Development of a methodological approach to determine regression equations in the study of the technology for manufacturing tablets based on quercetin

When developing the technology of tablets and tablet masses one of the tasks is to determine the amounts of excipients required to obtain pharmaco-technological quality indicators that meet the requirements of the State Pharmacopoeia of Ukraine (SPhU).

**Aim.** To develop an algorithm for determining the type of three-factor mathematical models with dependent variables.

**Materials and methods.** The study object was experimental observations of the quantitative effect of excipients in the tablet composition based on a solid dispersion of quercetin on the pharmacopeial characteristics of this dosage form, in particular on its flowability. The experimental data were processed by the planned experiment using Mathcad 15 and MS Excel software according to the algorithm proposed.

**Results and discussion.** It has been found that the identification of mathematical models in pharmaceutical studies with three dependent factors, which total value is determined by the quantitative composition of the dosage form and fixed at a certain level, is difficult to perform due to the difficulty of interpreting the multiple regression parameters as characteristics of factors in isolation through their correlability. It has been proven that the replacement of variables leads to the determination of a mathematical model that does not reveal the mechanism of the factors' action and is a static description of their overall impact on the indicator studied.

**Conclusions.** As a result of the research conducted, regression equations were found to determine the effect of the amount of these excipients on the tablet flowability. It has been found that the most influential role in determining the target in the factor space studied is played by the interaction of factors. The algorithm for determining the mathematical description of the dependence with three variables has been proposed. Based on the fact that the research is conditioned by strict conditions in the quantitative composition, the chosen mathematical model of the dependence of the quadratic equation with two factors, for which there are limitations, has no correlation. The model determined is characterized by the possibility of providing the graphical interpretation and simplification of analysis and can be used for forecasting and optimization.

**Key words:** quantitative effect of excipients; three-factor mathematical model; dependent mixture factors; identification algorithm

O. V. Kutova, I. V. Kovalevska, N. V. Demchenko

National University of Pharmacy of the Ministry of Health of Ukraine
**Introduction.** Quantitative research in the development of the tablet technology involves finding the optimal composition of a dosage form (fills, disintegrants, binders, glidants and lubricants, etc.), parameters of technological modes determined by the type of a dosage form (compression pressure, compression rate, film-forming solution temperature), the drying time of the coating, etc., as well as ensuring the values of pharmacotechnological quality indicators, in accordance with the requirements of the State Pharmacopoeia of Ukraine (SPhU).

It should be noted that there is often a need to find a significant amount of excipients when developing the technology of tablet drugs. Multifactoriality is one of the specific features of these studies at the experimental stage. In this study, the determination of regression relationships is associated with three quantitative factors, which total value is fixed at a certain level. Interpretation of multiple regression parameters in this case is complicated by their correlation.

Various modifications of the planned experiment are used to construct multifactor models [1-5]. In the study of three factors in the presence of certain restrictions on two of them, the third factor becomes dependent and changes under the conditions of the passive experiment. As a result, the levels of the dependent factor $x_3$ in the experiments are only recorded, but not specified. In this case, the researcher is not able to control the dependent factor and becomes a passive observer of the phenomenon occurring. We have to deal with the values of the factors, which change unevenly. Due to the uneven change in the corresponding range of the dependent factor, the usual methods of finding empirical dependences are unsuitable, and there is a need to develop the effective one in the conditions studied [6].

The aim of the article was to develop an algorithm for determining the type of three-factor mathematical models with dependent variables.

**Materials and methods.** The algorithm for identifying the regression equation for three dependent quantitative factors is shown on the example of the development of the tablet technology based on solid dispersion of quercetin [7]. According to the standard procedure at the first stage as a result of studies, the effect of excipients on the pharmacological characteristics of tablets was determined by their qualitative composition. Three excipients significantly affecting the tablet mixture and tablets — microcrystalline cellulose ($x_1$ factor), nesulin ($x_2$ factor) and sodium starch glycolate ($x_3$ factor) were selected. At the second stage it was necessary to determine the quantitative content of excipients under the following conditions: $12 < x_1 < 14, 12 < x_2 < 14, x_1 + x_2 + x_3 = 29$.

The tablet weight and finished tablets were monitored according to five pharmaco-technological indicators (reviews, target functions): $y_1$ — the Kara index, $y_2$ — the tablet mass flowability, $y_3$ — the Hausner coefficient, $y_4$ — the tablet strength, $y_5$ — the tablet disintegration time.
In classical regression analysis, the structure considered is known before planning, but in pharmaceutical research it is not always so. In many cases, the purpose of building a regression model is to elucidate the relationship between independent factors and the dependent pharmacopoeial index. Determining this structure is one of the tasks in the regression analysis. It is usually formed as a search for an equation with the best estimates of static significance. This is not entirely true as the aim of the study is not the goal of mathematical statistics alone.

The strategy for constructing the equation determines the requirements for planning the experiment. In this case, the formulation of the problem does not ensure the independence of factors, and as a result, there will be no good conditionality of the matrix full of factor planning.

To achieve the factor independence, the variable $x_3$ was excluded from consideration [8-10]. This allowed us to apply the experimental areas of the standard form $2^2$ and $3^2$.

The mathematical processing of the results of the study was performed using Mathcad 15 software [11], the least squares method (LSM). The estimation of the significance of regression equations and their coefficients was performed using statistical functions of the MS Excel spreadsheet [12].

**Results and discussion.** For this study, the effect of the third excipient P – a loosening substance, which amount directly depends on the number of the first two excipients, is excluded from consideration as the most correlated factor. To propose a type of the regression equation, we used a graphical approach based on the visual analysis of location of the points of the experimental surface obtained according to Plan $2^2$ (Tab. 1) using Mathcad 15 software (Fig. 1a)

The assumption about the linearity of the mathematical description was refuted by the type of response surface. In addition, the coefficient of determination of the linear equation $y_2 = f(x_1,x_2)$ was equal to 0.28, the average relative error of calculations at base points exceeded the accepted 3 %. Based on the graphical interpretation of the experimental dependence and the experimental plan $2^2$, the regression equations with the interaction of factors were analyzed:

$$y_2(x_1,x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2.$$ \hspace{1cm} (1)

The plan with the minimum number of experiments (saturated plan) allows the method of least squares to determine the coefficients of the equation (1):

$$y_2(x_1,x_2) = 230.23 - 17.55 x_1 - 18.035 x_2 + 1.405 x_1 x_2,$$ \hspace{1cm} (2)

but the requirements for the quality of the evaluation of the function and its parameters are not met despite the maximum coefficient of determination ($R^2 = 1$), the absence of a relative calculation error at base points and satisfactory coincidence of experimental and theoretical response surfaces (Fig. 2).

To make a final decision on the determination of the mathematical model, despite a good assessment of the adequacy

---

**Table 1**

<table>
<thead>
<tr>
<th>No.</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$y_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(+)</td>
<td>(–)</td>
<td>21.42 ± 0.4</td>
<td>4.5 ± 0.35</td>
<td>0.85 ± 0.05</td>
<td>0.56 ± 0.04</td>
<td>10 ± 0.5</td>
</tr>
<tr>
<td>2</td>
<td>(+)</td>
<td>(+)</td>
<td>12.03 ± 0.6</td>
<td>7.92 ± 0.86</td>
<td>1.39 ± 0.07</td>
<td>0.58 ± 0.03</td>
<td>13 ± 0.7</td>
</tr>
<tr>
<td>3</td>
<td>(–)</td>
<td>(+)</td>
<td>30.23 ± 0.9</td>
<td>2.68 ± 0.4</td>
<td>1.31 ± 0.055</td>
<td>0.61 ± 0.04</td>
<td>9 ± 0.66</td>
</tr>
<tr>
<td>4</td>
<td>(–)</td>
<td>(–)</td>
<td>17.0 ± 0.3</td>
<td>4.83 ± 0.43</td>
<td>1.15 ± 0.003</td>
<td>0.52 ± 0.03</td>
<td>5 ± 0.45</td>
</tr>
</tbody>
</table>

---

Fig. 1. The experimental response surfaces: a) by complete factorial experiment $2^2$, b) by complete factorial experiment $3^2$
of two statistical indicators, it was proposed to analyze the separate influence of factors on the target indicator within the interval studied. In this study, the increase in the content of microcrystalline cellulose with different amounts of neusilin in different ways affects the target: if \( x_2 = 12 \% \), then an increase in \( x_1 \) leads to an increase in \( y_1 = f(x_1) \) (Fig. 3a).

An increase in the content of neusilin at \( x_1 = 12 \% \) leads to a decrease in the value of the mass flowability, at \( x_1 = 14 \% \) – to an increase in this indicator (Fig. 3b).

The main effect of the factors is masked and gives the final increasing linear dependence, hiding the non-linear behavior of variables within the factor space. Thus, the available information indicates the impossibility of using the equation (1) as a mathematical description of this study despite a good assessment of its adequacy on two statistical indicators, and indicates the need to consider a second-order regression model.

Increasing the experimental base by adding the experiment in the center of the plan \( 2^2 - y_2(13.13) = 7.0 \) allows obtaining statistical estimates of the equation (1): the coefficient of determination \( R^2 = 0.77 \) shows a sufficient approximation accuracy; the reliability according to the level of significance of the Fisher criterion is 0.58 – the model is not significant; the p-values for the coefficients of the equation \( 0.35 \), and it proves their insignificance. The results obtained do not allow us to consider the regression equation (1) as an adequate mathematical model.

There is a need to plan an experiment for a quadratic model. It is advisable to build the optimal plan for the quadratic model in such a way that it includes the points of the optimal plan for the linear model, maintaining orthogonality (Tab. 2).

The graphical interpretation of the experimental response surface \( y_2 \) obtained from the results of the experimental matrix according to Plan \( 3^2 \) differs from option \( 2^2 \) (Fig. 1b).

Calculations by the 2nd order regression equation, which coefficients are determined using the *regress* function in Mathcad 15:

\[
y_2(x_1,x_2) = -93.3472 - 17.8683x_1 + 32.325x_2 + 1.405x_1x_2 + 0.0133x_1^2 - 1.9367x_2^2,
\]

give a slight average relative error at base points not exceeding 0.5 %. Experimental and theoretical response surfaces, which are given in Fig. 4, match quite well.
The factor experiment planning matrix $3^2$ and the research results

<table>
<thead>
<tr>
<th>No.</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$y_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14 (+)</td>
<td>12 (–)</td>
<td>21.42 ± 0.4</td>
<td>4.15 ± 0.35</td>
<td>0.85 ± 0.05</td>
<td>0.56 ± 0.04</td>
<td>10 ± 0.5</td>
</tr>
<tr>
<td>2</td>
<td>14 (+)</td>
<td>14 (+)</td>
<td>12.03 ± 0.6</td>
<td>7.42 ± 0.86</td>
<td>1.39 ± 0.07</td>
<td>0.58 ± 0.03</td>
<td>13 ± 0.7</td>
</tr>
<tr>
<td>3</td>
<td>12 (–)</td>
<td>14 (+)</td>
<td>30.23 ± 0.9</td>
<td>3.18 ± 0.4</td>
<td>1.31 ± 0.055</td>
<td>0.61 ± 0.04</td>
<td>9 ± 0.66</td>
</tr>
<tr>
<td>4</td>
<td>12 (–)</td>
<td>12 (–)</td>
<td>17.00 ± 0.3</td>
<td>5.53 ± 0.43</td>
<td>1.15 ± 0.03</td>
<td>0.52 ± 0.03</td>
<td>5 ± 0.45</td>
</tr>
<tr>
<td>5</td>
<td>14 (+)</td>
<td>13 (0)</td>
<td>16.7 ± 0.35</td>
<td>7.8 ± 0.8</td>
<td>1.17 ± 0.32</td>
<td>0.62 ± 0.04</td>
<td>10.2 ± 1.2</td>
</tr>
<tr>
<td>6</td>
<td>12 (–)</td>
<td>13 (0)</td>
<td>23.6 ± 0.62</td>
<td>6.2 ± 0.64</td>
<td>1.19 ± 0.3</td>
<td>0.61 ± 0.05</td>
<td>5.5 ± 0.6</td>
</tr>
<tr>
<td>7</td>
<td>13 (0)</td>
<td>14 (+)</td>
<td>21.11 ± 1.1</td>
<td>5.35 ± 0.52</td>
<td>1.35 ± 0.35</td>
<td>0.56 ± 0.05</td>
<td>10.1 ± 0.9</td>
</tr>
<tr>
<td>8</td>
<td>13 (0)</td>
<td>12 (–)</td>
<td>19.2 ± 0.8</td>
<td>4.8 ± 0.4</td>
<td>1.17 ± 0.4</td>
<td>0.57 ± 0.07</td>
<td>5.5 ± 0.6</td>
</tr>
<tr>
<td>9</td>
<td>13 (0)</td>
<td>13 (0)</td>
<td>20.15 ± 0.95</td>
<td>7.0 ± 0.58</td>
<td>1.26 ± 0.45</td>
<td>0.58 ± 0.04</td>
<td>6.4 ± 0.54</td>
</tr>
</tbody>
</table>

Statistical estimates of the equation (3) are as follows: the coefficient of determination – $R^2 = 0.999$ shows an excellent approximation accuracy; the reliability according to the level of significance of the Fisher criterion is 0.0000429, that is, the model is significant; the p-value for all coefficients of the equation is much lower than 0.05, except for the value of 0.77 for the coefficient at $x^2$. The presence of alternating positive and negative deviations indirectly confirms the absence of a systematic error in the regression equation constructed. Confidence intervals of the function at base points are: $y_2(14,12) = 4.1586 ± 3.82; y_2(14,14) = 7.4402 ± 3.82; y_2(12,14) = 3.145 ± 3.82; y_2(12,12) = 5.4836 ± 3.82; y_2(14,13) = 7.736 ± 3.17; y_2(12,13) = 6.25 ± 3.17; y_2(13,14) = 5.2794 ± 3.17; y_2(13,12) = 4.808 ± 3.17; y_2(13,13) = 6.980 ± 3.17.

The results obtained allow us to consider the regression equation (3) as an adequate mathematical description.

By the equation (3) it is impossible to determine the true mechanism of influence of factors $x_1, x_2, x_3$ on the result trait due to the reasons previously mentioned. Therefore, the real mechanism of action and interaction of factors can be proposed by comparing the possible mathematical structures of the equation (3), and using the substitution $x_3 = 29 – x_1 - x_2$. The equation (5) is identical to three parametric dependences:

$$y_2(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_1 x_2 + a_5 x_1 x_3 + a_6 x_2 x_3.$$  \hfill (4)

Taking into account that the assessment of the influence of the coefficients of the equation (5) using MS Excel, shows no significance of the component $x_1^2$ (p-value is equal to 0.77), it is possible not to consider also the interaction of variables at level $x_1 x_3$. Thus, we can offer the form of the equation as a mathematical description:

$$y_2(x_1, x_2, x_3) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_1 x_2 + a_5 x_2 x_3 + a_6 x_2^2,$$  \hfill (5)

with the coefficients determined:

$$y_2(x_1, x_2, x_3) = -603.72 + 17.522 x_1 + 3.124 x_1 x_2 + 1.719 x_2 x_3 - 0.218 x_2^2.$$  \hfill (6)

In the equation (8), all the coefficients are significant, the average calculation error is 0.6 %, the coefficient of determination $R^2 = 0.9994$ shows an excellent approximation accuracy; the reliability according to the level of significance of the Fisher criterion is 0.0000562, the graphical interpretation of the given equation proves the coincidence of the response surfaces similar to Fig. 4.

Confidence intervals of the function at base points were determined using the Student’s test with 95 % reliability: $y_2(14,12) = 4.15 ± 3.82; y_2(14,14) = 7.45 ± 3.82; y_2(12,14) = 3.15 ± 3.82; y_2(12,12) = 5.49 ± 3.82; y_2(14,13) = 7.74 ± 3.17; y_2(12,13) = 6.256 ± 3.17; y_2(13,14) = 5.3 ± 3.17; y_2(13,12) = 4.826 ± 3.17; y_2(13,13) = 6.99 ± 3.17.

![Fig. 4](image-url) The graphical interpretation of equation (5) and determination of the calculation error using it.
In the same way, regression equations were determined for all targets studied in order to further optimize the technology of manufacturing tablets [13].

Conclusions and prospects for further research

The essence of the methodological approach proposed to the determination of regression equations is reflected in the example of a mathematical description of one of the five pharmacopoeial characteristics of tablets studied – the flowability of the tablet mass based on quercetin.

To reduce the required number of experiments, taking into account the dependence of the variables affecting the target indicator, the mathematical processing of the research results has been performed using two factors, which content in the dosage form studied is strictly limited by the research conditions.

A satisfactory estimate of the regression equation determined by the fractional plan with two variables has been revised after the analysis of the isolated influence of factors. There is a need to expand the information base within the factor space in order to consider the regression equation of the 2nd order.

The result of the mathematical processing is the determination of a regression equation in the form of an incomplete algebraic polynomial of the 2nd degree, in which the constants determined do not reflect the actual degree of influence of the variables. After the mathematical transformation, the dependence of three variables with refined coefficients has been obtained, revealing the real mechanism of the influence of factors on the indicator studied. Thus, the presence of the excipients in combination with the active ingredients in the formulation of tablets developed has a significant effect on the flowability of the test dosage form. Increasing the amount of neusilin from 12 to 14 % can lead to a decrease in flowability, but in the mixture with the amount of microcrystalline cellulose from 12 to 14 % and sodium starch glycolate from 1 to 3 % the target value increases.

The methodological approach proposed is applied to all pharmacopoeial characteristics studied. The regression equations obtained can be used to optimize the technology of manufacturing tablets.

Conflict of interests: authors have no conflict of interests to declare.

REFERENCES


REFERENCES


Information about authors:
Kutova O. V., Candidate of Engineering Science (Ph.D.), associate professor of the Department of Technologies of Pharmaceutical Preparations, National University of Pharmacy of the Ministry of Health of Ukraine. E-mail: paxtoxt@gmail.com. ORCID: https://orcid.org/0000-0002-3761-2831

Kovalevska I. V., Doctor of Pharmacy (Dr. habil.), associate professor of the Department of Industrial Technology of Drugs, National University of Pharmacy of the Ministry of Health of Ukraine. E-mail: inga.kovalevskaya@gmail.com. ORCID: https://orcid.org/0000-0001-5610-8334

Demchenko N. V., Candidate of Economics (Ph.D.), Associate Professor of the Organization and Economics of Pharmacy Department, National University of Pharmacy of the Ministry of Health of Ukraine. E-mail: demchenata@ukr.net. ORCID: https://orcid.org/0000-0001-5915-0087

Відомості про авторів:
Кутова О. В., кандидатка технічних наук, доцент кафедри технологій фармацевтичних препаратів, Національний фармацевтичний університет Міністерства охорони здоров’я України. E-mail: paxtoxt@gmail.com. ORCID: https://orcid.org/0000-0002-3761-2831

Ковалевська І. В., доктор фармації, наук, доцентка кафедри заводської технології ліків, Національний фармацевтичний університет Міністерства охорони здоров’я України. E-mail: inga.kovalevskaya@gmail.com. ORCID: https://orcid.org/0000-0001-5610-8334

Демченко Н. В., кандидат економічних наук, доцент кафедри організації і економіки фармації, Національний фармацевтичний університет Міністерства охорони здоров’я України. E-mail: demchenata@ukr.net. ORCID: https://orcid.org/0000-0001-5915-0087

Сведения об авторах:
Кутова О. В., кандидат техн. наук, доцент кафедры технологий фармацевтических препаратов, Национальный фармацевтический университет Министерства здравоохранения Украины. E-mail: paxtoxt@gmail.com. ORCID: https://orcid.org/0000-0002-3761-2831

Ковалевская И. В., доктор фармац. наук, доцент кафедры заводской технологии лекарств, Национальный фармацевтический университет Министерства здравоохранения Украины. E-mail: inga.kovalevskaya@gmail.com.

Демченко Н. В., кандидат экон. наук, доцент кафедры организации и экономики фармации, Национальный фармацевтический университет Министерства здравоохранения Украины. E-mail: demchenata@ukr.net. ORCID: https://orcid.org/0000-0001-5915-0087

Надійшла до редакції 02.12.2021 р.