A STATISTICAL PROPOSAL FOR SEQUENTIAL CLINICAL TRIALS IN DIFFERENT CANCER LOCATIONS

A. Koubková, F.T. Barbosa, G. Molenberghs

Keywords: Phase II clinical trials, decision function, response rate distribution, cancer location.

Abstract: This work concerns phase II cancer clinical trials design, which create an important step in drug testing, before it can be used in practice. We specialize on anticancer drugs.

Since different cancers are similar, the idea of this work is to use to estimate the activity of the drug in some specific cancer based on the results of testing the drug in another cancer. The main aim is then to determine an optimal sequence of the cancer locations, how they should be tested, in order to gain as much as possible.

1 Introduction

This contribution is about drug testing and it is a summary of the prepared paper [3].

When a new drug is developed, it needs to come through a number of tests before it is accepted for use in clinical practice. First are the chemical tests in laboratories, then come tests on animals and the last stage consists of tests on humans, so called clinical trials. There are four stages (phases) of clinical trials. In phase I clinical trials the toxicity of the new agent is screened, in phase II its efficacy is estimated, in phase III the drug is compared with the standard treatments and in phase IV rare side effects are monitored.

This work deals with phase II cancer clinical trials. The main aim is to estimate the efficacy of the drug described by its response rate. It is the number of patients for which the drug shows desired activity divided by the total number of patients treated. There are strong criteria on patients, who can enter the tests. The result of phase II clinical trial is either to recommend the drug for phase III trials or to reject it from any further study.

For most of the cancers, there exist some active drugs. These drugs determine the minimal response rate, based on which the new agent is evaluated as active or not. The new agent needs to be preferable than the standard one and it should show at least the same activity.

The phase II clinical trials are one arm studies, in which only moderate number of patients (about 30) with one specific cancer type is treated. Since different cancers behave similarly, one anticancer drug can be active against more of them. Herson [1] came with an idea of a multi-stage design including patients with different tumor types. He wanted to estimate the predictive
probabilities of response using some prior information on the response rate as well as on the degree of tumor non-specificity in response, which would be adjusted as the trial would go on.

In this work we use a similar idea, i.e. to involve the similarity between the cancers in prediction of the drug efficacy. We propose a series of sequentially conducted clinical trials in different tumor locations. To this purpose we develop a decision function determining order of the tumors in the sequence, so that one can gain as much as possible by conducting a trial in the first tumor type. This function takes into account similarity between the cancer types, similarity between the drugs, aggressivity and incidence of the cancers, penalty for treating a patient successfully or unsuccessfully and the results from the clinical trials finished with the new agent. The functional values are updated each time a trial with the new agent is finished.

2 Correlation matrices between the tumor types

Since no easy measure of similarity between the cancers is known, we estimate it by ourselves. There exist more than 100 different cancers. Since many of them are very rare, we used only 25 selected cancer types instead of the particular diseases. These are adrenal, bladder, brain, breast, cervix, colon, endometrium, esophagus, kidneys, acute lymphocytic leukemia (ALL), acute myelogenous leukemia (AML), chronic lymphocytic leukemia (CLL), chronic myelogenous leukemia (CML), liver, Hodgkin’s lymphoma, non-Hodgkin’s lymphoma, small cells lung, non-small cells lung, ovaries, pancreas, prostate, skin, stomach, testes, and urethra cancers.

We studied these cancer types from 4 different points of view based on which we calculated 5 correlation structures. Finally, we add one more matrix concerning the healthy body locations corresponding to the above cancers, instead of the diseases themselves.

The matrix assuming the healthy body organs names Biological correlation matrix and it is calculated as a sample correlation matrix based on 51 features like similarity in structure, cellular base and functions of the organs. Since about 90% of the cancers develop in the epithelial tissue, special emphasis is laid on it.

Next we create Two matrices based on drug response. The first one takes into account 99 different drugs and 105 drug combinations, whereas the second one concerns only the 99 drugs. For the first matrix, the activity of the drugs was described by a weighted average of response rates coming from the recent clinical trials. For the second one we use an indicator describing, whether the drug is used alone in the location, or it is used only in combinations, or it is not used at all.

All the cancers are caused due to mutations in genes involved in cell growth and division. There are two kinds of such mutations, the first one activates the genes for rapid growth (oncogenes), the second one inactivates the genes controlling the growth. The Genetic correlation matrix takes into
account properties of the mutations, their locations and expression of some markers.

The last two matrices, each concerns mutation only in one special gene. The \textit{P53 correlation matrix} characterizes mutations in P53 gene, which controls the cell growth and division. The \textit{EGFR correlation matrix} is calculated based on expression of Epidermal Growth Factor Receptor, which is involved in cellular growth.

Each of the proposed matrices shows quite different pattern. Based on \textit{First correlation matrix based on drug response}, the tumors seem almost independent (the coefficients take values around 0.05). On the other hand, in the \textit{EGFR correlation matrix}, there are more than 68 coefficients between different tumor types equal to one. The coefficients of the other matrices take values from quite wide range, but mostly they are around 0.2.

3 Decision function

Denote $\theta_i$ the true response rate of the new agent in the $i$-th tumor location, $g_i, h_i$ the gain and loss for treating a patient with $i$-th tumor type successfully or unsuccessfully respectively and $G(i)$ the value of the gain function for $i$-th location. The appropriate gain function consists of three parts, and its idea is partially based on [2].

\textit{Part I}. The first part of the function describes the gain obtained in the phase II trial. For phase II trials, it is reasonable to assume the same sample size in all the tumor locations and we choose $n_1 = 30$. In the trial performed in the $i$-th tumor location, there will be $\theta_i n_1$ patients treated successfully and $(1 - \theta_i)n_1$ unsuccessfully, so the gain will be

$$G(\text{phase II}, i) = g_i \theta_i n_1 - h_i (1 - \theta_i)n_1 = (g_i + h_i)\theta_i n_1 - h_i n_1.$$ 

\textit{Part II}. With probability

$$\sum_{x=k_{1,i}+1}^{n_1} \binom{n_1}{x} \theta_i^x (1 - \theta_i)^{n_1-x}$$

the new agent will enter the phase III trials. The critical values $k_{1,i}$ are calculated based on the equality $P_{\theta_0, i}\{	ext{the drug is evaluated as ineffective}\} \leq \beta = 0.1$ describing the type II error.

In phase III trials the sample sizes are different for the particular tumor types and reflect their incidence. We assume the sample sizes from the range $[200, 600]$ and calculate them as $n_{2,i} = 200 + 1.81 \cdot \text{(incidence per year)/1000}$. In the phase III trial conducted in the tumor type $i$, there will be $\theta_i n_{2,i}$ successfully treated patients and $(1 - \theta_i)n_{2,i}$ unsuccessfully treated patients. So the gain will be

$$G(\text{phase III}, i) = g_i \theta_i n_{2,i} - h_i (1 - \theta_i)n_{2,i} = (g_i + h_i)\theta_i n_{2,i} - h_i n_{2,i}.$$
Part III. To the clinical practice the new agent will come with probability
\[
\sum_{x=k_2,i+1}^{n_2,i} \binom{n_2,i}{x} \theta_x^i (1 - \theta_i)^{n_2,i-x}.
\]

For calculation of the critical values \(k_{2,i}\) we used the threshold response rates \(\theta_{0,i}\) as before, but this time we determine them from the equality for the type I error rate (with \(\alpha = 0.05\)).

The numbers of patients treated with the new agent in clinical practice, \(n_{3,i}\), can highly vary for the different tumor types and they will depend on the incidence of the diseases. We assume \(n_{3,i} = 1/2 \cdot (\text{incidence per year})\) yearly treated patients per year. It means that in the \(i\)-th tumor location, \(\theta_in_{3,i}\) patients will be treated successfully and \((1 - \theta_i)n_{3,i}\) unsuccessfully per year. This leads to the last part of the gain function
\[
G(\text{practice}, i) = g_i \theta_i n_{3,i} - h_i (1 - \theta_i)n_{3,i} = (g_i + h_i)\theta_i n_{3,i} - h_i n_{3,i}.
\]

The three parts together give the complete gain function. Since we don’t know the true response rates \(\theta_i\), we need to estimate them. If we know the distributions of \(\theta_i\), we can use their expectations, or their sample means, or we can integrate the gain function over the parameter \(\theta_i\) using its density.

Finally, the pure gain is often not of interest. Preferentially, the more aggressive cancers are treated first. So we add one more term to our gain function. This term is \(k t_i\), where \(k\) is an appropriate constant used to influence the decision function in the desired way, and \(t_i\) is an aggressivity coefficient, for example the death rate. The final gain function is
\[
G(i) = (g_i + h_i)\theta_i n_1 - h_in_1 + \left( \sum_{x=k_1,i+1}^{n_1} \binom{n_1}{x} \theta_x^i (1 - \theta_i)^{n_1-x} \right)
\times \left[ (g_i + h_i)\theta_i n_{2,i} - h_in_{2,i} + \left( \sum_{y=k_2,i+1}^{n_2,i} \binom{n_2,i}{y} \theta_y^i (1 - \theta_i)^{n_2,i-y} \right) \right.
\times \left. ((g_i + h_i)\theta_i n_{3,i} - h_in_{3,i}) \right] + k \cdot t_i.
\]

4 Multimodal response rate distribution

To evaluate the above decision function, the distributions of the response rates \(\theta_i\) need to be estimated. The calculation is divided into two parts. In the first one we assume only the ability of the tumors to respond and the similarity between the drugs, based on which the basic prior distribution, \(q(\theta_i)\) is determined. In the second part we adjust the basic prior for the information of the finished phase II trials performed with the new agent and we get the advanced response rate distribution \(r(\theta_i|X)\), where \(X\) denotes the
A statistical proposal for sequential clinical trials ... results from the finished trials. The distribution \( r(\theta | X) \) can be either prior or posterior depending on whether a trial in tumor type \( i \) was done or not.

Since the response rates \( \theta_i \) are in fact proportions of the binomial distribution, their appropriate prior distribution is a beta distribution. The first part of the basic prior distribution of \( \theta_i \), \( f_1(.) \) is calculated based on the ability of the tumor to respond. If the tumor didn’t respond to any drug in the past, there is a high probability that it will not respond to the new agent either. This suggests a beta distribution with the parameters \( a < 1 \) and \( b > 1 \) and since we want to have an uninformative prior, we choose \( a = 0.8 \), \( b = 1.2 \).

If the tumor responded in the past to any drug, denote the probability of responding (i.e. the number of active drugs divided by the number of all drugs) by \( \eta_i \) and the average response rate of the active drugs as \( \bar{\theta}_i \). Then we can assume that the new agent will have the true response rate \( \bar{\theta}_i \) with the probability \( \eta_i \) and the other values will be less probable. The distribution \( f_1(\theta_i) \) is calculated as a solution of the system of two equations:

\[
\begin{align*}
    f_1' (\bar{\theta}_i) &= \frac{1}{B(a,b)} \theta_i^{a-2} (1 - \bar{\theta}_i)^{b-2} ((a-1)(1 - \bar{\theta}_i) - (b-1)\bar{\theta}_i) = 0, \\
    f_1 (\bar{\theta}_i) &= \frac{1}{B(a,b)} \theta_i^{a-1} (1 - \bar{\theta}_i)^{b-1} = 1 + \eta_i,
\end{align*}
\]

which correspond to the conditions that \( f_1(\theta_i) \) should have mode at \( \bar{\theta}_i \) and its functional value here should be \( f_1(\bar{\theta}_i) = 1 + \eta_i \).

The second part of the basic prior distribution, \( f_2(.) \), describes the similarity of the new agent to drugs with known response rates. Denote the correlation coefficient between the new agent and the known drug as \( D \) (if there are more similar known drugs, it is an average of the corresponding correlation coefficients) and the response rate of the similar known drug in tumor type \( i \) as \( \theta_{D,i} \) (again, for more similar drugs, it is an averaged response rate). The distribution \( f_2(\theta_i) \) should have mode at \( \theta_{D,i} \) and its functional value here should be \( f_2(\theta_{D,i}) = 1 + D \), which leads to the following system of equations determining \( f_2(\theta_i) \)

\[
\begin{align*}
    f_2' (\theta_{D,i}) &= \frac{1}{B(a,b)} \theta_{D,i}^{a-2} (1 - \theta_{D,i})^{b-2} ((a-1)(1 - \theta_{D,i}) - (b-1)\theta_{D,i}) = 0, \\
    f_2 (\theta_{D,i}) &= \frac{1}{B(a,b)} \theta_{D,i}^{a-1} (1 - \theta_{D,i})^{b-1} = 1 + D.
\end{align*}
\]

The basic prior distribution is a standardized sum of \( f_1(.) \) and \( f_2(.) \), i.e.

\[
q(\theta_i) = \frac{f_1(\theta_i) + f_2(\theta_i)}{f_1' (f_1(\theta_i) + f_2(\theta_i)) d\theta_i} = \frac{f_1(\theta_i) + f_2(\theta_i)}{2}.
\]

If a trial is conducted in the tumor location \( i \), then the basic prior \( q(\theta_i) \) should be replaced by a basic posterior distribution.
\[
q(\theta_i|X_i) = \frac{q(\theta_i)L(X_i|\theta_i)}{\int_0^1 q(\theta_i) L(X_i|\theta_i) \, d\theta_i},
\]

where \(L(X_i|\theta_i)\) is the likelihood from the trial.

These basic distributions should be adjusted by information on the trials conducted with the new agent in other tumor locations. We will add a part of the corrected likelihoods to the basic distributions. The amount added depends on the similarity between the tumor types. Assume that the trials were conducted in tumor types \(l_1, \ldots, l_m\) and denote the correlation matrix between these tumors as \(R\). Next denote \(S_i\) the vector of correlation coefficients between the examined tumor type \(i\) and the tumors \(l_1, \ldots, l_m\) already tested and calculate the product \(RS_i\). The likelihoods from the conducted trials should be corrected with respect of the signs of the terms in \(RS_i\). Denote the vector of corrected likelihoods as \(L^c\).

Denote the vector of corrected likelihoods as \(L^c = (L^c_1, \ldots, L^c_m)\), where \(L^c_j = L(\theta_i|X_j)\) if \(\{RS_i\}_j > 0\) and \(L^c_j = 1 - L(\theta_i|X_j)\) if \(\{RS_i\}_j < 0\). Then the advanced distribution of the response rate \(r(\theta_i|X)\) is calculated as

\[
r(\theta_i|X) = \frac{q(\theta_i) + (L^c)^T |RS_i|}{\int_0^1 q(\theta_i) + (L^c)^T |RS_i| \, d\theta_i},
\]

where \(|RS_i|\) denotes the absolute value of \(RS_i\) taken by terms and where \(q(\theta_i)\) is replaced by \(q(\theta_i|X_i)\), when a trial in tumor type \(i\) was conducted.

5 Unimodal response rate distribution

The final estimated distribution proposed in the previous section is in general multimodal. Sometimes it is more convenient to have a unimodal one. In such a case, it should be a beta distribution with parameters \(a_{F,i}\) and \(b_{F,i}\), which need to be estimated. The estimation procedure consists of three parts.

**Part I.** Here is estimated the base of the distribution concerning the ability of the tumor types to respond. If the tumor \(i\) didn’t respond in the past to any drug, the parameters of the beta distribution are \(a_{B,i} = 0.8\) and \(b_{B,i} = 1.2\).

If the tumor responded to some chemical agent, the parameters \(a_{B,i}\) and \(b_{B,i}\) are calculated by solving the system of equations (1).

**Part II.** The parameters of a beta distribution, \(a_{D,i}\) and \(b_{D,i}\), calculated in this part describe the similarity of the new drug with the known drugs. These parameters are established by solving the system of equations (2).

**Part III.** This part concerns on the phase II clinical trials possibly conducted with the new agent. Assume the trials were performed in tumor types \(l_1, \ldots, l_m\) and as before, denote by \(R\) the correlation matrix between them and by \(S_i\) the vector of correlation coefficients between them and the examined tumor type \(i\). Next denote the sample size in the trial conducted in tumor location \(j\) as \(n_j\) and the number of responses observed there as \(r_j\). Now create two vectors \(v_i\) and \(u_i\) of size \(m\), such that \(v_{i,j} = r_j\) and \(u_{i,j} = n_j - r_j\)
if the $j$-th term of $\mathbf{R}_i$ is positive, and $v_{i,j} = n_j - r_j$ and $u_{i,j} = r_j$, if it is negative. The parameters $a_{T,i}$ and $b_{T,i}$ are then calculated as

$$a_{T,i} = |\mathbf{R}_i|v_i, \quad b_{T,i} = |\mathbf{R}_i|u_i,$$

where $|\mathbf{R}_i|$ denotes the vector of absolute values.

The parameters of the final response rate distribution are then simply the sums of the three parts:

$$a_{F,i} = a_{B,i} + a_{D,i} + a_{T,i}, \quad b_{F,i} = b_{B,i} + b_{D,i} + b_{T,i}.$$ 

The only one exception is the case, where $a_{B,i} = a_{D,i} = 0$. Then the final parameters are $a_{F,i} = 0.8 + a_{T,i}$ and $b_{F,i} = 1.2 + b_{T,i}$.

6 Simulations

A small simulation study to show the properties of the above introduced procedure was conducted. A drug with the following response rates was assumed: 0.6 in breast cancer and $2|s_i|(0.5 + \text{sign}s_i)(0.6 - 0.5)$ in the other cancers, where $s_i$ are the correlation coefficients with breast cancer from the Biological correlation matrix. This drug was assumed to be dissimilar with any known drug and the Biological correlation matrix was taken for the calculations. The gains and losses $g_i = h_i = 1$, $g_i = 1 < h_i = 2$ and $g_i = 2 > h_i = 1$ for treating a patient were compared. Also the multimodal and unimodal response rate distributions were compared.

The simulation procedure was as follows. At first a tumor type for the first test was selected. Next a trial was simulated here and the decision function was updated to choose a tumor type for second trial. Then again a trial was simulated and the decision function was updated once more.

The procedure using equal gain and loss for treating a patient and the one preferring gain recommended the tumor types with high incidence. The third procedure gives some balance between the incidence and aggressivity of the disease. The main difference between the two types of the response rate distribution is that the unimodal distribution is much more informative than the multimodal one.

7 Conclusion

A decision procedure was proposed to determine an optimal order of tumor types for sequential phase II clinical trials with a new anticancer drug. At first a prior response rate distribution for each tumor type was estimated. It can be either unimodal or multimodal distribution. Using this distribution a gain function is evaluated and so the order in the tumor sequence determined.

The gain function takes into account properties of the diseases, such as its incidence, aggressivity, similarity with other tumors and evaluation for treating a patient. Also the characteristics of the new drug, such as its similarity with the known drugs or target, for which it is designed, are assumed. This function is updated each time a clinical trial with the new agent is finished.
The properties of the procedure are shown by a small simulation study, which compares the two types of response rate distribution and also different penalties for treating a patient.

Appendix

Here is a graph depicting how the distribution of the response rate is evolved. The first two lines correspond to the multimodal response rate distribution and the second two lines to the unimodal one. In both cases, the distributions of response rate in breast cancer and in prostate cancer is depicted.

References


Acknowledgement: The work was supported by the Research plan MSM 113200008.

Address: A. Koubková, Department of Statistics, Charles University, Prague, Czech Republic
F.T. Barbosa, G. Molenberghs, Center for Statistics, Limburgs Universitair Centrum, Diepenbeek, Belgium

E-mail: koubkova@ksi.ms.mff.cuni.cz