

Discussion of input-output data in the computer implementation of the PSA-level test during the analysis of prostate cancer

Mária Molnárné Nagy

University of Debrecen, Doctoral School of Information science and Technology
nagymaria1@gmail.com

Abstract

The article analyzes the initial and result dataset of the stochastic process based on the PSA-level measuring used for the detection of prostate cancer. It identifies the required inputs as well as the processing method. It specifies the minimum data set and outlines the opportunities which can be implemented by adding further data.

In case of dates listed in the input data the necessary and sufficient accuracy and the normalization of the time factor are presented.

The article describes the structure and internal consistency of the internal data in the algorithm to implement. The output data of the PRO-FILTER screening package –which will be implemented as a computer service– and one possible way of its computation is presented.

It describes the preparation of the PRO-FILTER database, the steps of progressive data collection, and it gives the points where the continuously expanding database can influence the algorithms.

Keywords: prostate cancer, PSA-level, data analysis

MSC: 60G05, 92C99

1. Introduction

Prostate cancer is one of the most common tumors. It has the second most cancer case incidence in men worldwide. [2] In Hungary it has about 1900 casualties yearly. It is a slow progressing disease, but of course, the earlier it is diagnosed, the better the chances are. The prostate cancer is age related, the tumor rarely appears in men under 45, but the incidence rate grows between age 45 and age 70. The patients diagnosed over 70 are often not treated radically because the risk of the procedures and the expectable life span. In these cases only palliation applied,

and the cause of death is usually another disease. These facts corroborate the need for screening in the men population between 45 and 70.

The currently used screening methods are:

- Prostate Specific Antigen (PSA)-level test –simple relative cheap blood test
- Rectal Digital Examination (RDE) –simple cheap examination performed by specialist
- Multiple Transrectal Ultrasound Guided Biopsy –expensive test with higher complication rate

The first two methods are usable (used) in wide range screening and if these methods give positive result, than is the biopsy performed. “The PSA test is not enough sensitive and not enough specific.” [4]

Our research’s goal is to estimate the characteristics of the PSA-level change process and with these estimations to help the specialist’s work. This article discusses the first implementable algorithm from the aspect of data.

2. Stochastic model of cancer cell growth

The stochastic model of cancer cell growth was described by Andor Dobó in [3], this is the ground line of the algorithm.

„The PSA as probability variable is taken as a time dependent variety, with other words as a stochastic process „

2.1. Describing the process

The $F(t)$ expectation value is approximated with $p(t)$ polynomial Taylor-series.

$$F(t) \approx p(t) = F(0) + \frac{F'(0)}{1!}t + \frac{F''(0)}{2!}t^2 = b + vt + \frac{a}{2}t^2.$$

The usage of quadratic polynomial is enough for this purpose, because the

- PSA-level is always positive
- PSA-level is expectably ascending
- to use higher degree polynomials we need more points, meaning more PSA measurements

For the description of the process we estimate a, b, v variables from at least 3 PSA-level measurements using the least squares method. The measurements are described with $(t_0, p_0), \dots, (t_n, p_n)$ pairs where $t_0 = 0$ and t_i is defined with the accuracy of month. The velocity and the acceleration are important qualifiers of the process.

Velocity: $\dot{p}(t) = \frac{dp}{dt} = v + at$

Acceleration: $\ddot{p}(t) = \frac{d^2p}{dt^2} = a$

2.2. Determining the cancer development time

The development time of the cancer is marked by τ . The c is a lower limit of $p(\tau)$, it represents the PSA-level, which divide negative and positive cases, it will determine if the patient should have a biopsy.

$$p(\tau) = c = b + v\tau + \frac{a}{2}t^2$$

$$\tau_1 = \tau_1(c) = \frac{-v + \sqrt{v^2 - 2a(b - c)}}{a} < 0$$

The value of c will be determined by experimental knowledge of the database, as a start value 0 is usable. The more data is in the database the more accurate value of the c is possible, which can decrease the number of unnecessary biopsies. In the medical practice the PSA-level where the biopsy is proposed is age and race specific, if the c value became accurate enough it can replace this constant level.

2.3. Describing the cancer development

Important property of the cancer development is the doubling time and doubling velocity. These show the aggressiveness of the cancer. The doubling time (t_d) is the time when the last measured value doubles. The doubling velocity (v_d) is the velocity of the process at t_d .

$$t_d = \frac{-v + \sqrt{v^2 + 2ab}}{a}$$

$$v_d = v + at_d = \sqrt{v^2 + 2ab}$$

3. Data sets

This section describes the dataset. It discusses not only the necessary data, but also the possible additions, what can result a advanced service.

3.1. Minimum input dataset

The minimum dataset is the unique id, birth date, blood test results and race. The unique id is the tool for keeping anonymity of the database, when the patient can add further data later, to have a new evaluation. The id can be chosen by the user or can be randomly generated. The date of birth is necessary to calculate the age. The blood test results are entered in measure date, PSA-level pairs. At least three pair is needed to perform the calculations. Race is necessary to determine the accurate limit values for the patient, if the examined population has no high racial diversity the usage of default value can be offered to the user to speed up input process.

3.2. Result dataset

The planned service's result is a printable page which contains the inputs as well the outputs of the calculations. The result page contains the estimated function's variables (a, b, v) , which describe the process. The derived progress properties will be also presented, like velocity, acceleration, τ , doubling time, doubling velocity. The race and age based limit values of biopsy proposal are also presented. These numerical values will be shown on a chart too, where the progress of PSA increase is more easily understandable.

At the end of the result page a medical proceeding proposal will be printed, this proposal will be based on the calculated properties of the process and these proceeding will be determined and validated by oncology and urology specialists. In case the proceedings doesn't propose immediate intervention, but active following, than the service should provide the date of the next blood test, based on the estimated function.

3.3. Further input data possibilities

There are lots of other factors, which can influence the medical proceedings. This section identifies some of them as possible additions for a future algorithm.

The known prostate cancer in family history is a risk factor, which indicates more attention to the PSA-level, for example earlier screening or more often PSA-level check.

If a patient has elevated PSA-level but also other known diseases, the proceeding can be very different, for example some procedures cannot be performed. These correlations could be determined with the help of specialists and could be built in the proposals.

The life expectancy influence the medical proceedings also, if the patient probably dies from other cause before the prostate cancer became fatal, then the practitioner can propose only palliation.

The described method works only patients whom had no intervention made to stop or slow the cancer. If an intervention was performed usually the PSA-level drops and the speed of the elevation shows the success of the intervention. The second elevation could be described by a second similar function or a different type of function. This possibility needs further investigation.

The follow up data of the patients can help the improvement of the algorithm, like the result of biopsy, time of death, cause of death. These data on a big database can give statistical results, which can be built in.

4. Evaluation of the algorithm

When a big enough database is available about the results of the algorithm, the method should be evaluated and adjusted according to the statistical results. This evaluation should contain the following elements. The follow up PSA-levels should

be compared to the estimated ones, to determine accuracy. The positivity rate of the performed biopsies should be calculated. The database should be statistically evaluated clustered by the follow up results. The algorithm should be evaluated regularly and adjusted according to the results.

Acknowledgements. This article would not be possible without

- Mr Andor Dobó , who started the research and described the process of cancer from the mathematical viewpoint
- Dr Károly Nagy, who provided the medical background of the work
- Péter Fésüs, who is the moving spirit of the research and works on the improvement of the algorithm

References

- [1] DR. NAGY KÁROLY, A prosztatatarák kezelése. 16.01.2013 Budapest, Manuscript
- [2] Cancer Research UK (2014), <http://www.cancerresearchuk.org/>
- [3] DOBÓ ANDOR, prosztatatarák terjedésének matematikai leírása, 07.03.2012 Budapest, Manuscript http://doboandor.files.wordpress.com/2012/04/proszatarc3a1k-terjedc3a9sc3a9nek-matematikai-lec3adrc3a1sa_vs8_20120425085752.pdf
- [4] DR. NAGY KÁROLY, A prosztatata rosszindulatú betegségének rövid összefoglalása, 31.13.2012, Budapest, Manuscript, <http://doboandor.files.wordpress.com/2012/04/pca.pdf>