ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, СИСТЕМНИЙ АНАЛІЗ ТА КЕРУВАННЯ

UDC 519.2, 519.8
DOI: 10.20535/1810-0546.2017.5.108577

I. Nesteruk*
Institute of Hydromechanics, NASU, Kyiv, Ukraine
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

STATISTICS BASED MODELS FOR THE DYNAMICS OF CHERNIVTSI CHILDREN DISEASE**

Background. Simple mathematical models of contamination and SIR-model of spreading an infection were used to simulate the time dynamics of the unknown before children disease, which occurred in Chernivtsi (Ukraine). The cause of many cases of alopecia, which began in this city in August 1988 is still not fully clarified. According to the official report of the governmental commission, the last new cases occurred in the middle of November 1988, and the reason of the illness was reported as chemical exogenous intoxication. Later this illness became the name "Chernivtsi chemical disease". Nevertheless, the significantly increased number of new cases of the local alopecia was registered almost three years and is still not clarified.
Objective. The comparison of two different versions of the disease: chemical exogenous intoxication and infection. Identification of the parameters of mathematical models and prediction of the disease development.
Methods. Analytical solutions of the contamination models and SIR-model for an epidemic are obtained. The optimal values of parameters with the use of linear regression were found.
Results. The optimal values of the models parameters with the use of statistical approach were identified. The calculations showed that the infectious version of the disease is more reliable in comparison with the popular contamination one. The possible date of the epidemic beginning was estimated.
Conclusions. The optimal parameters of SIR-model allow calculating the realistic number of victims and other characteristics of possible epidemic. They also show that increased number of cases of local alopecia could be a part of the same epidemic as "Chernivtsi chemical disease".

Keywords: model identification; parameter identification; statistical methods; mathematical modeling of infection diseases; SIR-model; contamination models.

Introduction

All mathematical models – but particularly models for life-science applications – raise the question whether the selection of regarded effects is suitable for the description of the observations under consideration. It is especially difficult to select the proper model for an unknown phenomenon as in the case of mysterious children disease, which occurred in Chernivtsi (Ukraine). The cause of many cases of alopecia, which began in this city in August 1988 and lasted almost three years, is still not fully clarified.

In such cases, different versions or mathematical models must be taken into account. The corresponding unknown parameters (which are used in any model) have to be identified with the use of real data, which are always available only with restrictions. First, there is the quantitative restriction that new data is not available or at least available for the price of an immense effort only. Second, there is a qualitative restriction: life-science data oftentimes contains large measurement errors. Third, there is a causal restriction: the data alone does not allow distinguishing between correlation and causal relations.

Here, we consider the development of the unknown before children disease, which occurred in Chernivtsi (Ukraine) [1–4]. In particular, children in the age between 1 and 15 years suffered total hair loss (alopecia); mucosal lesions; hematological, neuro-psychic and cutaneotrophic disorders.

In October–November 1988, the governmental commission investigated the disease. In that time many scientists from the Chernivtsi State University and other research institutions were trying to find the reason the disease.

According to the official report of the governmental commission, the last new cases occurred in the middle of November 1988 [2], and the reason of the illness was reported as chemical exogenous intoxication. Later this illness became the name "Chernivtsi chemical disease". The registered number of victims was 162 (this value was later corrected due to difficulties of diagnostics of the unknown disease). Nevertheless, the chemical agent/agents, which cause the disease, are not found till now, in spite of some reports of possible thallium intoxication [2]. In that time there were also investigators who supported the infection version of the disease (e.g., Prof. Dekhan-Khodzhaeva from Tashkent, who re-

* corresponding author: inesteruk@yahoo.com
** The author gratefully acknowledges the support of the project AMMODIT funded within EU H2020-MSCA-RICE

ported the presence of a fungus in the blood of ill children; Dr. I. Nesteruk, who in that time worked at Chernivtsi State University and investigated the statistics of the disease).

In this paper, we will try to analyze ones more the statistical data of the Chernivtsi disease with the use of different mathematical models, to identify the optimal values of their parameters with the use of statistical methods and to get closer to the real causes of the phenomenon. In particular, we will try to answer the question which version (chemical exogenous intoxication or infection) is more reliable from the mathematical point of view.

Problem formulation

The comparison of different models of the Chernivtsi children disease dynamics with the use of analytical solutions and the linear regression for the identification of the model parameters.

Is the popular contamination version realistic?

We shall analyze the data, which origins from governmental commission. Numbers of concerned children are given in Table 1 and present the cases with total alopecia.

***Table* 1.** Number of children concerned by Chernivtsy children disease ("Chernivtsy chemical disease")

Date	t_j	A_j
05. Aug. 88	1	2
10. Aug. 88	2	2
15. Aug. 88	3	2
20. Aug. 88	4	2
25. Aug. 88	5	4
30. Aug. 88	6	7
04. Sep. 88	7	9
09. Sep. 88	8	10
14. Sep. 88	9	12
19. Sep. 88	10	12
24. Sep. 88	11	17
29. Sep. 88	12	19
04. Oct. 88	13	31
09. Oct. 88	14	40
14. Oct. 88	15	45
19. Oct. 88	16	56
24. Oct. 88	17	61
29. Oct. 88	18	73

The statistics about the further development of the disease is not very reliable for our analysis, since:

1) In November 1988 it was a panic in the city; the parents evacuated their children in different regions of the USSR (mostly near Chernivtsi); the new cases were not registered in proper way or information about these cases came with a big delay.

2) It was difficult to separate the information about cases with total and partial alopecia. The partial hair loss is a known illness and the official governmental commission didn't recognize any connection between the cases with total and local alopecia.

Table 1 shows that the precise time of the disease beginning t_0 is unknown. Therefore, the optimization procedures have to determine the optimal value of this parameter as well as for other parameters, which will be used in different models.

The exogenous intoxication was and still is the most popular version of the Chernivtsi disease. Let us analyze this conclusion of the official commission with the use of statistical data presented in Table 1.

Since we don't know the location, the beginning time and other characteristics of the contamination, we have to use the simplest robust models. For example, a person becomes ill, after obtaining a certain amount of the poison. It means that the time derivative of the number of victims $V(t)$ must be proportional to the concentration of the toxic substance $G(t)$. In the simplest case of linear dependence the corresponding equation can be written as

$$\dot{V} = \beta G. \tag{1}$$

Usually the poison concentration depends also on the location of the contamination source and can be calculated with the use of complicated partial differential equations, which are different in cases of air, water or soil contamination. Knowing nothing about the nature of possible contamination, let us consider at first the simplest case of constant poison concentration $G = G_0$, which appears at some unknown moment t_0 and exists during some period of time.

Then equation (1) has a trivial linear solution:

$$V(t) = \beta G_0 (t - t_0). \tag{2}$$

In order to check how the registered points fit the straight line (2), let us use the linear regression [5]. For the data presented in Table 1, the equation of the linear regression V on t (the optimal straight line, minimizing the sum of squared distances between registered and theoretical points) looks

$$V(t) = 4.007t - 15.619 \tag{3}$$

and is shown in Fig. 1 (line *3*).

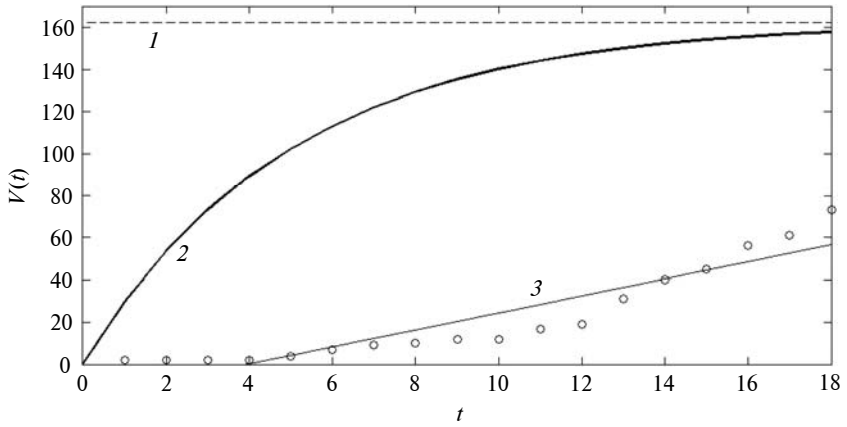

Fig. 1. Models of contamination. The experimental points corresponding to Table 1 are shown by circles. The saturation level N_C = 162 (according to the report of the governmental commission) is shown by the dashed line *1*. Contamination with dissipation (eq. (6)) is shown by solid line *2*. The case of the constant poison concentration $G = G_0$ (eq. (2)) and linear regression for the experimental points are presented by line *3*

The regression coefficient is rather high $r \approx 0.922$. We can use the F-test to check the null hypothesis that says that a proposed linear model fits well the experimental data, presented in Table 1. The experimental value of the Fisher function can be calculated with the use of the formula

$$F = \frac{r^2(n-m)}{(1-r^2)(m-1)} \qquad (4)$$

where $n = 18$ is the number of observations, $m = 2$ is the number of parameters in the regression equation. The corresponding value $F = 90.73$ has to be compared with the critical value $F_c(k_1, k_2)$ of the Fisher function at a desired significance or confidence level. The critical values for $k_1 = m - 1 = 1$, $k_2 = n - m = 16$ are presented in Table 2.

***Table* 2.** Critical values of the Fisher function for different significance levels

Significance level	0.05	0.01	0.001	0.0001	0.00001
$F_c(1,16)$	4.49	8.53	16.12	26.36	40.07

It can be seen that the calculated value $F = 90.73$ is greater than the critical ones at all the significance levels presented in Table 2. It means that experimental points fit well the linear dependence (3). Nevertheless, comparing (2) and (3) and estimating the starting time of the contamination yields the unrealistic value $t_0 \approx 3.9$, since the first cases of the disease were registered approximately 15 days before. Therefore, we have to doubt the used model with the constant poison concentration.

It is possible to have the contamination dynamics similar to the registered one in the case of an increasing concentration of a poison only, which drops to zero after middle of November 1988. We will estimate the necessary increase rate of a poison concentration in order to fit the registered data later.

Usually the concentration of a poison decreases with time. This fact allows developing more realistic models of the contamination dynamics. In particular, it is possible to have a restricted number of victims in comparison with the constant poison concentration model, which yields the unbounded linear increase of the ill persons number (see (2)).

Spots of contamination can move with air of water flows. For example, a radioactive contamination of soil at a fixed place decreases rapidly after a severe rain and is transferred to other places. Let us consider the most unfavorable situation with a slow contamination decrease and neglect these flows. Then the dissipation of the poison is connected with diffusion or natural decreasing of a number of radioactive atoms. In both cases, the time derivative of the poison concentration is proportional to its current value [6]:

$$\dot{G} = -\varsigma G \qquad (5)$$

where ς is the degradation rate.

Let us consider the dynamics of a one-time contamination by the toxic amount G_0 which occurred at the moment of time t_0. Then the set of equations (1) and (5) has the simple analytic solution

$$V(t) = \frac{\beta G_0}{\varsigma}(1 - e^{-\varsigma(t-t_0)}). \qquad (6)$$

The combination of parameters $N_C = \beta G_0 / \varsigma$ yields the asymptotical value of the contaminated persons, since V approaches N_C at great values of time. An example of solution (6) for $N_C = 162$, $\varsigma = 0.2$, $t_0 = 0$ is shown in Fig. 1 by curve *2*. The horizontal asymptote $N_C = 162$ is represented by dashed line *1*.

It doesn't look that the registered points can fit the solution (6) at some reliable values of para-

meters N_C, ς, and t_0. To be sure, let us first consider the case $N_C >> V$. Then the function

$$F_* = \ln(1 - V/N_C) = -\varsigma(t - t_0) \qquad (7)$$

can be approximated as follows:

$$F_* \approx -V/N_C = -\varsigma(t - t_0)$$

and for the number of victims we have the simple linear relationship $V = \varsigma N_C(t - t_0)$ which coincides with (2). We have already concluded that function (2) can fit the registered data only at unrealistic values of the starting time.

According to (7), F_* can be considered as a random variable with the linear distribution. At any fixed value of N_C we can apply the linear regression. E.g., for $N_C = 100$ the contamination starting time t_0 was calculated to be approximately 4.52; $r = -0.896$; $F = 65.14$. With increasing of N_C the value of t_0 diminishes and tends to 3.9 already estimated for the large N_C. Therefore, all the estimations of the starting time are unrealistic.

These facts allow us to doubt the contamination version of the Chernivtsi children disease. It must be noted, that presented analysis concerns the exogenous intoxication when the toxic substance comes to the human body outside and its concentration, degradation, and transfer do not depend on the number of contaminated people. There are also other types of contamination. For example, in the accident occurred in 1987 at Goiania (Brazil) an old radiotherapy source was stolen from a hospital. It was subsequently handled by many people, resulting in four deaths and 249 were found to have significant levels of radioactive material in or on their bodies [7]. The dynamics of such contamination is similar to one of the infectious disease. We will concentrate on the infection version of the Chernivtsi disease in the next Sections.

Infection version. Exact and approximate solutions of SIR-equations

The SIR-model for an infectious disease can be written as follows [8, 9]:

$$\dot{S} = -\alpha SI, \qquad (8)$$

$$\dot{I} = \alpha SI - \rho I, \qquad (9)$$

$$\dot{R} = \rho I. \qquad (10)$$

The number of susceptible persons is S, infected – I, removed – R; the infection and immunisation rates are α and ρ respectively.

Since $\dot{S} + \dot{I} + \dot{R} = 0$ (see, eqs. (8)–(10)), the sum $N = S + I + R$ must be constant for all moments of time and can be treated as the amount of susceptible persons before the outbreak of an epidemic, since $I = R = 0$ at $t < t_0$. It must be noted that the constant N is not the volume of population N_{total} but only the initial number of people sensitive to some specific disease. In particular, the ratio N/N_{total} may be rather small.

To determine the initial conditions for the set of equations (8)–(10), let us suppose an epidemic started at some moment of time t_0, when the first infected person appeared. Then

$$I(t_0) = 1, \quad R(t_0) = 0, \quad S(t_0) = N - 1. \qquad (11)$$

There are situations when an epidemic starts with several or many victims. For example, when many people have eaten an infected food or a bacterial weapon has been used, the initial value I_0 and $S_0 = N - I_0$ must be used in (11).

In the case if Chernivtsi disease the small amount of victims in August 1988 (see Table 1) allows using the initial conditions (11) and removing the unknown parameter I_0. The same initial conditions can be used when a new modification of influenza virus has been developed inside a population.

Very important properties of epidemic can be derived from the set of differential equations (8)–(10) without solving, see [8, 9]. In particular, it follows from (8) and (9) that

$$\frac{dI}{dS} = \frac{\nu}{S} - 1, \quad \nu = \frac{\rho}{\alpha}. \qquad (12)$$

Integration of (12) with the initial conditions (11) yields:

$$I = \nu \ln S - S + N - \nu \ln(N - 1). \qquad (13)$$

Function I has a maximum at $S = \nu$ (it follows from (9)) and tends to zero at infinity, see [8, 9]. In comparison, $S_\infty > 0$ and can be calculated from a non-linear equation

$$S_\infty = (N - 1)e^{\frac{S_\infty - N}{\nu}}. \qquad (14)$$

Formula (14) follows from (8), (10), and (11) and coincides with the relationship from [8, 9] at $S_0 = N - 1$. An approximate solution of (8)–(10) was found by Kermack&McKendrick [8], an exact solution was proposed by Kendall (see [10]).

We solve (8)–(10) by introducing the function $V(t) = I(t) + R(t)$, corresponding to the number of victims, see [11]. It follows from (9), (10) and (13) that:

$$\dot{V} = \alpha SI$$

$$= \alpha(N - V)[\nu \ln(N - V) + V - \nu \ln(N - 1)]. \quad (15)$$

Integration of (15) yields:

$$t = \frac{F_1(V, N, \nu) + \alpha t_0}{\alpha}, \quad (16)$$

$$F_1 = \int_1^V \frac{dU}{(N - U)[\nu(N - U) + U - \nu \ln(N - 1)]}. \quad (17)$$

Thus, for every set of parameters N, ν, α, t_0, and a fixed value of V, the integral (17) can be calculated and the corresponding moment of time can be determined from (16). Then I can be calculated from (13) by putting $S = N - V$, and function R from $R = V - I$.

If $N >> V \geq 1$, the obtained solution of the set of differential equations (8)–(10) can be simplified with the use of two different approximations for the function $\ln(N - U) - \ln(N - 1)$. If we assume

$$\ln(N - U) - \ln(N - 1) \approx 0,$$

then

$$F_1 = \ln V/N$$

and

$$V = e^{\gamma(t - t_0)}, \quad \gamma = \alpha N. \quad (18)$$

According to (18) epidemics start exponentially and only two parameters ($\gamma = \alpha N$ and t_0) describe the process.

To follow next stages, let us use the more exact approximation:

$$\ln(N - U) - \ln(N - 1) \approx (1 - U)/N.$$

Then integral (17) can be expressed as follows:

$$F_1 = \frac{N}{\nu + (N - \nu)N}$$

$$\times \left[\ln \frac{NV + \nu(1 - V)}{N} - \ln \frac{N - V}{N - 1} \right],$$

and the solution has the form:

$$V = \frac{EN - \nu}{E + N - \nu},$$

$$E = \frac{N}{N - 1} \quad (19)$$

$$\times \exp \frac{\alpha(t - t_0)[\nu + (N - \nu)N]}{N}.$$

The second approximation (19) yields limited value of victims, since $V = I + S$ tends to N at infinity.

The difference between the exact solution of (16), (17) (solid line *3*), first approximation (18) (dotted line *1*), and second approximation (19) (dashed line *2*) can be seen in Fig. 2. In this example the values of parameters $N = 162$, $\nu = 1$, $\alpha = 0.001443$ and $t_0 = -1.13$ were taken for calculations. The Fig. 2 illustrates that the registered points from Table 1 can fit rather good both approximate and exact solutions of the SIR epidemic model and can give reliable estimations of the starting time t_0, in comparison with the contamination model presented in the previous Section and in Fig. 1. In order to decide is an unknown disease infectious or not, it is enough to have experimental data for the initial stage of disease (as in Chernivtsi case) and to use the simplest exponential approximation (18).

Unfortunately, any model of exponential grows yield unreliable long time prediction, since any epidemic stops with a limited number of victims (as shown in Fig. 2 by lines *2* and *3*). The accuracy of the second approximation can be also insufficient for the long time predictions (especially at large values of ν). To illustrate this fact let us calculate $W = (N - S_\infty)/N$ from nonlinear equation (14).

The second approximation (19) yields $W = 1$ (i.e., the epidemic stops when all the susceptible persons are infected), but the solutions of (14) shown in Fig. 3 for different values of parameter $\nu = 100$; 300; 500; 700 (curves *1*–*4* respectively) can strongly differ from $W = 1$. Thus, the range of application of the second approximation is also limited and in many cases, only exact solution or numerical methods can ensure a good parameter identification.

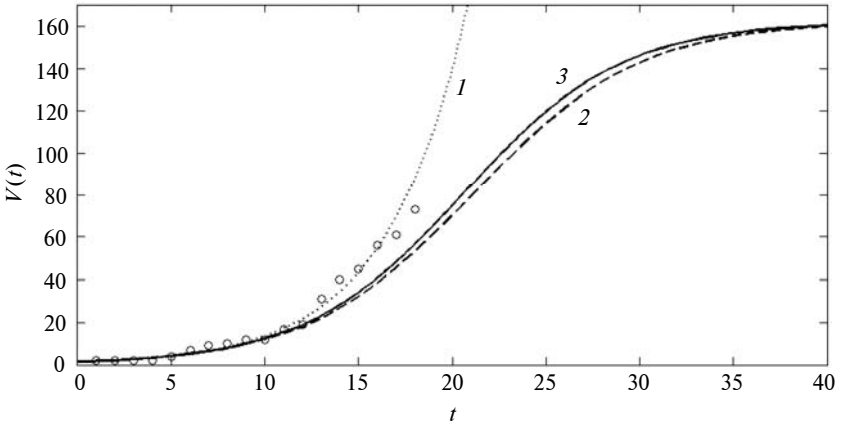

Fig. 2. Infection version. First (eq. (18)) and second (eq. (19)) approximations are shown by dotted line *1* and dashed line *2* respectively. The exact solution (eqs. (16), (17)) is shown by the solid line *3*. The experimental points corresponding to the Table 1 are shown by circles

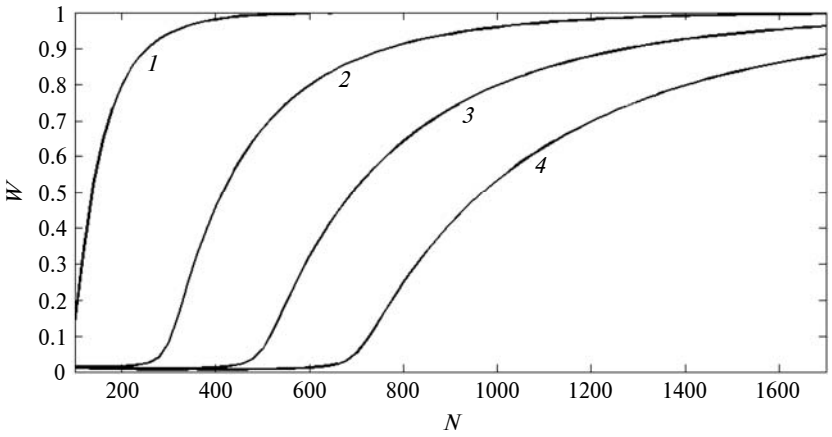

Fig. 3. Solutions of nonlinear equation (14) for the values of parameter ν = 100; 300; 500; 700 (curves *1*–*4* respectively)

Parameter identification in Chernivtsi case

If for the initial stage of an unknown disease, the experimental data about the number of victims A_j is available at different moments of time t_j, it is necessary to determine do they fit an exponential growth (see (18)) or not. For this purpose, different methods can be used.

For example, in 1988 the author applied the linear regression (see, e.g., [5]) to the function

$$F_2 = \ln V = \gamma(t - t_0). \qquad (20)$$

Data presented in Table 1 yields the optimal values $\gamma \approx 0.2337$, $t_0 \approx -1.13$ with the very high correlation coefficient $r = 0.987$. The corresponding line is shown in Fig. 2 by curve *1*. This statistics based method treat the function (20) as a random variable and applies the minimization of the function

$$J_1(\gamma, t_0) = \sum_1^n [\gamma(t_j - t_0) - \ln A_j]^2. \qquad (21)$$

With the use of summed squared error between the model predictions and the measured data, e.g. [11]

$$J_2(\gamma, t_0) = \sum_1^n [e^{\gamma(t_j - t_0)} - A_j]^2, \qquad (22)$$

other optimal values of parameters can be obtained: $\gamma \approx 0.199$ and $t_0 \approx -3.8$.

It can be seen that different minimization methods yield rather different optimal values of parameters. The corresponding lines can differ very slightly in the region, where the registered data are available, but can give very different long time predictions.

In the case of contamination the exponential grows of the victims number is possible only in the case of exponential grows of the poison concentration. Really, equation (1) has the solution

$$V(t) = \exp[\gamma(t - t_0)],$$

coinciding with (18), if

$$G(t) = \frac{\beta}{\gamma} \exp[\gamma(t - t_0)]. \qquad (23)$$

With the use of (23) we can estimate the ratio of poison concentrations at moments of time corresponding to the first and last data registration points (August 5, 1988 and October 29, 1988 accordingly, see Table 1). Then

$$G(t = 1)/G(t = 18) = \exp(-17\gamma).$$

We have already defined the optimal values of the parameter γ with the use of data from Table 1 and functions (21) and (22) (γ = 0.2337 and γ = 0.199 respectively). It means that concentration of the toxic substance had to increase approximately 53 times (or 29 times for the second estimation) during this period of time and then drop to zero after the middle of November 1988.

We can imagine some hypothetical situations, when it could be possible. For example, an explosion at a chemical plant occurred in the beginning of August 1988; a leakage of a poison increased yielding its increasing concentration in the air; after the middle of November 1988 this leakage was stopped and a wind removed a contamination cloud. We could also imagine another case with the increasing poison concentration. For example, somebody added a poison to the drink water very intensively in order to increase its concentration and rapidly stopped to do this after the middle of November 1988. Both scenarios look unrealistic and allow us to concentrate on the infection version.

Since the first and second approximations (see (18) and (19)) have limited accuracy and ability for long time predictions, let us concentrate on the exact solution (16), (17) and use the fact that the random function $F_1(V, N, \nu)$ has a linear distribution (see (16)). Then we can apply the linear regression for every pair of parameters N and ν and calculate the corresponding values of t_0 and α. The optimal (the most reliable) values of N and ν correspond to the maximum value of the function F presented by eq. (4), since in this case the confidence level will be also maximal. According to formula (4), the maximum of F corresponds to the ma-

ximum of the correlation coefficient r, since n and m are fixed.

Calculations show, that the correlation coefficient has a global maximum $r_{max} = 0.999451150131$ at $N = 635$ and $\nu = 358.97$. The corresponding optimal values of other two parameters are $t_0 = 0.7623$ and $\alpha = 0.00077922$. Then we can calculate the optimal value of the parameter $\rho = \alpha\nu = 0.27972$, the saturation level $V_\infty = 458$ (with the use of (14)), the Fisher criterion value $F = 825.8$ (with the use of (4)).

Results and discussion

The exact solution (16), (17) corresponding the found optimal values of parameters are shown in Fig. 4 by the curve. It can be seen that the saturation level – the total number of possible victims – could be much larger than 162 (as reported by the governmental commission and shown in Fig. 4 by circles) and possible epidemic could have much longer duration.

This fact forced to draw our attention to the dynamics of the local alopecia cases (shown in Fig. 4 by triangles). The number of new cases of the local alopecia rapidly increased during first 700 days (120×5) and is much greater than the number of sporadic cases typical for the population of 50 000 children in Chernivtsi. The average monthly rate of 9 new cases of local alopecia was taken for calculations (typical for European cities with the same children population, see [12, 13]). The crosses represent the difference between registered cases of local alopecia and typical for the 50 000 children population. The sum of these unexpected cases of the local alopecia and cases of the total alopecia ("Chernivtsi chemical disease" shown by circles) are represented in Fig. 4 by squares.

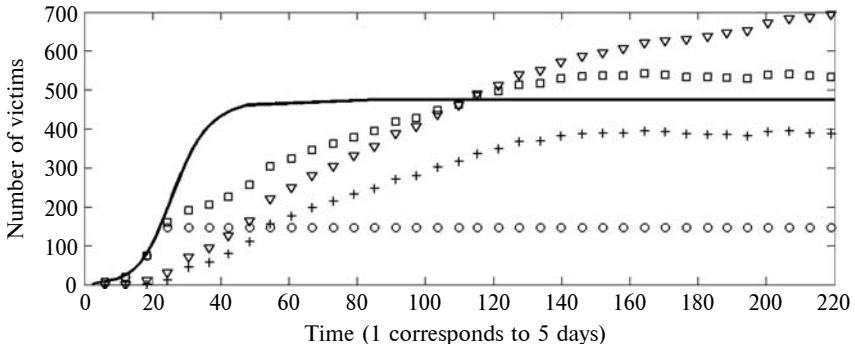

Fig. 4. Results of parameter identification in Chernivtsi case and comparison with the experimental data for the total and local cases of alopecia in Chernivtsi during 1100 days of observation. The curve represents the exact solution (16), (17). Triangles correspond to the new registered local alopecia cases. The crosses represent the difference between registered cases of local alopecia and typical for the 50 000 children population. The circles show the cases of the total alopecia ("Chernivtsi chemical disease"). The sum of the unexpected cases of the local alopecia and cases of the total alopecia are represented squares

It can be seen that this sum of the cases of the total alopecia and the unexpected cases of the local alopecia (squares) has the saturation level, which is close to the calculated value of 458. Deviations from the theoretical curve for the period between 100 and 550 days can be explained by the mass children evacuation from Chernivtsi.

At the beginning of a possible epidemic (when only slight dependence on N occurs and the exponential growth with $\gamma = \alpha N \approx 0.35$ is valid), the number of victims increased twice every $\ln 2/\gamma \approx 1.98$ time intervals, i.e., approximately every 10 days. Therefore a possible infection does not spread very fast. This fact supports the fungus version of disease proposed by Prof. Dekhan-Khodzhaeva.

We can estimate also the possible date of the epidemic beginning – August 3 or 4, 1988, since the corresponding optimal value $t_0 = 0.7623$. The new disease could happen because of a mutation of an infectious agent, which always exist in populations (as a new severe influenza is a result of a mutation of a common virus).

The big value of the Fisher criterion ($F = 825.8$) obtained for the optimal values of the parameters (it is much greater than the critical values shown in Table 2) demonstrates very high confidence level. Nevertheless, this level is also high for the values of parameters located in the vicinity of the optimal point $t_0 = 0.7623$; $\alpha = 0.00077922$, since the maximum of regression coefficient at this point is not sharp. This fact can question the procedure of the parameter identification.

Another weak feature of the method is connected with the fact that the estimation of susceptible children, who are still present in the population

$$S_\infty = N - V_\infty = 635 - 458 = 177,$$

is rather large. It means that these children can catch the infection. Such situation needs more precise analyses with the use of more complicated models (see, e.g., [14]).

Conclusions

Simple mathematical models for the time dynamics of the Chernivtsi children disease showed that the infectious version is more reliable in comparison with the popular contamination one. The optimal parameters of SIR-model allow calculating the realistic number of victims and other characteristics of

possible epidemic. They also show that increased number of cases of local alopecia could be a part of the same epidemic as "Chernivtsi chemical disease".

Probably, the further research should focus on using more complicated mathematical models and on finding the infectious agent, which can cause sporadic cases of the local alopecia as well.

Acknowledgement

The author is grateful to Prof. Dirk Langemann (Technische Universitaet Braunschweig) and Prof. Juergen Prestin (Universitaet zu Luebeck) for their support and for very useful discussions of the results. The study was supported by EU-financed Horizon-2020 project AMMODIT (Grant Number MSCA-RISE 645672).

References

[1] D.D. Zerbino *et al.*, "The results of a study of chemical-induced disease in children in Chernovtsy", *Vrach. Delo*, vol. 8, pp. 88–91, 1991 (in Russian).

[2] D.D. Zerbino and A.M. Serdiuk, *Chernivtsi Chemical Disease, New Ecologic Pathology? (Essays about epidemiology, clinic revealings, etiology, versions of genesis, documents)*. Lviv, Ukraine: Missioner, 1998 (in Ukrainian).

[3] V.K. Patratii *et al.*, "Chemical intoxication syndrome in children with diffuse alopecia", *Pediatriia*, no. 12, pp. 52–55, 1991 (in Russian).

[4] V.K. Tatochenko *et al.*, "An epidemic outbreak of diffuse alopecia in children", *Pediatriia*, no. 12, pp. 67–71, 1990 (in Russian).

[5] N.R. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. New York: John Wiley, 1998.

[6] *Structured-Population Models in Marine, Terrestrial, and Fresh-Water Systems*, S. Tuljapurkar and H. Caswell, eds. Springer, 1997.

[7] *The Radiological Accident in Goiania*. Vienna, Austria: International Atomic Energy Agency, 1988.

[8] W.D. Kermack and A.G. McKendrick, "A contribution to the mathematical theory of epidemics", *J. Royal Statistical Society, Ser. A*, vol. 115, pp. 700–721, 1927.

[9] J.D. Murray, *Mathematical Biology I/II*. New York: Springer, 2002.

[10] N.T.J. Bailey, *The Mathematical Theory of Epidemics*. London, UK: Griffin Book Co., 1957.

[11] D. Langemann *et al.*, "Comparison of mathematical models for the dynamics of the Chernivtsi children disease", *Mathematics in Computers and Simulation*, vol. 123, pp. 68–79, 2016. doi: 10.1016/j.matcom.2016.01.003

[12] W. Pschyrembel, *Klinisches Woerterbuch*. Berlin, Germany: de Gruyter, 2013.

[13] Chantal Bolduc (2017, May 08). Alopecia Areata [Online]. Available: http://emedicine.medscape.com/article/1069931-overview

[14] P. Waltman, *Deterministic Threshold Models in the Theory of Epidemics*, in *Lecture Notes in Biomathematics*, vol. 1. Springer, 1974.

І.Г. Нестерук

СТАТИСТИЧНІ МОДЕЛІ ДЛЯ ДИНАМІКИ ЧЕРНІВЕЦЬКОЇ ДИТЯЧОЇ ХВОРОБИ

Проблематика. Прості математичні моделі забруднення та SIR-модель поширення інфекції використовувались для вивчення розвитку в часі невідомої раніше дитячої хвороби, що виникла в Чернівцях (Україна). Причина багатьох випадків алопеції, що почалися в цьому місті в серпні 1988 р., досі докладно не з'ясована. Відповідно до офіційного звіту урядової комісії останні нові випадки трапились усередині листопада 1988 р., а причиною хвороби названо хімічну екзогенну інтоксикацію. Пізніше ця хвороба отримало назву "Чернівецька хімічна хвороба". Однак майже три роки реєструвалась значно збільшена кількість нових випадків локальної алопеції, що досі не отримало свого пояснення.

Мета дослідження. Порівняння двох різних версій хвороби: хімічної екзогенної інтоксикації та інфекції; визначення параметрів математичних моделей та прогнозування розвитку хвороби.

Методика реалізації. Отримано аналітичні розв'язки диференціальних рівнянь для моделей забруднення середовища та SIR-моделі для епідемії. За допомогою лінійної регресії знайдено оптимальні значення параметрів моделей.

Результати досліджень. З використанням статистичного підходу було визначено оптимальні значення параметрів моделей. Розрахунки показали, що інфекційна версія хвороби є більш імовірною порівняно з популярною гіпотезою про отруєння. Зроблено оцінку можливої дати початку епідемії.

Висновки. Оптимальні значення параметрів SIR-моделі дають змогу розрахувати реалістичну кількість жертв та інші характеристики можливої епідемії. Вони також свідчать, що збільшена кількість випадків локальної алопеції могла би бути частиною тієї ж епідемії, що і "Чернівецька хімічна хвороба".

Ключові слова: визначення моделі; ідентифікація параметрів; статистичні методи; математичне моделювання інфекційних хвороб; SIR-модель; моделі забруднення.

И.Г. Нестерук

СТАТИСТИЧЕСКИЕ МОДЕЛИ ДЛЯ ДИНАМИКИ ЧЕРНОВИЦКОЙ ДЕТСКОЙ БОЛЕЗНИ

Проблематика. Простые математические модели загрязнения и SIR-модель распространения инфекции использовались для изучения динамики неизвестной раньше детской болезни, возникшей в Черновцах (Украина). Причина множества случаев аллопеции, начавшихся в этом городе в августе 1988 г., до сих пор до конца не ясна. В соответствии с официальным отчетом

правительственной комиссии последние новые случаи произошли в середине ноября 1988 г., а причиной болезни названа химическая экзогенная интоксикация. Позже эта болезнь получила название “Черновицкая химическая болезнь”. Однако почти три года регистрировалось значительно увеличенное число новых случаев локальной аллопеции, что до сих пор не получило своего объяснения.

Цель исследования. Сравнение двух разных версий болезни: химической экзогенной интоксикации и инфекции; определение параметров математических моделей и прогнозирование развития болезни.

Методика реализации. Получены аналитические решения дифференциальных уравнений для моделей загрязнения и SIR-модели для эпидемии. С помощью линейной регрессии найдены оптимальные значения параметров моделей.

Результаты исследований. С использованием статистического подхода были определены оптимальные значения параметров моделей. Расчеты показали, что инфекционная версия болезни более вероятна по сравнению с популярной гипотезой об отравлении. Сделана оценка возможной даты начала эпидемии.

Выводы. Оптимальные значения параметров SIR-модели позволяют рассчитать реалистичное число жертв и другие характеристики возможной эпидемии. Они также свидетельствуют, что увеличенное число случаев локальной аллопеции могло быть частью той же эпидемии, что и “Черновицкая химическая болезнь”.

Ключевые слова: определение модели; идентификация параметров; статистические методы; математическое моделирование инфекционных болезней; SIR-модель; модели загрязнения.

Рекомендована Радою
факультету прикладної математики
КПІ ім. Ігоря Сікорського

Надійшла до редакції
31 січня 2017 року