



Przegląd Epidemiologiczny

Epidemiological Review



QUARTERLY
JOURNAL OF THE NATIONAL INSTITUTE OF PUBLIC HEALTH
-NATIONAL INSTITUTE OF HYGIENE
AND THE POLISH SOCIETY OF EPIDEMIOLOGY AND INFECTIOUS DISEASES

Index Copernicus 100,98 pkt

MNiSW 40 pkt

VOLUME 74	2020	No 2
------------------	-------------	-------------

CONTENTS

Full text: www.przglepidemiol.pzh.gov.pl

EDITORIAL / ARTYKUŁ REDAKCYJNY

M Gromulska: **Centenary of Przegląd Epidemiologiczny / Stulecie Przeglądu Epidemiologicznego** 155

HISTORY OF MEDICINE / HISTORIA MEDYCYNY

B Cichy: **An analysis of deaths in Tuliszków parish caused by a cholera epidemic in 1831/ Analiza zgonów w parafii Tuliszków spowodowanych epidemią cholery w 1831 roku** 159

R E Paliga: **Evolution of epidemiological thought in medicine taking into account the evolution of Polish epidemiology in the 20th century / Ewolucja myśli epidemiologicznej medycynie z uwzględnieniem kształtowania się polskiej epidemiologii w XX wieku** 169

R E Paliga: **Quarantine as a tool of epidemic fight / Kwarantanna jako narzędzie do walki z epidemiami** ... 180

EPIDEMIOLOGICAL CHRONICLE / KRONIKA EPIDEMIOLOGICZNA

A Wiktor, M Stępień: **Hepatitis B in Poland in 2018 / Wirusowe zapalenie wątroby typu B w Polsce w 2018 roku** 196

K Zakrzewska, M Stępień, M Rosińska: **Hepatitis C in Poland in 2018 / Wirusowe zapalenie wątroby typu C (wzw C) w Polsce w 2018 roku** 209

K Szmulik-Misiurek, M Niedźwiedzka-Stadnik, M Rosińska: **HIV and AIDS in Poland in 2018 / Zakażenia HIV i zachorowania na AIDS w Polsce w 2018 roku** 223

M Korzeniewska: **Tuberculosis in Poland in 2018 / Gruźlica w Polsce w 2018 roku** 239

COVID-19 / COVID-19

A Jarynowski, M Wójta-Kempa, V Belik: **Trends in interest of COVID-19 on Polish Internet / Trendy w zainteresowaniu tematem COVID-19 w polskim Internecie** 258

M Fathi, K Vakili, N Deravi, S Yaghoobpoor, E Ashan, M Mokhtari, M Moshfeghi, M Vaezjalali: Coronavirus diseases and pregnancy: COVID-19, SARS, and MERS / COVID-19, SARS, MERS a ciąża	276
G Ręka, A Korzeniowska, H Pieciewicz-Szczęsna: The influence of vaccination against tuberculosis with the Bacillus-Calmette-Guérin (BCG) vaccine on COVID-19 incidence and mortality – review of the literature / Wpływ szczepienia przeciwko gruźlicy szczepionką Bacillus-Calmette-Guérin (BCG) na zachorowania i umieralność z powodu COVID-19 – przegląd literatury.....	290
A Wawrzyk, D Rybitwa, P Cywiński, R Pióro, M Łobacz: Prevention of SARS-COV-2 coronavirus spread at the Auschwitz-Birkenau State Museum in Poland. The most visited Memorial site in the world during the COVID-19 pandemic / Zapobieganie rozprzestrzenianiu się kpronawirusa SARS-COV-2 w Państwowym Muzeum Auschwitz-Birkenau w Polsce, najczęściej na świecie odwiedzanym miejscu pamięci podczas pandemii COVID-19	303

PROBLEMS OF INFECTIONS / PROBLEMY ZAKAŻEŃ

E Bojkiewicz, K Toczyłowski, A Sulik: Tick-borne encephalitis – a review of current epidemiology, clinical symptoms, management and prevention / Kleszczowe zapalenie mózgu - przegląd aktualnej epidemiologii, objawów klinicznych, postępowania i profilaktyki	316
M Paciorek, C Bieńkowski, D Krogulec, A Bednarska, M Kowalczyk, M Makowiecki, D Bursa, A Skrzat-Klapaczyńska, K Perlejewski, M Radkowski, T Laskus, A Horban: Differences and similarities in clinical manifestations of <i>Listeria monocytogenes</i> and <i>Mycobacterium tuberculosis meningitis</i> / Podobieństwa i różnice w obrazie klinicznym u pacjentów z zapaleniem opon mózgowo-rdzeniowych o etiologii <i>Listeria monocytogenes</i> i <i>Mycobacterium tuberculosis</i>	326
R Słowik, M Wałaszek, W Zieńczuk, A Różańska, J Wójkowska-Mach, M Kołpa, Z Wolak, Ł Kawik: Analysis of the incidence of surgical site infections after open reposition of long bone fractures and closed fracture settings in a 7-year follow-up in an orthopedic and trauma ward in southern Poland / Analiza występowania zakażeń miejsca operowanego po zabiegach otwartej repozycji złamań kości długich i zamkniętych nastawieniach złamań w 7-letniej obserwacji w oddziale ortopedyczno-urazowym w południowej Polsce.....	336
S Bohdanov, Y Polyvianna, T Chumachenko, D Chumachenko: Forecasting of salmonellosis epidemic proces in Ukraine using autoregressive integrated moving average model / Prognoza epidemicznego procesu salmonelozy na Ukrainie za pomocą modelu rejestracji autoregresyjnej średniej przesuwnej	346
J Nowowiejska, A Baran, I Flisiak: Rare coexistence of unilateral erythema nodosum with erysipelas in the area of previous adder bite / Rzadki przypadek współwystępowania jednostronnego rumienia guzowatego i róży w miejscu pierwotnego ukąszenia przez żmiją zygzakowatą.....	355

PUBLIC HEALTH / ZDROWIE PUBLICZNE

R Wierzejska, B Wojda: Folic acid supplementation in pregnancy and prevention of fetal neural tube defects / Przyjmowanie kwasu foliowego przez kobiety ciężarne w profilaktyce wad cewy nerwowej płodu	362
P Grabowska, M Lewoc, O Jakubowicz-Zalewska, P Mierzejewska, A Baran, I Flisiak: What do you know about sexually transmitted infections? Survey concerning knowledge and risky behaviors among students of Białystok universities / Co wiesz o infekcjach przenoszonych drogą płciową? Badanie ankietowe dotyczące wiedzy i ryzykownych zachowań wśród studentów białostockich uczelni wyższych.....	370
INSTRUCTIONS FOR AUTHORS / INSTRUKCJA DLA AUTORÓW	383

Serhii Bohdanov¹, Yulia Polyvianna², Tetyana Chumachenko², Dmytro Chumachenko¹

FORECASTING OF SALMONELLOSIS EPIDEMIC PROCESS IN UKRAINE USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL

PROGNOZA EPIDEMICZNEGO PROCESU SALMONELOZY NA UKRAINIE ZA POMOCĄ MODELU REJESTRACJI AUTOREGRESYJNEJ ŚREDNIEJ PRZESUWNEJ

¹National Aerospace University "Kharkiv Aviation Institute"¹,

²Kharkiv National Medical University

³Department of Microbiology, Jagiellonian University, Kraków, Poland, Polish Society of Hospital Infections

¹Narodowy Aerokosmiczny Uniwersytet „Charkowski Awiacyjny Instytut”,

²Charkowski Narodowy Uniwersytet Medyczny

ABSTRACT

The article highlights the problem of salmonellosis among the population of the Kharkov region, Ukraine. Three time series were used for calculations: a series of incidence rates for men, a series of incidence rates for women and a series of incidence rates for the general population, each of the series was an ordered set of monthly values from December 2015 to December 2018. It was revealed that the most effective tool for analyzing these statistical data is the use of the autoregressive moving average model (ARIMA). The following steps were used: identification and replacement of outliers, the use of smoothing and decomposition of the series. The developed model allows you to explicitly indicate the order of the model using the `arima()` function or automatically generate a set of optimal values (p, d, q) using the `auto.arima()` function. The validated model allows to calculate the predicted values of the incidence of salmonellosis for 50 days. In certain cases, models of exponential smoothing are able to give forecasts that are not inferior in accuracy to forecasts obtained using more complex models.

Key words: *salmonellosis incidence, time series, prognosis, autoregressive moving average model (ARIMA), autocorrelation graphs, validated model, and exponential smoothing model*

STRESZCZENIE

W artykule opisano problem infekcji salmonelozą w świecie, a także konieczność przewidywania chorób zakaźnych na przykładzie występowania salmonelozy wśród ludności obwodu Charkowskiego, Ukraina. Do obliczeń zostały wykorzystane trzy tymczasowe szeregi: szereg poziomu zachorowalności mężczyzn, szereg poziomu zachorowalności kobiet i szereg poziomu występowania w ogólnej populacji. Każdy z rzędów stanowił uporządkowany zestaw miesięcznych wartości od grudnia 2015 do grudnia 2018 roku. Stwierdzono, że najbardziej skutecznym narzędziem do analizy tych danych statystycznych jest wykorzystanie modelu autoregresyjnej średniej kroczącej (ARIMA). Zostały wykorzystane takie kroki: identyfikacja i wymiana odstających, wykorzystanie wygładzania i dekompozycja szeregów. Opracowany model pozwala określić kolejność modelu za pomocą funkcji `arima()` lub automatycznie wygenerować zestaw optymalnych wartości (p, d, q) za pomocą funkcji `auto.arima()`. Zatwierdzony model pozwala na obliczenie prognozy wartości częstości występowania salmonelozy na 50 dni. W niektórych przypadkach modele wygładzania wykładniczego są w stanie generować prognozy tak dokładne, jak te uzyskiwane przy użyciu bardziej złożonych modeli.

Słowa kluczowe: *częstość występowania salmonelozy, szereg czasowy, prognoza, model autoregresyjnej średniej kroczącej (ARIMA), grafiki autokorelacji, zatwierdzony model i model wykładniczego wygładzania*

INTRODUCTION

Salmonella is a genus of bacteria that belongs to the Enterobacteriaceae family and consists of microorganisms that have related phenotypic and genotypic properties. These are gram-negative bacilli, which are intracellular facultative anaerobes. The main reservoirs and sources of salmonella are many species of animals and birds. Animals are infected by the fecal-oral mechanism of transmission, at the same time transovarial transmission of salmonella is possible in birds. A human is susceptible to this infection and can become infected through eating foods that have been contaminated with salmonella (primary or secondary), as well as after contact with animals (patients or carriers). Under favorable conditions, microorganisms of the genus *Salmonella* can live outside the living organism for several months, and even multiply in some products (1 – 3). The industrial poultry industry and the food industry have contributed to the growth of salmonellosis worldwide in recent years, as evidenced by the registration of outbreaks of salmonellosis in different countries (4, 5, 6).

In the United States for the period from 2013 to 2018, the average annual incidence rate of salmonellosis was 15.4 per 100 thousand people (7).

In 2017, 19.7 cases of salmonellosis per 100 thousand people were recorded in the countries of the European Community. However, the incidence of salmonellosis differs in various countries of this region. During 2013 - 2017 statistical reports from the countries of the European Community have not shown that the incidence of salmonellosis is increasing or decreasing. Thus, seven countries of the community reported about increase of cases of this disease, while four countries reported about decrease of salmonellosis cases over the period 2013 - 2017 (8).

Treatment of cases of salmonellosis is much more expensive, less successful and more dangerous for the patient than its prevention. This highlights the tasks of organizing measures for the identification and treatment of patients with salmonellosis, as well as for optimal spending of funds in such events (9). One of the most effective methods for solving such problems is to build a mathematical model that describes the processes of infection spreading in a population, the development of a disease, and predicting the epidemic process (10). The modern development of information technology allows to achieve high accuracy of the constructed models (11). This will help in making of effective decisions by epidemiologists to reduce the incidence of disease (12).

WPROWADZENIE

Salmonella – rodzaj bakterii, który należy do rodziny Enterobacteriaceae i składa się z mikroorganizmów, które mają fenotypowe i genotypowe właściwości. Są to Gram-ujemne pałeczki, które są wewnątrzkomórkowymi fakultatywnymi beztlenowcami. Głównymi źródłami *Salmonelli* jest wiele rodzajów zwierząt i ptaków. Zwierzęta zarażają się przez kałowo-ustny mechanizm transmisji, u ptaków jest możliwa transowarialna transmisja *Salmonelli*. Człowiek reaguje na te infekcje i może zarazić się przez jedzenie żywności, która została zanieczyszczona salmonellami (pierwotnie lub wtórnie), a także po kontakcie ze zwierzętami (chorymi lub nosicielami). W sprzyjających warunkach mikroorganizmy z rodzaju *Salmonella* mogą żyć poza żywym organizmem kilka miesięcy, a w niektórych produktach nawet rozmnażać się (1-3). Przemysłowa branża drobiarska i przemysł spożywczy przyczyniły się do wzrostu salmonelozy na całym świecie w ostatnich latach, o czym świadczy rejestracja wybuchu salmonelozy w różnych krajach (4-6).

W Stanach Zjednoczonych od 2013 do 2018 r. średnioroczny wskaźnik zachorowalności na salmonelozę wyniósł 15,4 na 100 tys. ludności (7).

W 2017 roku w krajach Unii Europejskiej zarejestrowano 19,7 przypadków salmonelozy na 100 tys. ludności. Jednak poziom zachorowalności na salmonelozę różni się w różnych krajach tego regionu. W ciągu lat 2013 – 2017 sprawozdania statystyczne krajów Unii Europejskiej nie wykazały, że częstość występowania salmonelozy ulega zwiększeniu lub zmniejszeniu. Siedem krajów wspólnoty informowało o tendencji do wzrostu przypadków tej choroby, podczas gdy cztery kraje o tendencji do zmniejszenia salmonelozy za okres 2013 – 2017 (8).

Leczenie przypadków salmonelozy jest znacznie kosztowniejsze, jednak mniej skuteczne i bardziej niebezpieczne dla pacjenta niż profilaktyka zakażeń. Z tego powodu dużą uwagę skupia się na organizacji działań rozpoznawania i leczenia chorych na salmonelozę, a także na optymalizację wydatków w ramach takich działań (9). Jedną z najskuteczniejszych metod rozwiązywania takich problemów jest zbudowanie modelu matematycznego opisującego procesy rozprzestrzeniania się infekcji w populacji, rozwoju choroby i prognozowania przebiegu epidemii (10). Obecny rozwój technologii informatycznych pozwala na osiągnięcie wysokiej precyzji zbudowanych modeli (11). Pomoże to epidemiologom w podejmowaniu skutecznych decyzji o zmniejszeniu zapadalności (12).

OBJECTIVE

Thus, the objective of the study is to calculate the predicted incidence of salmonellosis in Ukraine based on statistical data.

MATERIAL AND METHODS

To build the model, the data on the new cases of salmonellosis in the Kharkiv oblast (Eastern Ukraine) from December 2015 to December 2018, provided by the Kharkiv Oblast Laboratory Center of the Ministry of Health of Ukraine, were used. Time series of new cases for the population in Kharkiv Oblast has been considered in this work. The daily number of new cases was grouped into the monthly incidence, which was included in the model. Each row is an ordered set of monthly values from December 2015 to December 2018 (data sampling is presented in Fig. 1).

CEL

Celem badań jest opracowanie modelu i obliczenia prognozowanej zachorowalności na salmonelozę w Ukrainie, na podstawie danych statystycznych.

MATERIAŁY I METODY

Do budowy modelu wykorzystano dane o zachorowalności na salmonelozę w obwodzie Charkowskim (Wschodnia Ukraina) z grudnia 2015 do grudnia 2018, dostarczone przez Charkowskie Wojewódzkie Laboratorium Centrum Ministerstwa Zdrowia Ukrainy. W artykule rozważono szeregi czasowe przypadków dla populacji regionu Charkowa. Dzienna liczba przypadków została pogrupowana w miesięczną zachorowalność, która została uwzględniona w modelu. Każdy szereg stanowi uporządkowany zbiór miesięcznych wartości od grudnia 2015 do grudnia 2018 roku (próbka danych jest przedstawiona na Ryc. 1).

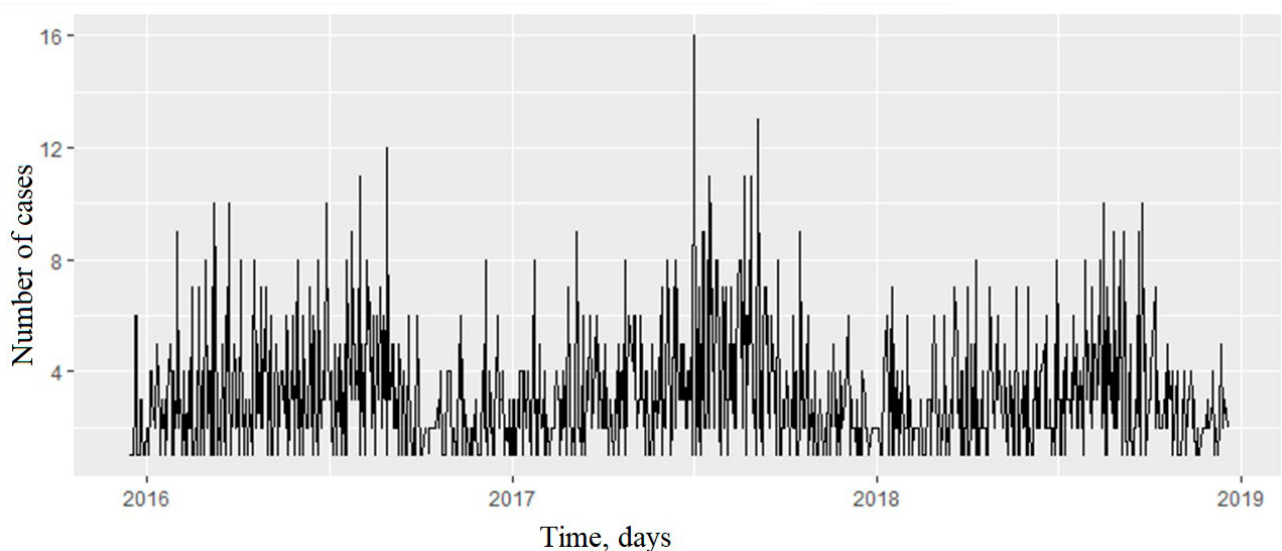


Fig. 1. The incidence of salmonellosis in the Kharkiv region from 2015 to 2018

Ryc. 1. Zachorowalność na salmonelozę w obwodzie Charkowskim od 2015 do 2018 roku

The analysis of statistical methods (13-15), conducted in the study, showed that the most effective is the use of the autoregressive moving average model (ARIMA) [16].

ARIMA stands for autoregressive integrated moving average and is determined by three order parameters: (p, d, q). The fitting process for the ARIMA model is sometimes called the Box-Jenkins method [17].

The autoregressive component (AR (p)) refers to the use of past values in the regression equation for the Y series. The autoregressive component p indicates the number of lags used in the model.

Analiza metod statystycznych (13-15), przeprowadzonych w badaniu, pokazała, że najbardziej skuteczne jest zastosowanie modelu autoregresyjnej średniej kroczącej (ARIMA) (16).

ARIMA oznacza autoregresyjną zintegrowaną średnią krocząca i zależy od trzech parametrów: (p, d, q). Proces dopasowania modelu ARIMA czasami nazywany jest metodą Boxa-Jenkins (17).

Autoregresyjny składnik (AR (p)) odnosi się do korzystania z poprzednich wartości w równaniu regresji dla serii Y. Autoregresyjny parametr p oznacza rząd autoregresji, wykorzystywanych w modelu.

For example, AR (2) or, equivalently, ARIMA (2,0,0), is represented as

$$Y_t = c + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \dots + e_t \quad (1)$$

where φ_1, φ_2 are the parameters for the model.

The value of d represents the degree of difference in the integrated (I (d)) components. Differentiation of a series includes a simple subtraction of its current and previous value d times. Differentiation is often used to stabilize a series when the assumption of stationarity is not fulfilled.

The moving average component (MA (q)) represents the model error as a combination of previous error terms, etc. The order q determines the number of terms for inclusion in the model by the formula:

$$Y = c + \theta_1 e_{(t-1)} + \theta_2 e_{(t-2)} + \dots + \theta_q e_{(t-q)} + e_t$$

It should be noted that the model provides for non-seasonal series, which means that it may be necessary to remove the seasonality of the series before modeling. ARIMA models can also be specified through a seasonal structure. In this case, the model is defined by two sets of order parameters: (p, d, q), as described above, and parameters describing the seasonal component of m periods. The ARIMA methodology has its limitations. These models are directly dependent on past data and therefore work better on long and stable series. ARIMA approximates historical patterns and therefore does not aim to explain the structure of the underlying data mechanism.

The analysis found that the number of cases of salmonellosis in the summer months exceeded the number of cases in the winter, therefore, to increase the accuracy of the forecast, seasonal decomposition of the series was carried out by independent exponential smoothing with the additional parameter δ (delta).

For software implementation and visualization, the programming language R is used in the RStudio software environment.

RESULTS

In the visual sense, it can be assumed that there are a lot of estimated outliers in the sample that can affect the model, distorting statistical reports.

Following the necessary methods for removing outliers of time series, the following steps are developed: identification and replacement of outliers, the use of smoothing and decomposition of series.

Even after outlier removal, daily data is still quite volatile. Visually, it is possible to draw a line through the row, tracing its large troughs and peaks, smoothing

Na przykład AR (2) jest równoważne z zapisem ARIMA (2,0,0) i przedstawia się jak:

$$Y_t = c + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \dots + e_t \quad (1)$$

gdzie φ_1, φ_2 – parametry modelu.

Wartość d przedstawia stopień różnicy w zintegrowanym (I (d)) składnikach. Różnicowanie szeregu zawiera proste odejmowanie jego bieżącej i poprzedniej wartości d-razy. Różnicowanie jest często używane do ustabilizowania szeregu, gdy założenie stacjonarności nie jest spełnione.

Składnik średniej ruchomej (MA (q)) przedstawia błąd modelu jako kombinację poprzednich terminów błędów itd. Kolejność q określa liczbę terminów, które mają zostać uwzględnione w modelu za pomocą wzoru:

$$Y = c + \theta_1 e_{(t-1)} + \theta_2 e_{(t-2)} + \dots + \theta_q e_{(t-q)} + e_t$$

Należy zwrócić uwagę, że poniższy model przewiduje serie niesezonowe, co oznacza, że może być konieczne usunięcie sezonowości serii przed modelowaniem. Modele ARIMA można również określić poprzez strukturę sezonową. W tym przypadku model jest określony dwoma zestawami parametrów zlecenia: (p, d, q), jak opisano powyżej, i parametrami opisującymi poprawną składową m okresów. Metodologia ARIMA ma swoje ograniczenia. Modele te są bezpośrednio zależne od poprzednich danych i dlatego lepiej działają na długich i stabilnych odcinkach. ARIMA przybliża wzorce historyczne i dlatego nie ma na celu wyjaśnienia struktury podstawowego mechanizmu danych.

Dokonując analizy założono, że liczba przypadków salmonelozy w miesiącach letnich przekroczyła liczbę przypadków w zimie, dlatego w celu zwiększenia dokładności prognozy był przeprowadzony sezonowy rozkład szeregu poprzez niezależne wygładzanie wykładnicze z dodatkowym parametrem δ (delta).

Do programowej realizacji i wizualizacji używano języka programowania R w środowisku RStudio.

WYNIKI

W sensie wizualnym można założyć, że w próbie występuje wiele oszacowanych wartości odstających, które mogą wpływać na model, zniekształcając raporty statystyczne.

Postępując zgodnie z metodami usuwania wartości odstających szeregów czasowych opracowano następujące kroki: identyfikacja i wymiana wartości odstających, wykorzystanie wygładzania i rozkładania szeregów.

out noisy vibrations. This line can be described by one of the selected methods – as a moving average, with which points are averaged over several periods of time, thereby smoothing and observing the data in a more stable predicted series. The smoothed time series is shown in Figure 2.

Nawet po usunięciu wartości odstających codzienne, dane nadal są dość zmienne. Wizualnie jest to możliwe: należy narysować linię przez szereg, śledząc jego doły i szczyty, wygładzając drgania. Linia ta może być opisana jedną z wybranych metod - jako średnia ruchoma, za pomocą której są uśredniane punkty kilku okresów czasu, tym samym zacierając i obserwując dane w bardziej stabilny przewidywany szereg. Wygładzony szereg jest przedstawiony na rysunku 2.

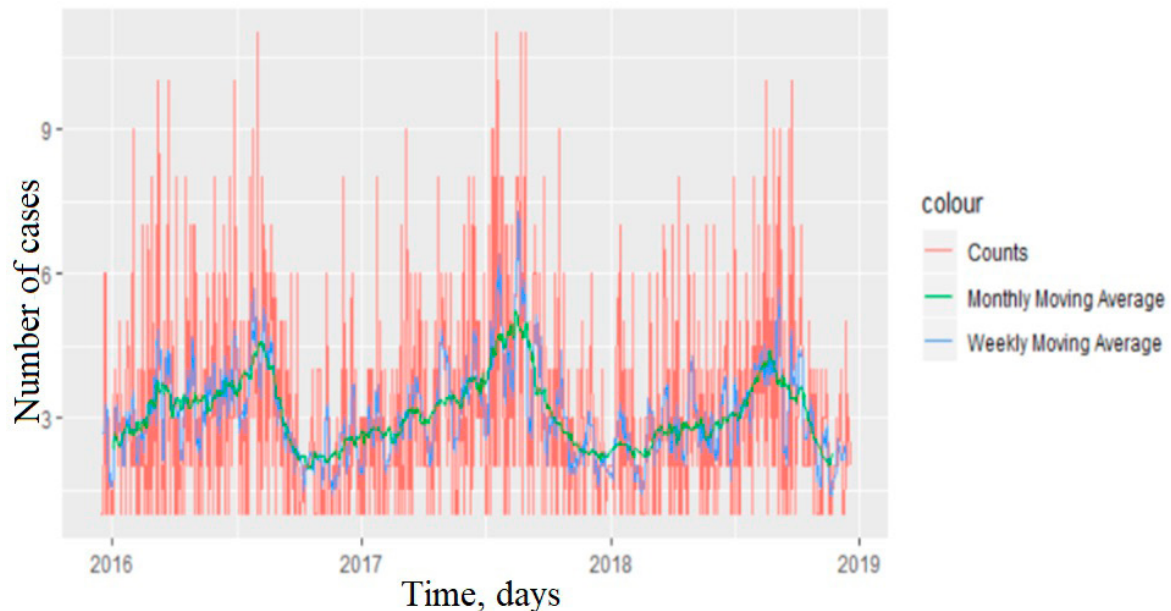


Fig. 2. A smoothed series of incidence of salmonellosis

Ryc. 2. Wygładzone szeregi zachorowalności na salmonellozy

The additive model is usually more suitable when the seasonal or trend components are not proportional to the level of the series, since we can simply superimpose the components on top of each other to restore the series. On the other hand, if the seasonality component changes depending on the level or trend of the series, simply superimposing the components will not be enough to restore the series. In this case, the multiplicative model may be more appropriate.

To use the ARIMA model, the row must be stationary.

A series is called stationary when its average value, variance, and autocovariance are independent of time. This assumption has an intuitive meaning: since ARIMA uses the previous series logs to model its behavior, modeling stable series is less uncertain.

If the series is non-stationary, then such series can be corrected using a transformation, such as differentiation. Converting a series can help in removing its trend or cycles. The idea underlying differentiation is that if the original data series does not

Model addytywny jest zwykle bardziej odpowiedni, gdy komponenty sezonowości lub trendu nie są proporcjonalne do poziomu szeregu, ponieważ możemy po prostu nałożyć komponenty, aby zrekonstruować szereg. Z drugiej strony, jeśli składnik sezonowości zmienia się w zależności od poziomu lub trendu szeregu, zwykłe nałożenie się składników nie wystarczy do odtworzenia szeregu. W takim przypadku bardziej odpowiedni może być model multiplikatywny.

Do korzystania z modelu ARIMA, szereg musi być stacjonarny.

Szereg nazywamy stacjonarnym, gdy jego wartość średnia, wariancja i autokowariancja nie zależą od czasu. Założenie to ma sens intuicyjny: ponieważ ARIMA używa poprzednich rzędów opóźnień do modelowania swoich zachowań, modelowanie stabilnych rzędów jest niestacjonarnością w mniejszym stopniu.

Jeśli szereg jest niestacjonarny, to może być skorygowany za pomocą konwersji, takiej jak różnicowanie. Konwersja szeregu może pomóc w usunięciu jego trendu lub cyklu. Idea różnicowania polega na tym, że jeśli oryginalny szereg danych nie ma stałych właści-

have constant properties over time, then a transition from one period to another can occur.

Autocorrelation graphs are a visual tool for determining if a series is stationary. These graphs can also help you select ARIMA model parameters. Partial autocorrelation graphs reflect the correlation between a variable and its lags.

The developed model allows you to explicitly indicate the order of the model using the *arima()* function or automatically generate a set of optimal values (p , d , q) using the *auto.arima()* function. This function scans combinations of model parameters and selects a set that optimizes the model's matching criteria. There are a number of such criteria for comparing the quality of compliance for several models, in our case, the Akaike information criteria [18] and the Bayesian information criteria (19). These criteria are closely related and can be interpreted as an estimate of how much information will be lost if this model is selected.

Testing the model showed that for 150 days ahead the forecast is characterized by sufficient accuracy. However, for the timely diagnosis of the beginning deterioration of the epidemic situation, it is advisable to predict for a shorter period. Forecasting 50 days ahead showed even higher accuracy and, in our opinion, is optimal. Therefore, in our opinion, the forecast for 50 days ahead is optimal. Such a forecast allows the redistribution of available resources and timely preventive measures. Also, the absence of the need for forecasting for more than 50 days is due to the fact that as a result of the forecast, preventive measures are taken that change the nature of the dynamics of the epidemic process of salmonellosis. The calculated forecast is shown in Figure 3.

wości w czasie, może się zdarzyć przejście od jednego okresu do drugiego.

Wykresy autokorelacji są wizualnym narzędziem do określenia tego, czy szereg jest stacjonarny. Te grafiki mogą również pomóc wybrać parametry modelu ARIMA. Częściowe autokorelacyjne wykresy odzwierciedlają korelację między zmienną a jej opóźnieniami.

Opracowany model pozwala określić kolejność modelu za pomocą funkcji *arima()* lub automatycznie wygenerować zestaw optymalnych wartości (p , d , q) za pomocą funkcji *auto.arima()*. Ta funkcja monitoruje kombinacje parametrów modelu i wybiera zestaw, który optymalizuje kryteria zgodności modelu. Istnieje wiele takich kryteriów dla porównania jakości zgodności z kilku modeli, w naszym przypadku – informacyjne kryteria Akaike (18) i informacyjne kryteria Bajsa (19). Kryteria te są ściśle związane i mogą być interpretowane jako ocena tego, ile informacji będzie straconych, jeśli zostanie wybrany ten model.

Testowanie modelu wykazało, że prognozy na 150 dni naprzód charakteryzują się wystarczającą dokładnością. Jednak w celu szybkiego rozpoznania początkowego pogorszenia się sytuacji epidemicznej zaleca się przewidywanie na krótszy okres, więc naszym zdaniem prognozowanie na 50 dni naprzód jest optymalne. Taka prognoza pozwala przegrupować dostępne zasoby i we właściwym czasie przeprowadzić profilaktyczne zabiegi. Również brak potrzeby prognozowania przez ponad 50 dni wynika z faktu, że w wyniku prognozy podejmowane są działania prewencyjne zmieniające charakter dynamiki epidemicznego procesu salmonellozy. Wyznaczona prognoza jest pokazana na rycinie 3.

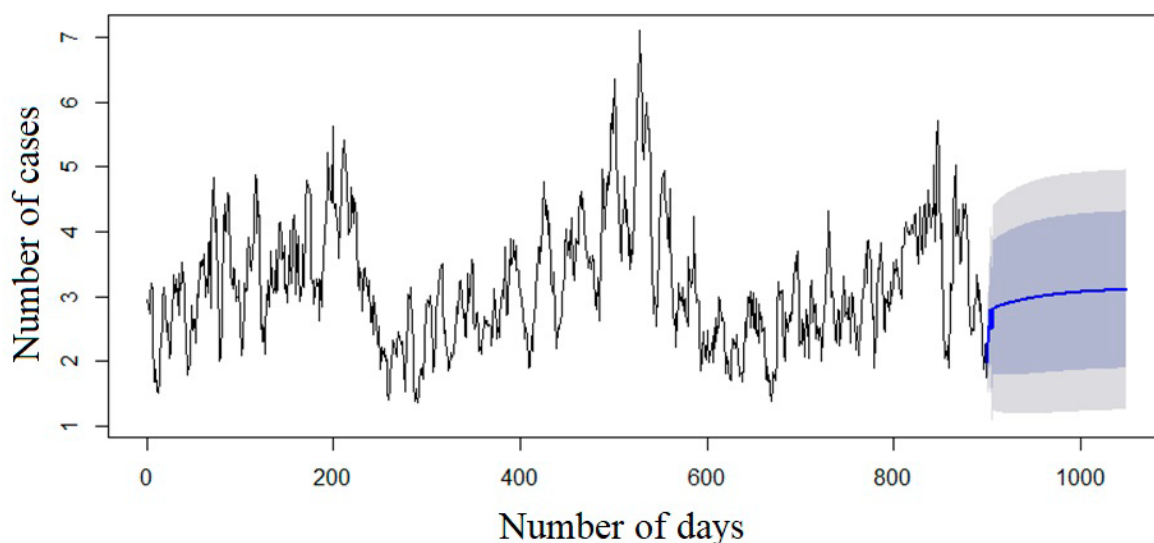


Fig. 3. Predicted incidence of salmonellosis in the Kharkiv region

Ryc. 3. Prognoza występowania zachorowań na salmonellozy w obwodzie Charkowskim

The graphs show the forecast and two intervals: where the predicted value will fall with a probability of 80% (blue color), and with a probability of 95% (light blue color). The blue line above shows the smoothed predicted incidence.

The developed model allows you to enter daily statistics on the case of salmonellosis, therefore, it allows you to build an actual forecast in real time.

In certain cases, the models of exponential smoothing are able to give forecasts that are not inferior in accuracy to the forecasts obtained using more complex models, while once again confirming the fact that the most complex model is far from always the best. The constructed model provides for a series without seasonality, and also differentiates output non-stationary data. In other words, the predictions made are based on the assumption that there will be no other seasonal fluctuations in the data, and the change in patients from one day to another is constant in terms of average value and variance.

CONCLUSIONS

Medical forecasting is an urgent task. The probability of the development and outcome of the disease, based on knowledge of the laws of epidemic processes and the course of the disease. Statistical data and their analysis make it possible to partially substantiate a hypothetical result, as well as provide for the frequency and nature of complications for a short period.

In this study, we analyzed methods and models for processing and analyzing statistical information of the epidemic process of the incidence of salmonellosis in the Kharkiv region. As the main methods for classifying and predicting data, various methods for predicting statistical analysis were considered. In particular, it was proposed to use ARIMA models and finite-difference differentiation, since they are the best suited for solving classification problems with linearly-divided classes. These methods make it possible to calculate the predicted values of the condition of patients according to the statistical results of clinical trials, if the dependent variable is measured in a categorical scale.

The model allows with daily case reporting to automatically build a short-term prognosis, predicting incidence for 50 days ahead and further, which makes it possible to rationally use the available financial, material and labor resources, timely conduct adequate preventive measures, which will help reduce the incidence of salmonellosis.

The study was funded by the Ministry of Health of Ukraine for the state budget in the framework of the research work on the theme "To develop a scientifically

Na wykresach przedstawiono prognozę i dwa przedziały: gdzie przewidywana wartość trafi z prawdopodobieństwem 80% (kolor niebieski) i z prawdopodobieństwem 95% (jasno-niebieski kolor). Niebieska linia powyżej pokazuje wygładzoną prognostyczną częstość występowania.

Opracowany model pozwala na wprowadzanie codziennych statystyk dotyczących przypadku salmonellozy, a zatem pozwala budować rzeczywistą prognozę w czasie rzeczywistym.

Modele wygładzania wykładniczego w niektórych przypadkach są w stanie dawać prognozy, które nie ustępują dokładności prognozom uzyskanym podczas korzystania z bardziej zaawansowanych modeli, a tym po raz kolejny potwierdza fakt, że najtrudniejszy model nie zawsze jest najlepszy. Zbudowany model jest prosty w wykonaniu i umożliwia uzyskanie dokładnego wyniku.

WNIOSKI

Prognozowanie medyczne jest pilnym zadaniem. Prawdopodobieństwo rozwoju i wyniku choroby opiera się na znajomości praw procesów epidemicznych i przebiegu choroby. Dane statystyczne i ich analiza pozwalają częściowo uzasadnić hipotetyczny wynik, a także określić częstotliwość i charakter powikłań w krótkim okresie.

W tym badaniu przeanalizowaliśmy metody i modele przetwarzania i analizy informacji statystycznych na temat epidemicznego procesu występowania salmonellozy w regionie Charkowa. Jako główne metody klasyfikacji i prognozowania danych uwzględniono różne metody prognozowania analizy statystycznej. W szczególności zaproponowano użycie modeli ARIMA i różnicowania różnic skończonych, ponieważ najlepiej nadają się one do rozwiązywania problemów klasyfikacyjnych z klasami rozdzielonymi liniowo. Metody te pozwalają obliczyć przewidywane wartości stanu pacjenta zgodnie ze statystycznymi wynikami badań klinicznych, jeśli zmienna zależna jest mierzona w skali kategoryjnej.

Model umożliwia przy codziennej rejestracji przypadków automatycznie budować krótkoterminowe prognozy, przewidując występowanie zachorowalności na 50 okresów w przyszłości i dalej, co umożliwia racjonalne wykorzystanie dostępnych zasobów finansowych, materiałowych i siły roboczej, terminowo przeprowadzać odpowiednie środki zapobiegawcze, które pomogą zmniejszyć częstość występowania salmonellozy.

Badanie zostało sfinansowane przez Ministerstwo Zdrowia Ukrainy w budżecie państwa w ramach prac badawczych na temat „Opracować naukowo oparty system nadzoru epidemiologiczno-epizootycznego

substantiated system of epidemiological and epizootic surveillance of infections that are shared between animals and people on the example of salmonellosis in the context of the strategy of “One Health”

nad zakażeniami powszechnymi u ludzi i zwierząt, wykorzystując salmonelozę jako przykład, w kontekście strategii „Wspólne zdrowie”.

REFERENCES

1. Michael P. Spector Resistance and survival strategies of *Salmonella enterica* to environmental stresses / Michael P. Spector, William J. Kenyon. // Food Research International: Volume 45, Issue 2, March 2012, Pages 455 – 481
2. Murray C.J. *Salmonellae* in the environment / Murray C.J. Australian *Salmonella* Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide // Rev Sci Tech. 1991 Sep;10(3):765-85
3. Mollie D. Role of Nonhost Environments in the Lifestyles of *Salmonella* and *Escherichia coli* / Mollie D. Winfield and Eduardo A. Groisman. American society for microbiology. Applied and Environmental Microbiology // Journal List - Appl Environ Microbiol. 2003 Jul; 69(7): 3687–3694. [Electronic source]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC165204/>
4. Debbie Lee. Precipitation and Salmonellosis Incidence in Georgia, USA: Interactions between Extreme Rainfall Events and Antecedent Rainfall Conditions / Debbie Lee, Howard H. Chang, Stefanie Ebelt Sarnat, and Karen Levy. Environmental Health Perspectives // Journal List - Environ Health Perspect. 2019 Sep; 127(9): 097005. [Electronic source]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6792369/>
5. Laura Ford. Seven *Salmonella* Typhimurium Outbreak in Australia Linked by Trace-Back and Whole Genome Sequencing / Laura Ford, Qinning Wang, Russell Stafford and other // Foodborne pathogens and disease – Volume 15, number 5, 2018;285-286.
6. European Food Safety Authority. Multi-country outbreak of *Salmonella* Enteritidis infections linked to Polish eggs / European Food Safety Authority European Centre for Disease Prevention and Control: doi: 10.2903/sp.efsa.2017.EN-1353 Technical report // Stockholm - 2017 December 2017
7. Foodborne Diseases Active Surveillance Network (FoodNet). CDC/ Central for Disease Control and Prevention – CDC 24/7: Saving Lives, Protecting People // [Electronic source]: <https://wwwn.cdc.gov/foodnetfast/>
8. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017 / European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC). Scientific report: doi: 10.2903/j.efsa.2018.5500 // EFSA Journal 2018;16(12):5500. - 19 November 2018 - P. 22-23 [Electronic source]: <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2018.5500>
9. Meniailov, I., Bazilevych, K., Fedulov, K., et. al. Using the K-means method for diagnosing cancer stage using the Pandas library. CEUR Workshop Proceedings, vol. 2386, 2019, pp. 107-116.
10. Chumachenko, D., Chumachenko, T.. Intelligent Agent-Based Simulation of HIV Epidemic Process. Advances in Intelligent Systems and Computing, vol. 1020, 2020, pp. 175-188.
11. Chumachenko, D., Chumachenko, K., Yakovlev, S. Intelligent Simulation of Network Worm Propagation using the Code Red as an Example. Telecommunications and Radio Engineering, Vol. 78, Iss. 5, 2019, pp. 443-464.
12. Polyvianna, Yu., Chumachenko, D., Chumachenko T. Computer Aided System of Time Series Analysis Methods for Forecasting the Epidemics Outbreaks, 2019 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), 2019, pp. 7.1-7.4.
13. Chumachenko, D., Balitskii, V., Chumachenko, T., et. al. Intelligent expert system of knowledge examination of medical staff regarding infections associated with the provision of medical care. CEUR Workshop Proceedings, vol. 2386, 2019, pp. 321-330.
14. Chumachenko, D. On Intelligent Multiagent Approach to Viral Hepatitis B Epidemic Processes Simulation. Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018, 2018, pp. 415-419.
15. Chumachenko, D., Meniailov, I., Bazilevych, K., et. al. Development of an intelligent agent-based model of the epidemic process of syphilis. 2019 IEEE 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 2019, pp. 42-45.
16. Chumachenko, D., Meniailov, I., Bazilevych, K., et. al. On intelligent decision making in multiagent systems in conditions of Uncertainty. 2019 XIth International Scientific and Practical Conference on Electronics and Information Technologies (ELIT), Lviv, 2019, pp. 150-153.
17. Vaishnav, V., Vajpai, J. Seasonal Time Series Forecasting by Group Method of Data Handling, 2018 IEEE International Students' Conference

- on Electrical, Electronics and Computer Science (SCEECS), Bhopal, 2018, pp. 1-5.
18. Bai, X., Kong, F., Zhao, H. Classification and Forecasting for Enterprise Data, 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 3742-3745.
19. Dridi, N., Hadzagic, M. Akaike and Bayesian Information Criteria for Hidden Markov Models, in IEEE Signal Processing Letters, vol. 26, no. 2, 2019, pp. 302-306.

Received: 24.12.2019

Accepted for publication: 6.08.2020

Otrzymano: 24.12.2019 r.

Zaakceptowano do publikacji: 6.08.2020 r.

Address for correspondence:

Adres do korespondencji:

Dmytro Chumachenko

National Aerospace University "Kharkiv Aviation Institute"

Chkalov str., 17, Kharkiv, Ukraine

+380937196225

dichumachenko@gmail.com