variables are sufficient to satisfy the requirement. We tested the null hypotheses that the time series of varicella incidence is stationary after differentiation. We concluded that the time series was stationary after differentiation without trend and drift in data-generating process (tau criteria=14.44; $p < 0.1$). However, the question remains: whether only one round

of differencing is sufficient to obtain stationary time series. To investigate this issue, we repeated the already conducted test procedure, but not with the original, but with the differentiated time series. It turns out that the regression residuals are uncorrelated even with one endogenous variable. Even in the first step of the procedure, the null hypothesis for post-differentiation stationarity is rejected at significance levels below 1%. According to the following methodology, this result is sufficient to conclude that the one round differentiated time series of varicella incidence is stationary.

Although the Dickey-Fuller test proves that the time series of varicella incidence is stationary after differentiation; however, as is well known, this test has a relatively small power.^[20] Other researchers^[21] suggest a different type of test procedure where the null hypothesis is stationarity against the alternative stationarity after differentiation. Such an approach is consistent with the conservative strategy that it is more natural the hypothesis of interest, in this case post-differentiation, to be the alternative. The test model assumes that the baseline time series is a superposition of trend, random walk, and Gaussian noise. The null hypothesis states that the dispersion of shocks in random walk is zero and therefore the order is trend-stationary. The alternative is non-zero dispersion or stationarity after differentiation. The options to consider are two: stationarity around a linear trend and stationarity to a constant average level. The asymptotic distributions of the relevant statistics are non-trivial and their percentiles are tabulated by the authors.^[21]

The test procedure requires fixed length of the Bartlett (triangular) window, used to calculate the one time step error dispersion. It turns out that the length of this window is short (= 3) delays. The results of the test itself are as follows: In the case of null hypothesis about stationarity around the constant, the value of the corresponding statistics is 2.21 and this hypothesis can be rejected at a significance level of 0.01%. Thus, the stationarity test we conducted rejected the null hypotheses of stationarity around the trend with a slope or around a constant.

It is possible that a time series will remain stationary around a trend with abrupt structural changes. We tested the time series of varicella incidence for this option by applying the so-called stationarity test after differentiation.^[20,22] The null hypothesis is stationarity after differentiation. The alternative is stationarity around a trend with one point of structural change. The options are three: the first – the jump is only in the trend constant; the second – the jump is only in the slope, and the third – the jump is in both the constant and the slope. The point of structural change itself is not predetermined, but assigned in the test procedure, so as to minimally favour the null hypothesis of random walk with drift. Test regressions include the already commented endogenous variables.

The results of this incidence time series test are as follows: in the case of the alternative for structural change only in the trend constant, two endogenous variables are

sufficient for the non-correlation of the residues. The year of the possible structural change is 1985. The value of the relevant statistics is 10.65 and the null hypothesis for post-differentiation stationarity is rejected at a 1% significance level. The conclusion is that the hypothesis of stationarity after differentiation of the time series of varicella incidence is rejected against the alternatives for stationarity with different types of structural change in the linear trend in the first stage of the analysis.

In the end, we will focus on the fourth option of non-stationarity – time series to be stationary after the filtration with fractional differentiation filter. The test procedure^[23] is similar to the rejection of the null hypothesis in the stationarity after differentiation tests, e.g. Dickey-Fuller, and at the same time resembles the rejection of the null hypothesis in the stationary tests, for example the Kwiatkowski-Phillips-Schmidt-Shin test. Herein, we confirmed this assumption. As it was proven, Dickey-Fuller's test rejected the null hypothesis for stationarity after differentiation. Hence, the fractional differentiation process seems unlikely compared with the likelihood of stationarity after differentiation.

In conclusion, the applied test procedures confirm the conclusion that the non-stationarity of varicella incidence time series is stationarity after differentiation type. The differentiation of the time series leads to stationarity, which seems to be not only without a slope, but also without drift.

Stochastic modelling

One-round differentiation of varicella incidence time series was sufficient to ensure stationarity. This result suggests Box-Jenkins modelling with the appropriate integrated autoregressive moving average (ARIMA).^[24,25] The procedure starts with a preliminary analysis of the differentiated time series, its autocorrelation function, and partial autocorrelation function. **Fig. 3** presents the differentiated varicella incidence time series for 1928-2019 period and the corresponding autocorrelation and partial autocorrelation functions. The maximum lag number for both is 25.

The visual examination of the differentiated time series leads to the conclusion that it is stable in terms of variations in dispersion over time. **Fig. 3** also demonstrates that autocorrelation function and partial autocorrelation function have apparent peaks above the lower confidence limit at lag 1, then subsided within the confidence band. Exceptions are single local peaks at lag 8 and lag 21 for autocorrelation and lag 21 for partial autocorrelation, which are within the limits of significance. These observations lead to the conclusion that the appropriate stochastic model is among the many of the ARIMA models with a single differentiation, the maximum order of the autoregressive polynomial – one, the maximum order of the moving average polynomial – also one.

At this stage of the analysis, a question of crucial importance is how to define the “best fit” among the already outlined set of models. To respond to the challenge, we took advantage of the opportunity offered by IBM's Statistical

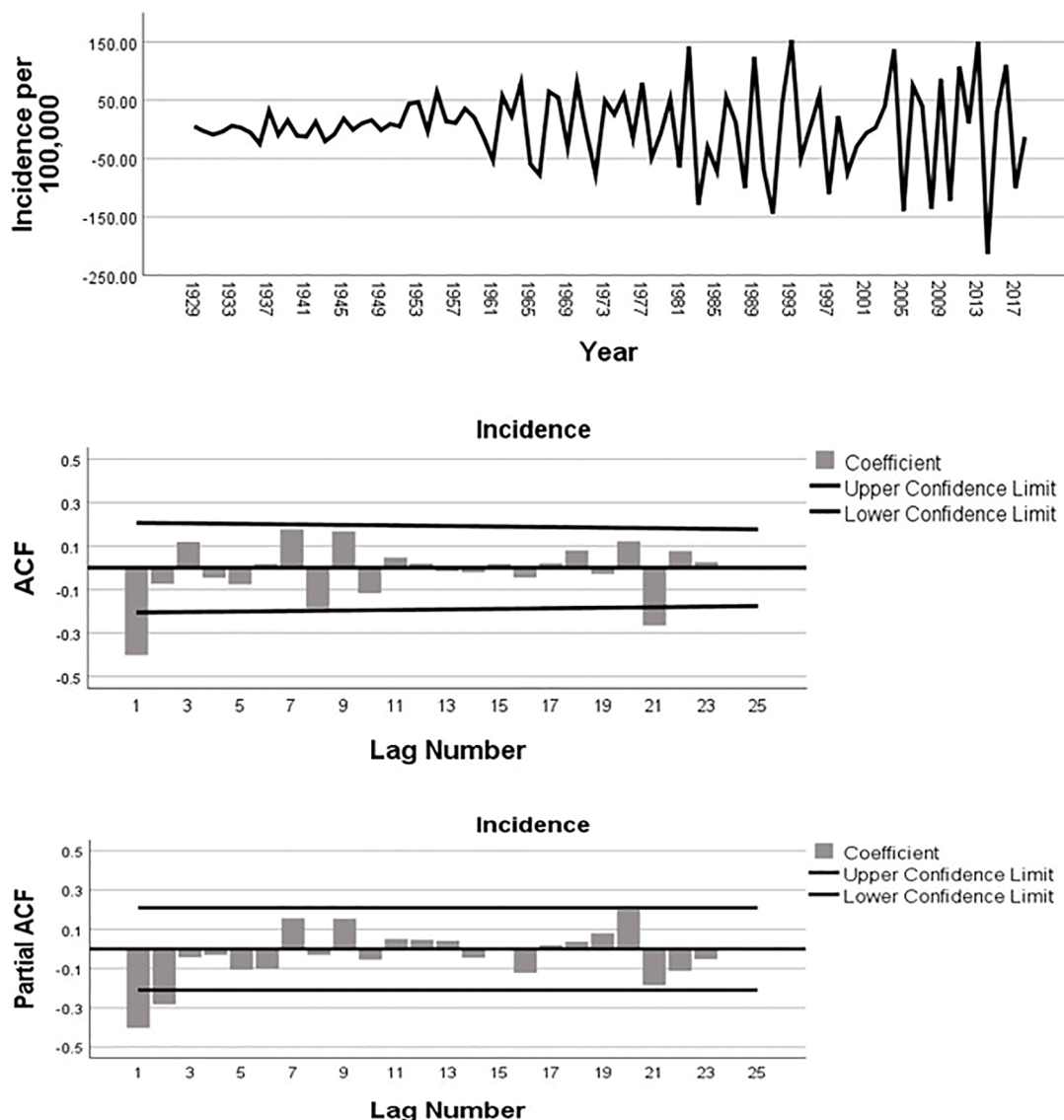


Figure 3. Time series (top), autocorrelation function (middle) and partial autocorrelation function (below) of the differentiated time series of varicella incidence.

Package for Social Sciences (SPSS). IBM SPSS Forecasting Analysis contains a Time Series Module in which Expert Modeller automatically identifies and estimates the best fitting ARIMA or exponential smoothing model for one or more dependent variable series. Although users can specify a custom ARIMA manually, Expert Modeller eliminates a great deal of the trial and error associated with doing so. As a result of the application of Expert Modeller, the ARIMA model (0,1,1) was favoured. The prediction equation for ARIMA (0,1,1) model can be written in a number of mathematically equivalent forms, one of which is the so-called “error correction” form, in which the previous forecast is adjusted in the direction of the error it made:

$$\hat{Y}_t = \hat{Y}_{t-1} + \alpha e_{t-1},$$

because $e_{t-1} = \hat{Y}_{t-1} - Y_{t-1}$ by definition, this can be rewritten as:

$$\hat{Y}_t = Y_{t-1} - (1-\alpha)e_{t-1} = Y_{t-1} - \theta_1 e_{t-1}$$

The estimation of the coefficients of the selected model using the maximum likelihood method results in:

$$\hat{Y}_t = Y_{t-1} - 0.404e_{t-1}$$

Model diagnostics

Stochastic ARIMA models suggest normal distribution and uncorrelated residues with a constant dispersion over time. The normal distribution was checked with the Shapiro-Wilk test. The value of W-statistics in the Shapiro-Wilk test is 0.99 and the corresponding p-value is 0.483. The hypothesis of normal distribution is not rejected. The final conclusion is that the residues are normally distributed. **Fig. 4** shows the autocorrelation function of the residues. It

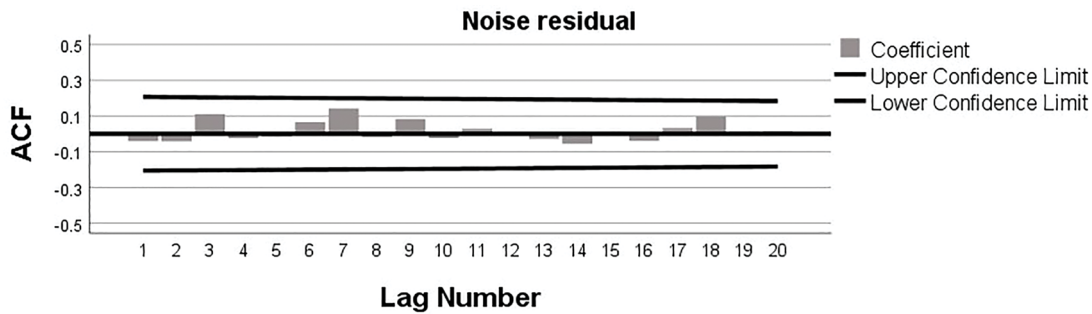


Figure 4. Diagnostics of the ARIMA model (0,1,1) residues.

is obvious that they are within the confidence limits, indicating non-correlation. On the bottom panel the *p*-values of the Ljung-Box-Pierce statistics are given. This test takes into account the magnitude of autocorrelations not separately for each lag, but as a group.^[26] The null hypothesis is uncorrelated residues and it is not rejected at the values obtained – Ljung-Box $Q=7.34, p=0.979$.

One-step in-sample forecast

Fig. 5 presents the time series of varicella incidence for a 92-year period and its approximation to the one-step forecast performed by the selected ARIMA model. The black curve represents the base case series, and the gray curve reflects the approximation, the dashed line sets the upper and lower limits of 95% CI. Apparently, the model provides a very good approximation to the varicella time course and well fits the incidence dynamics. This is one step in-sample forecast that demonstrates that the model is completely acceptable from a formal statistical point of view.

Extended out-of-sample forecast

As a general rule, the long-term forecasts of stationarity after differentiation processes are prone to failure due to the

indefinite accumulation of forecast errors over time. Here we will demonstrate a relatively extended forecast, but with a limited horizon outside the sample. In **Fig. 6**, the forecast is presented by a bold gray continuous line, and the dashed curves set the 95% CI limits. **Table 1** contains the values of the monitored data for the period from 2013 to 2019, the forecast values and the upper and lower limits of 95% CI. It is noteworthy that the actual incidence rates fall within the 95% CI range. Comparing our forecast (**Table 1**), and in particular the incidence rate for 2019 (375.21 per 100,000; 95% CI 167.43–640.06) with the officially registered one for the same year in the country rate (434.19 per 100,000), we could assume that the ARIMA (0,1,1) model is accurate and provides a good fit for the epidemiological data set of varicella incidence. We could also hypothesize that in the next 2-3 years we are expecting an increase in varicella incidence.

DISCUSSION

The proper identification procedures we applied indicated that ARIMA (0,1,1) model fits best our original time series of the annual varicella incidence for the period from 1928 to 2019. In trend-stationary models, characterised by given common properties, the time series returns to the trend,

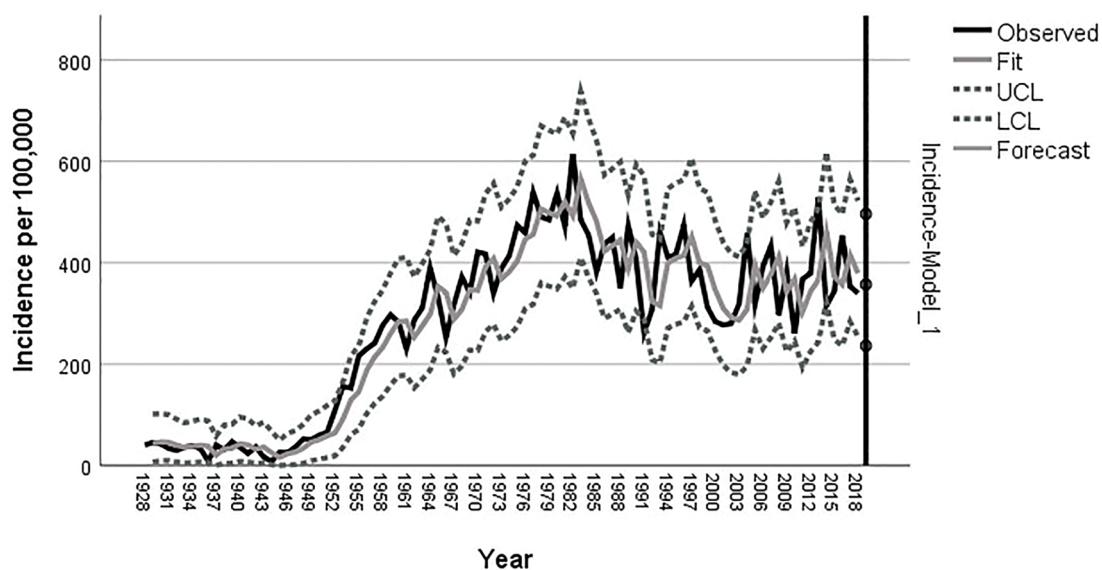


Figure 5. Time series and its ARIMA model (0,1,1) approximation.

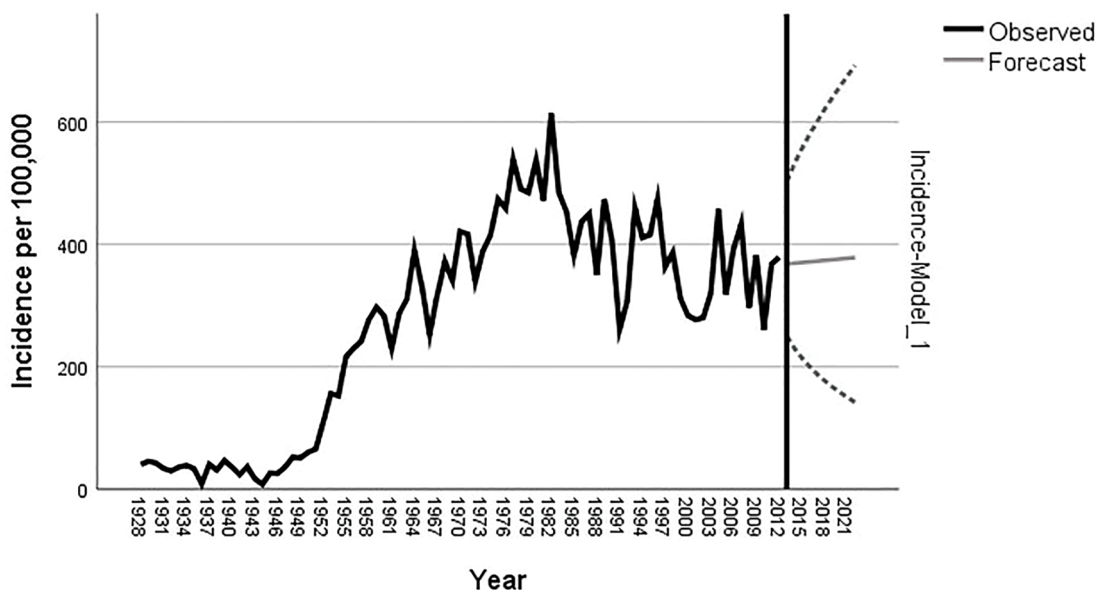


Figure 6. Time series and ARIMA model (0,1,1) 10 years ahead forecast.

Table 1. Time series, forecast, and 95% CI interval for varicella incidence

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Observed	528.97	315.28	343.88	453.94	353.12	339.90	434.19			
Forecast	368.32	369.47	370.61	371.76	372.91	374.06	375.21	376.36	377.50	378.65
ARIMA (0,1,1)										
95% LCL	250.02	230.58	214.47	200.59	188.36	177.39	167.43	158.31	149.91	142.11
95% UCL	502.96	531.49	556.68	579.64	600.96	621.02	640.06	658.26	675.75	692.63

in particular, at stationarity – to the average.^[27] The long-term forecast befits the trend. Unanticipated disturbances, commonly called shocks, naturally fade over time. On the contrary, if the time series is stationary after differentiation, the theory requires distinctly opposite interpretation.^[27] The time series does not return permanently to the trend or average. The dispersion of the prediction error grows over time. The shocks have long-lasting effect.

Our model is stationary after differentiation without drift. Hence, as long as the observed time series values do not tend to zero, the varicella incidence should not be steadily decreasing. The dispersion growth over time impacts the forecast accuracy. The effect of random shocks on incidence is long-lasting, but is smaller than the immediate $\psi(1)=0.09$.

Varicella incidence for the time period 1928-2019 demonstrates upward trend with an average of 280.14 (per 100,000). In the sub-period 1928-1951, the incidence was lower (between 40.40 and 65.60 per 100,000) with an average of 36.20 (per 100,000), most likely related to partial registration of the infection and incidence during that period. In 1952, the indicator for the first time exceeded 100 per 100,000 population, and in 1955 it exceeded 200 per 100,000 population (a rate below which the incidence did not fall throughout the overall sub-period from 1955 to 2019). During the 1955-2019 sub-period, the lowest registered rate was 230.40 (per 100,000) in 1956 and 1961, and the highest rate recorded was 614.29 (per 100,000) in

1982. Based on this evidence, the logical assumption is that the incidence rate will continue to increase in future decades. In the absence of varicella vaccination as a regular component for epidemic process modelling, we observed typical epidemics cycles at short intervals of 2-5 years.^[2] However, using the ARIMA model (0,1,1), the established trend demonstrates that the varicella incidence registered in Bulgaria increases over a longer interval of time – about 10 years (Fig. 4), which is also observed in the USA.^[28] It should be noted that the ten-year peaks in the United States data have been observed during a period of immune prophylaxis started in 1995. Accordingly, we could suggest that if the varicella vaccine is introduced in our immunisation schedule, the epidemic upsurges would be recorded even less frequently. Moreover, Bulgaria is one of the only seven countries in the European Union / European Economic Area that have no specific recommendations for varicella vaccination in the national immunisation programs.^[29]

CONCLUSIONS

Most of the available articles and reports studying varicella infection in Bulgaria are descriptive with analyses of rates (incidence, mortality, and lethality) and proportions (age distribution of the patients, number of hospitalised patients, etc.) and clinical focus on the severity of illness and

complications. To our knowledge, the presented ARIMA model is the first in this country to outline the dynamics of the epidemic process over a period of ninety-two years. Moreover, the established trend provides robust epidemiological forecast to inform health authorities and support evidenced based decision-making for the inclusion of varicella vaccination into the immunisation schedule of the country. The ARIMA model (0,1,1) in our study is an adequate tool for presenting the varicella incidence trend and is suitable to forecast near future disease dynamics, with acceptable error tolerance.

REFERENCES

1. Arvin AM. Varicella-zoster virus. *Clin Microbiol Rev* 1996; 9:361–81.
2. World Health Organization. Varicella and herpes zoster vaccines: WHO position paper, June 2014 – recommendations. *Vaccine* 2016; 34:198–9.
3. Helmuth IG, Poulsen A, Suppli CH, et al. Varicella in Europe – a review of the epidemiology and experience with vaccination. *Vaccine* 2015; 33(21):2406–13.
4. Riera-Montes M, Bollaerts K, Heininger U, et al. Estimation of the burden of varicella in Europe before the introduction of universal childhood immunization. *BMC Infect Dis* 2017; 17(1):1–6.
5. Stefanoff P, Polkowska A, D'Ancona FP. Varicella and herpes zoster surveillance and vaccination recommendations 2010–2011. Rome, Italy: VENICE II Consortium. 2011.
6. Kozhuharova M, Parmakova K, Kurchatova A, et al. [Acute infectious diseases in Bulgaria in 2008 (main epidemiological indicators)]. *Information Journal NCIPD* 2008 [Bulgarian].
7. National Centre for Infectious and Parasitic Diseases. Annual analyses. Available from: <https://www.ncipd.org/images/UserFiles/File/Epidemiologia/analizi%20epidemiologia.rar>
8. Muscat M, Marinova L, Mankertz A, et al. The measles outbreak in Bulgaria, 2009–2011: An epidemiological assessment and lessons learnt. *Euro Surveill* 2016; 21(9):30152.
9. Kozhuharova M, Kurchatova A, Vladimirova N, et al. [Acute infectious diseases in Bulgaria in 2010 (main epidemiological indicators)] *Information Journal NCIPD* 2010 [Bulgarian].
10. Ordinance No. 21 of 18 July 2005 regarding the registration, communication and reporting of infectious diseases force from 01.01.2006, issued by the Minister of Health (State Gazette, issue 62 of 29 July 2005, amended SG, issue 52 of 8 July 2011, amended and supplemented SG, issue 56 of 8 July 2014, amended and supplemented SG, issue 5 of 15 January 2019).
11. Addy CL, Longini IM, Haber MS. A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* 1991; 47:961–74.
12. Trottier H, Philippe P, Roy R. Stochastic modeling of empirical time series of childhood infectious diseases data before and after mass vaccination. *Emerg Themes Epidemiol* 2006; 3:9–24.
13. Funk S, Camacho A, Kucharski AJ, et al. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics* 2018; 22:56–61.
14. Ming RX, Liu J, Cheung WK, et al. Stochastic modelling of infectious diseases for heterogeneous populations. *Infectious diseases of poverty*. *Infect Dis Poverty* 2016; 5(1):107.
15. Georgiev P, editor. [Chronicles of Bulgarian Epidemiology in the 20th Century.] Sofia: Color Studio Publishing House; 2000 [Bulgarian].
16. Chapter 7. The ARIMA procedure. In: SAS/ETS19. 3 User's Guide. Cary, NC: SAS Institute Inc.; 2011 [ISBN 978-1-60764-911-3]
17. Brockwell PJ, Davis RA. Introduction to time series and forecasting. 2nd ed. New York: Springer-Verlag, Inc; 2002.
18. Pang T, Zhang D, Chong TT. Asymptotic inferences for an AR(1) model with a change point: Stationary and nearly non-stationary cases. *J Time Ser Anal* 2014; 35(2):133–50.
19. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 1979; 74(366a):427–31.
20. Pfaff B. Analysis of integrated and cointegrated time series with R. Springer Science & Business Media 2008.
21. Kwiatkowski D, Phillips PC, Schmidt P, et al. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J Econom* 1992; 54(1-3):159–78.
22. Zivot E, Andrews D. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *J Bus Econ Stat* 1992; 10(3):251–70.
23. Kirchgässner G, Wolters J, Hassler U. Introduction to modern time series analysis. 2nd ed. Berlin Heidelberg: Springer-Verlag; 2013.
24. Nelson BK. Time series analysis using autoregressive integrated moving average (ARIMA) models. *Acad Emerg Med* 1998; 5:739–44;
25. Sato RC. Disease management with ARIMA model in time series. *Einstein* 2013; 11:128–31.
26. Shumway RH, Stoffer DS. Time series analysis and its applications with R examples. Springer Science & Business Media 2006.
27. Hamilton JD. Time series analysis. Princeton University Press. 1994.
28. Bakker KM, Martinez-Bakker M, Helm B, et al. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proc Natl Acad Sci USA* 2016; 113(24):6689–94;
29. European Center for Disease Prevention and Control. Varicella vaccination in the European Union. Stockholm: ECDC, 2015.

Стохастическое моделирование скалярных временных рядов заболеваемости ветряной оспой за период 92 года (1928 – 2019 гг.)

Ралица Райчева¹, Ани Кеворкян², Йорданка Стоилова²

¹ Кафедра социальной медицины и общественного здравоохранения, Факультет медицины, Медицинский университет – Пловдив, Пловдив, Болгария

² Кафедра эпидемиологии и медицины чрезвычайных ситуаций, Факультет медицины, Медицинский университет – Пловдив, Пловдив, Болгария

Адрес для корреспонденции: Ралица Д. Райчева, Кафедра социальной медицины и общественного здравоохранения, Факультет медицины, Медицинский университет – Пловдив, бул. „Васил Априлов“ № 15А, 4002 Пловдив, Болгария; Email: dirdriem@gmail.com; Тел.: +359 899 965 439

Дата получения: 15 марта 2021 ♦ **Дата приемки:** 29 апреля 2021 ♦ **Дата публикации:** 31 августа 2022

Образец цитирования: Raycheva RD, Kevorkyan AK, Stoilova YD. Stochastic modelling of scalar time series of varicella incidence for a period of 92 years (1928-2019). Folia Med (Plovdiv) 2022;64(4):624-632. doi: 10.3897/folmed.64.e65957.

Резюме

Введение: Ветряная оспа – острое высококонтагиозное заболевание, характеризующееся генерализованной везикулярной экзантемой, вызванной первоначальной инфекцией вирусом ветряной оспы (ВВО), которым обычно страдают дети в возрасте от 2 до 8 лет.

Цель: Проанализировать динамику заболеваемости ветряной оспой в Болгарии за период 1928 – 2019 гг.

Материалы и методы: Анализ временных рядов основан на официальных данных о заболеваемости ветряной оспой (на 100 000 населения) в Болгарии за 92 года (1928 – 2019 гг.), полученных из трёх основных источников. Мы использовали этот метод для построения модели временных рядов общей заболеваемости (1928 – 2019 гг.) с использованием модуля моделирования временных рядов в SPSS v. 25. Мы выполнили все три шага стандартной методологии ARIMA для создания модели – идентификация, оценка параметров и диагностика. проверка.

Результаты: Было выполнено стохастическое скалярное моделирование временных рядов заболеваемости ветряной оспой с 1928 по 2019 год. Стохастическая ARIMA (0,1,1) была признана наиболее подходящей моделью. Разложение временных рядов заболеваемости ветряной оспой на стохастический тренд и стационарную составляющую было обосновано на основе определённой модели. Кроме того, мы оценили важность долгосрочного и немедленного эффекта одного шока. Обсуждался и долгосрочный прогноз.

Заключение: Модель ARIMA (0,1,1) в нашем исследовании является адекватным инструментом для представления тенденции заболеваемости ветряной оспой и подходит для прогнозирования динамики заболевания в ближайшем будущем с допустимой погрешностью.

Ключевые слова

модели ARIMA, динамика, эпидемиологическое прогнозирование, стационарные ряды, тренды
