Krynica Morska, 23rd–27th September 2012

DISEASE PROCESS AS A STOCHASTIC PROCESS - MODELING DISEASE COURSE BY STOCHASTIC DIFFERENTIAL EQUATION

Bronisław Sulkowski PhD.

Mathematic, Statistic and Informatic Laboratory, Medical University of Gdańsk, 15 Tuwim Str., 80-210 Gdańsk, bsmedin@amg.gda.pl

ABSTRACT

Looking for a mathematical model of the disease, based on which to model the course of the disease, estimate the patient's chances for a positive effect of prophylactic or therapeutic, and one sequence to assess the risks associated with the decision must take into account non-deterministic nature of the disease process. Therefore, searching for a suitable tool to predict the chance of the patient depending on the implemented procedure I attempted to create a model of the disease as a stochastic process described by stochastic differential equations (SDE). The literature review show there were some attempts to use a SDE in mathematical modeling in medicine [1], [2]. But I did not find this general concept of stochastic models of disease.

This article presents the concept model of the disease process as a stochastic process which is the solution defined SDE, with particular emphasis on the process of prevention. Considerations for other elements of the disease process and application examples can be found in [3].

INTRODUCTION

Decision-making in medicine is marked be a certain degree of uncertainty. We can not always be 100% sure while giving a diagnosis. We cannot usually predict precisely either the course of a disease process, the patient's response to treatment or the risk of any complications occurring.

While reflecting on the way to describe diseases mathematically, we were looking for a method which would let us model the whole course of the disease taking into account the factors that influence it (risk factors, types of therapy, prognostic factors).

This would enable the analysis the impact of risk factors and actions taken on the course of disease, which would determine the likelihood of certain outcomes, depending on the preventive measures taken. It would also be possible to estimate the long-term effects therapeutic intervention.

Since disease processes run in a non-deterministic way, we have made an attempt at their modeling by stochastic differential equations (SDE). Calculating the value of the corresponding functionals of the modeled SDE can create algorithms for predict the impact of taking certain actions, therapeutic, and prophylactic on the chance of the patient.

GENERAL STOCHASTIC DIFFERENTIAL EQUATION OF A DISEASE PROCESS

While analyzing the disease course one has to deal with a great variety of factors, which influence it. They can have an impact on either the onset, modification or end of the pathological process. A general form of a stochastic differential equation simulating a disease process should consider all possible elements and their influence on the disease course. Below is presented the disease model proposed by us. This model is described by formula:

$$\begin{aligned}\mathbf{X}^{d}(t) = \mathbf{X}_{0}^{d} + \int_{0}^{t} \mathbf{a}\left(s, \mathbf{X}^{d}(s)\right)ds + \int_{0}^{t} \boldsymbol{\sigma}\left(s, \mathbf{X}^{d}(s)\right)d\mathbf{W}^{n}(s) + \\ + \int_{0}^{t} \boldsymbol{\eta}\left(s, \mathbf{X}^{d}(s^{-})\right)d\hat{\mathbf{N}}^{\lambda_p(s)}(s) + + \int_{0}^{t} \boldsymbol{\gamma}\left(s, \mathbf{X}^{d}(s^{-})\right)d\hat{\mathbf{N}}^{\lambda_k(s)}(s)\end{aligned} \tag{1}$$

$X^d(t)$ is d-dimensional stochastic process, in which particular components, which can be variables as the risk factors, or variables influencing the disease course, or else describing physiological, morphological or prognostic values.

Dividing the process into components we can write it as:

$$X^{d} = \left[R_r^{d_1}, R_f^{d_2}, R_m^{d_3}, R_p^{d_4}, \Theta^{d_5}, \Psi^{d_6}\right], \tag{2}$$

where $d= d_1 + d_2 + d_3 + d_4 + d_5 + d_6$.

Particular components of the vector stochastic process are as follows:

- $R_r^{d_1}$ - process of changes in risk factors;
- $R_f^{d_2}$ - process of changes in physiological, pathophysiological and pathomorphological values, which are significant for the disease process;
- $R_m^{d_3}$ - process of changes in values modifying a disease course;
- $R_p^{d_4}$ - process describing changes in prognostic factors;
- Θ^{d_5} - vector defining the existence of a disease;
- Ψ^{d_6} - vector of variables defining reaching significant final states.

X_0^d means the initial probability distribution of vector X. If, at a particular moment, we can define a given factor, we assume the value of a variable in a deterministic way. If not, we can present a particular quantity in a form of a random variable with a specified distribution and in numerical modeling we can randomize initial values according to this distribution. . In practice, we deal very often with this situation, because even if the tests do not confirm the existence of malignant transformation, we have no absolute assurance that a seemingly healthy body does not have any cancer cells. This means that, in a theoretically healthy patient, parameter $\Theta^1=0$ only with a certain degree of probability, which is usually less than 1.

Function $\mathbf{a}\left(s, X^d(s)\right)$ is a vector d-dimensional measurable function, specifying physiological, pathophysiological and pathomorphological relations. It can be interpreted as a function defining an "expected, natural" history of a disease. For the parameters, which do not change in time X^i the value of $\mathbf{a^i}\left(s, X^d(s)\right) = 0.$ is If $X^i(s)$ is a component related to the progression of a disease process, it is obvious that it has to be 0 as long as the patient is healthy. In other words, the following condition has to be fulfilled:

$$a^{i}\left(s, X^{d}(s)\right) = 0 \text{ for } \Theta^1=0. \tag{3}$$

The equation component:

$$\int_0^t \sigma\left(s, X^d(s)\right) dW(s) \tag{4}$$

describes variability of the disease process. The function $\boldsymbol{\sigma}\left(s, \boldsymbol{X}^d(s)\right)$ is a matrix function measuring d×n, where n is a dimension of a defined semimartingale. One can take multidimensional Levy's process to describe variability of a disease process. It seems, however, that employing the Wiener's process, which is a particular type of that of the Levy's process would be sufficient, unless there are strong reasons to use the former. Furthermore, it is easier to be simulated. Of course, for all components of Xi with fixed values at specified time intervals we have:

$$\sigma^{i,j}\left(s, X^d(s)\right)=0, \tag{5}$$

and similarly to the case of components of vector function a, if $X_i(s)$ is a component related to progression in a disease process, it has to be 0 as long as the patient is healthy, i.e. the following condition must be fulfilled:

$$\sigma^{i,j}\left(s, X^d(s)\right) = 0 \text{ for } \Theta^1=0. \tag{6}$$

The equation component:

$$\int_0^t \boldsymbol{\eta}\left(s, \mathbf{X}^d\left(s^-\right)\right) d\hat{\mathbf{N}}^{\lambda_p(s)}(s)$$

is related to the onset and development of the disease in progress. Defining respectively: $d_p=d_1+d_2+d_3+d_4$ and $d_k=d_p+d_5$, in a majority of cases the value of the function $\eta^i\left(s, X^d(s^-)\right)$ can be written as:

$$\eta^i\left(s, X^d\left(s^-\right)\right)=\begin{cases} 0 \text{ for } i \le d_p \text{ or } i > d_k \\ 1 \text{ for } i \in \left(d_p, d_k\right] \end{cases} \tag{7}$$

or:

$$\eta^i\left(s, X^d\left(s^-\right)\right)=1 \tag{8}$$

for the components associated with the vector Θ and 0 for everyone else.
$\hat{N}^{\lambda_{p(s)}}(s)$ is an nonhomogeneous Poisson process, and function $\lambda_p(s)$ is usually reduced to one-dimensional function $\lambda_p\left(R_p^{d_4}(t)\right)$, which determines the expected value of onset of the disease or malignant transformation. And finally the last element:

$$\int_0^t \boldsymbol{\gamma}\left(s, \mathbf{X}^d\left(s^-\right)\right) d\hat{\mathbf{N}}^{\lambda_e(s)}(s) \tag{9}$$

defines the creation of a specific final state of the disease. Analogically to the abovementioned, one can specify: $d_p=d_1+d_2+d_3+d_4+d_5$ and $d_k=d_p+d_6$ and write:

$$\gamma^i\left(s, X^d\left(s^-\right)\right)=\begin{cases} 0 \text{ for } i \le d_p \\ 1 \text{ for } i \in \left(d_p, d_k\right] \end{cases} \tag{10}$$

or:

$$\gamma^i\left(s, X^d\left(s^-\right)\right)=1$$

for the components linked with the vector Ψ i 0 for all other components.

With these SDE one can model the course of a disease analyzing the impact of diverse factors on the change in incidence or progress of an illness, or else on the risk of complications. One can also employ the results to build a proper scheme of decisive analysis.

Due to the fact that in a general, stochastic equation of the illness, most of the coefficients will have a 0 value, it is purposeful to consider SDE of an illness as a system of SDEs referring to particular subprocesses.

DESCRIPTION OF THE PREVENTIVE PROCESS

Prophylaxis is a prevention of disease outbreak and development, recurrence and complications. An example of prophylaxis are treating reflux, which prevents the development of Barrett's oesophagus, the curing of which may prevent the appearance of oesophagus cancer. Prophylaxis is also screening which facilitates early detection of lesions, when radical treatment is still possible and effective.

The problem is that running prophylactic screening is costly and often cannot be justified economically. Up till now Japan has introduced general endoscopic screening of the alimentary tract, this being based on the greatest incidence in stomach cancer occurring in this country. In other regions the usefulness of widespread preventive endoscopic screening is doubtful. Still, it seems to be appropriate to select higher risk groups to carry out preventive examinations (endoscopic screening of the upper alimentary tract) in persons of increased risk.

It is a quite a difficult issue to practically select the patients for preventive screening and to specify the frequency of these examinations. Attempts at creating computer systems to evaluate the risk have so far yielded unsatisfactory results. Hence, our work is the next attempt to develop a system of this kind.

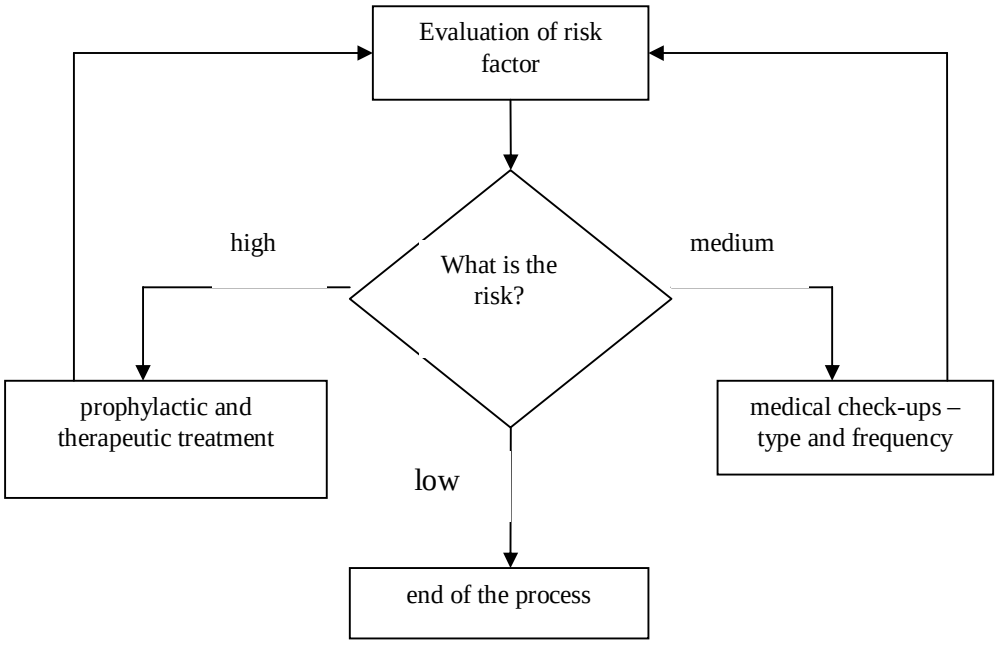

Fig. 1 Diagram of prophylactic process

The study will define the prevention process as referring to analysis and evaluation of the risk factors; this process consists of the following elements:

- describing and evaluating the risk factors;
- assessing threats of developing a disease;
- adjusting the method of further prevention to the level of threat.

From the practical and clinical point of view prophylactic process in neoplastic diseases consists in defining the risk level and, depending on its value, running periodical check-ups at selected patients or possibly taking up procedures reducing the current risk level.

EQUATION OF THE PROPHYLACTIC PROCESS

The most important task in modelling of stochastic prophylactic process is defining a vector of risk factors for illness X. The vector is:

$$\mathbf{R}_r^{d_1} = \left[r_1, r_2, ..., r_{d_1}\right],$$

where particular components of $\mathbf{R}_r^{d_1}$ refer to the values of specific risk factors. The risk factors can be divided depending on the way of variability into:

1. Constant risk factors (e.g. sex, postoperative state in stomach cancer, genetic load);
2. Risk factors defined by means of a function of time (e.g. age, time of exposure to harmful agent);
3. Risk factors described by means of a stochastic process (e.g. description of precancerous states – Barrett's process in oesophagus cancer or gastritis in stomach cancer).

While creating a system of stochastic differential equations of the prophylaxis process, we have to consider equations for risk factors defined in a form of stochastic process. We also have to take into account the main stochastic equation of the prophylaxis, which we will obtain after having reduced the equation (6.1). Since in the prophylaxis analysis, both the second and the third and the last element are usually equal to 0, we will receive:

$$\mathbf{X}^d(t) = \mathbf{X}_0^d + \int_0^t \boldsymbol{\eta}\left(s, \mathbf{X}^d(s^-)\right) d\hat{\mathbf{N}}^{\lambda_p\left(R^{d_1}(s)\right)}(s), \tag{11}$$

where $R^d(t)$ is a vector process of risk factors and $\lambda_p(R^d(t))$ is a dependence of parameter λ from risk factors. Variable parameter $\lambda_p(t)$ is an expected time length from now until the onset of an illness, or in other words, a mean total number of years lived by members of a given population per one case of an illness. While analyzing prophylaxis process, generally, only two components $R_p^{d_1}$ and Θ^{d_5} are taken into account, and Θ has usually only one element (unless one considers the problem of e.g. oesophagus cancer in general, not a specific form of that cancer).

To describe vector SDE one has to define equations for all the elements of risk described by means of a stochastic equation. These, in certain cases can take a form of disease course equation.

ASSESSING THE VALUE OF RISK

Having a specific equation of the prophylactic process, one can simulate trajectories of the process in a defined time length and determine expected values of specific functionals using Monte Carlo methods.

In the case of the abovementioned prophylactic process the functional can be defined as follows:

$$\Psi\left(X^d\right) = \begin{cases} 1 \text{ if } \mathrm{X}^{\mathrm{i}}(\mathrm{T}) = 1, \\ 0 \text{ if } \mathrm{X}^{\mathrm{i}}(\mathrm{T}) = 0. \end{cases}$$

where $X^i(t)$ mean illness in moment T. The value determined by means of simulation and Monte Carlo methods describes the risk of the disease within a given time length.

Another possible approach is simulating the course of trajectory till the moment t, where there is a change in the coordinate's value defining the onset of an illness. The proper functional should then be in a form of:

$$\Psi\left(X^d\right) = t.$$

The value $E\left(\Psi\left(X^{d}\right)\right)$ would then mean the expected time from now till the moment of an illness developing at a given patient.

CRITERIA FOR QUALIFYING FOR THE HIGH RISK GROUP

With the defined values expected for the abovementioned functionals, we can use them to create criteria of patient selection for screening programs. The simplest criterion is adopting a certain level of $E\left(\Psi\left(X^{d}\right)\right)$ as a border level. Beside the selection, there is an issue of frequency in preventive screening. In this matter our proposition is the following. One can estimate the time of the expected onset of the disease be means of the second functional, on the basis of which we can qualify the patient for a high risk group. Next, one can run simulations for a few time horizons and arrange the time for the next examination in such a way that the illness risk in a given patient would be at a level expected in an advance. It seems that this approach can enable optimalization of intended preventive screening and can enhance rationalization of the prophylaxis in neoplastic diseases.

REFERENCES

[1] Kennerly N., Mentor P.: „*Using stochastic differential equations to model the ups and downs of patients with bipolar disorder for clinical purposes*", http://www.ccbb.pitt.edu/BBSI/2006/proposal/kennerly.pdf, 2006.

[2] Lo C.F., Kwok M.Y.: *"Contingent Claims Analysis Approach for Modelling Suicide Risk in Later Life"*, http://www.realoptions.org/papers2004/LoKwokSuicide.pdf

[3] Sulkowski B., *„Zastosowanie metod modelowania stochastycznego w komputerowym systemie wspomagania procesów profilaktyczno-diagnostyczno-decyzyjnych w badaniach endoskopowych górnego odcinka przewodu pokarmowego."*, praca doktorska PG 2010