Article's History: Received: 06.11.2023; Revised: 15.01.2024; Accepted: 27.02.2024

Bulletin of Medical and Biological Research

Vol. 6, No. 1
2024

DOI: 10.61751/bmbr/1.2024.15

Journal homepage: https://bmbr.com.ua/en

UDC 616.347:577

Development and validation of a mathematical model for predicting the development of gastro-oesophageal reflux disease based on oesophagogastroduodenoscopy

Oleksandr Halushko*

Postgraduate Student
Volodymyr Vynnychenko Central Ukrainian State University
25000, 1 Shevchenko Str., Kropyvnytskyi, Ukraine
https://orcid.org/0009-0009-0275-6542

Yurii Hurtovyi

PhD in Physical and Mathematical Sciences, Associate Professor
Volodymyr Vynnychenko Central Ukrainian State University
25000, 1 Shevchenko Str., Kropyvnytskyi, Ukraine
https://orcid.org/0000-0002-1499-7089

Abstract. The purpose of this study was to identify a set of prognostic factors for the progression of gastro-oesophageal reflux disease for use in the development of a mathematical model for predicting this disease based on the results of oesophagogastroduodenoscopy. The study identified a range of prognostic factors for gastro-oesophageal reflux disease and a statistical method was employed to determine the level of their correlation with the development of the disease. The study found a link between certain clinical indicators and the occurrence of gastro-oesophageal reflux disease, which led to the formation of a set of prognostic factors for the progression of gastro-oesophageal reflux disease, including heartburn, frequent belching, regurgitation, damage to the mucous membrane of the oesophagus, stomach, duodenum, the presence of chronic gastroduodenitis, gastrointestinal dysfunction, bile reflux. In creating the mathematical prediction model, the logistic regression method was used to identify the correlation between the patient's clinical indicators and the occurrence of reflux disease and to determine the probability of its progression. To bring the clinical information in line with the statistical formula, it was assigned the values of independent variables, and the presence or absence of a particular indicator was coded using the binary number system. To test the developed model, recommendations were given to assess the statistical significance of the independent variables to determine its adequacy and to determine the predictive ability by testing on an independent sample of patients. The developed prognostic model is of great practical significance for patients, the healthcare industry, and the further development of the field, as it enables prompt detection of diseases and suitable prevention and treatment measures, increases the diagnostic potential of the industry, optimises the allocation of medical resources, and leverages machine learning and artificial intelligence capabilities based on the existing model

Keywords: heartburn; regurgitation; oesophageal mucosa; gastrointestinal tract; bile; logistic regression

INTRODUCTION

The need to develop a reliable model for predicting the development of gastro-oesophageal reflux disease arose due to problems associated with the late detection of the disease and, as a result, the complexity of its treatment. Gastro-oesophageal reflux disease (GORD) is an acute or chronic inflammation of the oesophagus (oesophagitis) that develops as a result of the backflow of stomach contents into the oesophagus (reflux). The disease is accompanied by a range of symptoms, including heartburn, belching, heaviness and/or pain in the oesophagus when swallowing food or saliva, hiccups, a lump in the throat, and vomiting. Depending on the stage of GORD, its clinical signs

Suggested Citation:
Halushko O, Hurtovyi Yu. Development and validation of a mathematical model for predicting the development of gastro-oesophageal reflux disease based on oesophagogastroduodenoscopy. Bull Med Biol Res. 2024;6(1):15–23. DOI: 10.61751/bmbr/1.2024.15

*Corresponding author

change – from erosive changes in certain areas of the oesophagus at the first stage to the formation of ulcers, stricture, stenosis, and metaplasia of the oesophageal mucosa (degeneration of its epithelial cells) at the fourth stage. Apart from having a considerable impact on the quality of life (reduced ability to work, changes in the usual regime, and dietary restrictions), the progression of GORD can be a factor in many complications, the most serious of which is the development of a malignant tumour of the oesophagus.

Currently, there are several major problems associated with the prediction of GORD. Thus, the high variability of its symptoms and the lack of reliable clinical markers considerably complicate the prognostic process and affect its accuracy. T. Reva *et al*. [1], K.M. Chue *et al*. [2], S. Shinozaki *et al*. [3] investigated this issue. In these studies, the researchers found the dependence of GORD variants and concomitant lesions on the level of reflux acidity, established the usefulness of the Hill classification for predicting the development of this disease after laparoscopic sleeve gastrectomy, and assessed the significance of such prognostic factors as smoking, functional dyspepsia, hiatal hernia, and severe gastric atrophy. The effect of different types of bile on the risk of oesophageal mucosal damage can be considered a prognostic factor for the progression of GORD, but focusing on only one parameter, without considering other clinical characteristics of the patient, can lead to a decrease in the accuracy of the prognosis. The Hill classification of the gastro-oesophageal junction has shown a close relationship with the development of postoperative GORD, but it is not advisable to use it as a universal indicator for predicting the disease in other clinical situations. The same statement applies to smoking, functional dyspepsia, hiatal hernia, and severe gastric atrophy, as these risk factors have been considered in the context of predicting the progression of GORD after successful eradication of *Helicobacter pylori*, while their significance for other clinical cases has not been fully established. E.M. Wessels *et al*. [4] determined the risk of developing GORD after oral endoscopic myotomy. Thus, oral endoscopic myotomy as a risk factor for GORD can be considered as a prognostic factor for the progression of the disease, but only for the group of patients who are indicated for the treatment of achalasia.

When studying ways to predict the development of GORD, certain questions arise about the methods of clinical trials of patients. D. Armstrong *et al*. [5], S.P. Lee *et al*. [6] investigated these issues. In the study, the researchers used a survey to identify the differences between clinicians in their approach to initial investigations to detect and predict the progression of GORD – some doctors favoured the use of upper endoscopy, while others preferred oesophageal manometry and pH monitoring; the researchers also noted differences in the interpretation of endoscopic images of the gastro-oesophageal junction between observers with different levels of experience (expert endoscopists and trainees). S. Myasoyedov & S. Andreieshchev [7], B. Shevchenko *et al*. [8] analysed the effectiveness of morphofunctional oesophagological monitoring and endoscopic ultrasound examination. The obtained results demonstrate the effectiveness of imaging methods due to their high information content and variability according to the individual clinical picture of the patient, but the possibility of error in the interpretation of the results due to the subjective perception of the observer is still a problem.

The analysis of the studies revealed the lack of universal reliable prognostic factors and a single research method for the prognosis of GORD, and therefore the purpose of this study was to establish clinical indicators that can be used in combination as prognostic factors for the development of this disease. The main objectives of this study were to develop a mathematical model for predicting the development of GORD based on the identified prognostic indicators obtained by oesophagogastroduodenoscopy and to establish reliable methods for validating this model.

MATERIALS AND METHODS

The creation of a mathematical model for predicting the progression of GORD was based on the use of clinical indicators that characterise the state of the gastrointestinal tract and have prognostic significance for the development of the disease, and the development of the most accurate method for identifying statistically significant correlations between these indicators and GORD.

The prognostically significant clinical indicators were identified by analysing the literature on the relevant topics, including the findings of theoretical and clinical studies on the prognosis, diagnostics, and treatment of GORD; the investigation of the profile of informativeness and reliability of oesophagogastroduodenoscopy as a method for collecting information on the clinical characteristics of the gastrointestinal tract of patients with GORD was based on the analysis of the technical characteristics of this method and the review of literature sources, including materials from clinical trials using it. The studies were searched for in PubMed, Web of Science, Google Scholar, and Scopus databases.

After clarifying the prognostically significant clinical indicators and determining the method of their obtaining, a mathematical model was developed to predict the progression of GORD. The first stage in building the model involved collecting clinical information from the patient, while the second involved statistical processing of the collected data. The patient's clinical information included a history and complaints to determine the presence of characteristic symptoms of GORD. Statistical processing was used to identify the correlation between the patient's clinical parameters and the development of reflux disease. The statistical method was chosen based on its ability to accommodate all identified clinical indicators, convert them into digital codes while maintaining clinical significance, and calculate the level of probability of disease occurrence or non-occurrence with maximum accuracy. Subsequently, the collected clinical information was coded for further analysis. The type of coding was chosen based on the statistical method chosen – logistic regression, regression analysis. In case of the logistic regression method, the dependent variable was reflux disease, while the independent variables were clinical data from the survey and endoscopic examination.

Assignment of parameters to clinical data and their coding according to the binary numbering system was as follows: GORD (p) – according to the results of the calculations, it would acquire one of two values: 0 – the disease is absent, 1 – the disease is present; the presence of risk factors for the progression of GORD ($x_1, \ldots, x_n$) – 0 – absent, 1 – present. Considering these parameters, the development of

a mathematical model for predicting the development of GORD was based on the following formula:

$$ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n\,, \qquad (1)$$

where p – the probability of the emergence of GORD; x_1, ..., x_n – defined risk factors for the development of GORD; β_0 – a constant indicating the log odds of the development of reflux disease when all independent variables are equal to zero; $\beta_1\beta_n$ – coefficients reflecting the change in the log odds of developing reflux disease when the corresponding independent variable (x_1,..., x_n) changes by one, if, conditionally, all other indicators stay unchanged.

To determine the statistical adequacy of the developed mathematical model, the following indicators were considered: coefficient of determination (R-square), Akaike's information criterion (AIC), Bayesian information criterion (BIC), t-test, and Wald test. To determine the effectiveness of this model, the study employed validation methods – checking an independent sample of patients divided into a training and a test group, and cross-validation, which helped to determine the parameters of accuracy, sensitivity, specificity of the created model, and calculate the error of the developed model using ROC analysis.

The practical significance of this study was determined by analysing aspects of the clinical relevance of the developed model. These aspects included value for the patient, value for the healthcare industry, and value for the further development of the field.

RESULTS

The development of a mathematical model for predicting the progression of GORD was based on the relationship between real clinical data obtained during the interview and examination of the patients' gastrointestinal tract and the presence of GORD. Notably, it is crucial to define a set of indicators to create a model, as this will enable a more accurate prediction of the probability of the disease occurring and help to avoid false results. Therefore, when considering the degree of prognosticity of clinical data, it is necessary to determine the type of their connection with GORD (direct or indirect), to determine the specificity and to pay attention to other aspects that may affect the error in determining the prognosis.

The classic symptoms of GORD – heartburn, frequent belching, and regurgitation – are detected during the collection of complaints and anamnesis (heartburn, frequent belching) and endoscopic examination (regurgitation). Heartburn is directly related to the progression of GORD, as it can lead to an increase in acid production in the stomach and provoke inflammation of the mucous membrane, which will weaken the contraction of the muscles around the oesophagus, causing the opening of the floodgate through which the acidic contents of the stomach will return up into the oesophagus [9]. Regurgitation is the result of upper reflux, i.e., in the context of GORD, it is not the cause of its development, but a consequence [10]. Heartburn and regurgitation as indicators of GORD are highly sensitive, but not specific due to their association with other gastrointestinal problems: ulcerative colitis, intestinal dysfunction. Despite the high sensitivity of these symptoms, some patients with a confirmed diagnosis of GORD who complain of chest pain do not have symptoms of heartburn or regurgitation.

Apart from visual examination of the oesophageal mucosa, the examination of the stomach and duodenum is informative. Damage to the oesophageal mucosa (erosion, ulceration) is a direct consequence of GORD and, at the same time, endoscopic signs of this disease [11]. Damage to the gastric mucosa affects the development of GORD in several ways: by increasing the production of gastric acid; by changes in gastric motility; by the development of ulcers or gastritis, which disrupt its protective barrier, reducing its effectiveness in preventing acid reflux into the oesophagus [12]. There are several mechanisms of influence of duodenal mucosa lesions on the development of GORD. The first one is triggered by inflammation, which can lead to relaxation of the lower oesophageal sphincter, causing acid from the stomach to enter the oesophagus (Fig. 1), while the second one is caused by changes in the balance of stomach acids, which can affect the motor activity of the stomach and disrupt the normal digestive process [13].

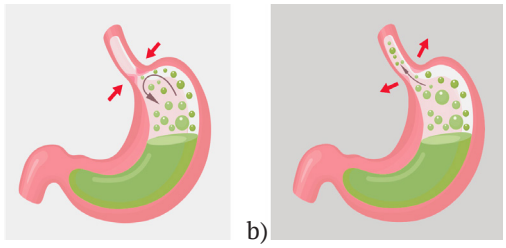

Figure 1. The mechanism of reflux during relaxation of the lower oesophageal sphincter
Notes: a) – sphincter is closed (normal); b) – sphincter is relaxed
Source: [14]

Thus, damage to the mucous membrane of the stomach and duodenum can be considered as a factor in the progression or exacerbation of GORD. Notably, gastroduodenitis, as an inflammatory process in the wall of the stomach and duodenum, can trigger several of the above mechanisms of gastric reflux, and its chronic form considerably increases the risk of developing GORD due to the prolonged irritating effect of bile on the oesophageal mucosa. Damage to the oesophageal mucosa is a more specific prognostic indicator of GORD than damage to the stomach and duodenum, but considering the presence of other factors that can lead to ulceration or erosion (e.g., thermal or chemical burns), the level of specificity cannot be considered sufficient to use it as the sole indicator of the probability of GORD progression.

Disorders of the gastrointestinal tract functioning are not specific to the progression of GORD, but in combination with the other clinical indicators mentioned above, they can more clearly predict the probability of progression or severity of this disease. Dysfunctions of the digestive system have several ways of influencing the progression of GORD. Gastritis and peptic ulcer disease can cause an increase in stomach acid production, which leads to its entry into the oesophagus [15]. Oesophageal sphincter dysfunction or achalasia causes a decrease in the tone of the sphincter, which is responsible for acid retention in the stomach [16]. Impaired peristalsis can lead to food

retention in the stomach, thus contributing to the development of GORD [17]. The direct impact of bile reflux on the progression of GORD is associated with the action of aggressive bile components that cause irritation and, with prolonged contact, inflammation of the oesophageal mucosa. The indirect effect is exerted through the release of bile into the stomach, which provokes an increase in the acidity of the gastric contents and damage to the barrier function of the lower oesophageal sphincter [18]. The presence of bile reflux may indicate certain problems of the gastrointestinal tract, and therefore, despite the low specificity of this indicator for the progression of GORD, its use as a prognostic factor is important, especially in combination with other clinical characteristics.

The effect of each type of bile on the development of GORD is related to its composition. Although the main components are bile acids, additional substances necessary for the normal digestive process may be included at different stages of the pathway. The principal components of bile produced by the liver are water, bile acids, proteins, fats, and electrolytes. When entering the biliary tract, mucus is added to these components, which provides additional binding of bile acids and protects the mucous membranes from erosion. In the gallbladder, bile is stored, concentrated, and fat-soluble substances are added to its composition for effective fat digestion [19]. Thus, when bile reflux occurs, the most aggressive irritation to the oesophageal mucosa is caused by bile that enters it from the gallbladder. Therefore, it can be assumed that in case of bile reflux, this type of bile is an additional indicator of GORD. Thus, to create a model for predicting the progression of GORD, indicators were selected that, in combination, allow the most accurate prediction of the probability of the disease. These indicators include heartburn, frequent belching, regurgitation, damage to the mucous membrane of the oesophagus, stomach, duodenum, chronic gastroduodenitis, gastrointestinal tract dysfunction, bile reflux (especially if bile enters the oesophagus from the gallbladder).

An important aspect that determines the reliability of a prognostic model is the choice of informative and accurate methods of patient examination. Apart from the symptom of heartburn, which is determined by the results of collecting complaints and patient history, other indicators of GORD progression are detected using instrumental imaging methods. The method of oesophagogastroduodenoscopy has shown high efficiency in the diagnosis of lesions of the mucous membranes of the gastrointestinal tract, which allows detecting endoscopic signs of GORD, including the type of damage (erosion, ulcer) and its volume. The informative and reliable nature of this method is ensured by a wide scope of its technical capabilities: the flexible design of the endoscopic tube allows the specialist to easily manipulate it during the passage through the internal organs; The optical system provides access to a clear image of the throat, oesophagus, stomach, and duodenum, which is displayed on the doctor's screen; the presence of channels for instruments allows for biopsies or other manipulations during the examination; the air supply function is used when it is necessary to expand the organs to facilitate the examination; thanks to the built-in video camera, images of the cavities of the examined organs in the format of photo or video materials can be saved for further analysis of the examination data [20, 21]. When analysing clinical trials on the factors of development, diagnosis, and treatment of GORD, it was found that oesophagogastroduodenoscopy was used in many of them as a method of detecting the key endoscopic signs of the disease [22-24]. An additional advantage of this method is to resolve the issue of the possibility of misinterpretation of the results of the endoscopic examination by the observer, depending on their experience or other subjective factors, which is the possibility of video recording the results of the procedure and their repeated analysis with the involvement of several specialists.

When creating a mathematical model for predicting the progression of GORD, the principle of logistic relationships between the patient's clinical parameters and the probability of the disease occurrence was used. The accuracy and validity of this model is achieved through statistical analysis methods to process real clinical data collected from the survey and video oesophagogastroduodenoscopy. This approach makes it possible to identify and consider clinical characteristics and pathological changes in the digestive tract that are essential for predicting the progression of GORD. The model for predicting the progression of GORD is based on obtaining clinical information from the patient, including the collection of complaints and patient history, which aims to establish the presence of the characteristic symptoms of GORD – heartburn and frequent belching, and endoscopic examination results to identify lesions of the mucous membrane of the oesophagus, stomach, duodenum (erosion, ulcers), and to determine the presence of sphincter dysfunction (regurgitation), and statistical processing of the data obtained.

To improve the understanding of the principle of developing a mathematical model for predicting GORD, it is worth considering an example using the logistic regression method, as its principle allows determining the correlation between the patient's clinical parameters and the development of this disease. In this case, according to Equation 1, the clinical data of the survey and endoscopic examination will acquire the values of independent variables x_1–x_9, while their coding according to the binary counting system will have the following form: the presence of heartburn complaints, frequent belching (x_1, x_2) – 0 – complaints absent, 1 – complaints present; lesions of the mucous membrane of the oesophagus, stomach, duodenum (x_3, x_4, x_5) – 0 – lesions absent, 1 – lesions present; sphincter disorders (regurgitation) (x_6) – 0 – disorders absent, 1 – disorders present; chronic gastroduodenitis (x_7) – 0 – disease absent, 1 – disease present; gastrointestinal tract disorders (x_8) – 0 – disorders absent, 1 – disorders present; bile reflux (x_9) – 0 – disorders absent, 1 – disorders present. Converting the Equation (1) in such a way as to calculate the probability of occurrence of the disease under study (p), the mathematical model for predicting the progression of GORD based on logistic regression will be as follows (2):

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_9 x_9)}}, \quad (2)$$

where p – the probability of GORD emergence; x_1 – the presence of heartburn complaints; x_2 – the presence of frequent belching complaints; x_3 – the presence of damage to the oesophageal mucosa; x_4 – the presence of damage to

the stomach mucosa; x_5 – the indicator of damage to the oesophageal mucosa; x_6 – the indicator of sphincter dysfunction (regurgitation); x_7 – the indicator of chronic gastroduodenitis; x_8 – the indicator of gastrointestinal tract dysfunction; x_9 – the indicator of bile reflux; β_0 – a constant indicating the log odds of developing reflux disease when all independent variables are zero; β_1-β_9 – coefficients reflecting the change in log odds of developing reflux disease when the corresponding independent variable changes (x_1-x_9) per unit, if, conditionally, all other indicators stay unchanged; e – the Euler number, approximately equal to 2.71828.

In this case, the graphical representation of the results of assessing the probability of GORD progression using the proposed mathematical model will be represented by a logistic curve (Fig. 2).

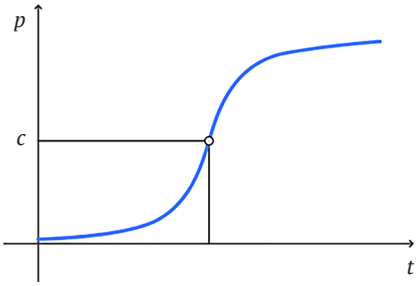

Figure 2. Graphical representation of the probability of GORD progression represented by a logistic curve
Notes: p is the probability of an event occurring (progression of GORD); c is the point after which the p function increases most rapidly (the probability of the disease progression increases); t is time (the study period)
Source: [25]

An increase in the value of the dependent variable means that an increase in the value of the corresponding independent variable is correlated with an increase in the chances of GORD progression. If, for instance, x_5 acquires a value of one, then the presence of damage to the mucous membrane of the oesophagus increases the risk of this disease. If the dependent variable decreases in response to an increase in the independent variable, it means that there is no association between a certain clinical indicator and the development of the disease. However, when developing the proposed mathematical model, clinical indicators were selected based on their direct or indirect association with the risk of GORD, and therefore the dependent variable is positively correlated with all independent variables. It is important to assess the overall adequacy of the developed mathematical model. For this, it is necessary to pay attention to the level of dependence of the dependent variable variation on the independent variables variation, which is determined based on the coefficient of determination (R-squared), the influence of out-of-sample variables on the model error – based on the AIC, the quality of the model – based on the balance between its accuracy and complexity BIC, and the statistical significance of independent variables – based on the t-test or Wald test.

It is vital to determine the accuracy and predictive ability of the mathematical model for predicting the progression of GORD by means of its validation. This stage is crucial, as the created model has a concrete practical significance and is aimed at performing a responsible task in clinical practice. Therefore, to determine its predictive capability, it is necessary to conduct the necessary tests under conditions that will help to critically assess the advantages and identify possible disadvantages. For this, it is recommended to conduct a clinical trial with an independent sample of patients, dividing them into training and test groups, or through cross-validation. Based on the test sample, one can estimate the accuracy, sensitivity, and specificity parameters of the created model and conduct a ROC analysis by calculating the area under the error curve. The key criterion for the development of this model was its suitability for use in clinical practice. By defining a set of prognostic indicators, establishing the optimised method for their detection and applying relevant statistical methods for processing clinical information, compliance with the set criterion was achieved and a mathematical model was created that allows identifying patients at high risk of GORD progression before it occurs or before the development of a more severe degree of severity (erosive form or Barrett's syndrome).

First of all, this model is useful for patients, as it can accommodate their individual clinical characteristics and provide results promptly and with high accuracy. This helps to expand access to the procedure, reduce diagnostic time, and determine the treatment strategy depending on the results. Thus, the use of a mathematical model to predict the progression of GORD can affect the effectiveness of therapy for patients in different groups as follows: patients at risk of GORD progression will be able to avoid the effects of the disease by following the doctor's recommendations regarding the adjustment of their diet and lifestyle; patients with a promptly diagnosed disease will be able to increase the chance of successful medical treatment, reduce the risk of complications, and get the opportunity to regularly monitor its course; patients with a diagnosed severe reflux disease will be able to receive an assessment of the risks and, according to the treatment plan developed according to them, undergo the necessary procedures, which may include surgical procedures. The developed mathematical method for predicting GORD is of medical and economic significance for the healthcare system. Based on the principles of evidence-based and personalised medicine, this method can increase the diagnostic potential of gastroenterology and improve public health by being used during routine or unscheduled examinations of patients with problems or even minor complaints related to the functioning of the digestive tract. At the same time, improving the diagnostic potential of the industry will help optimise the allocation of medical and economic resources, focusing on patients who need them most and avoiding the cost of unnecessary procedures for patients who are not at risk of developing reflux disease.

It is also essential to evaluate the role of the developed model for the further development and improvement of diagnostic systems in various healthcare sectors. The area of modelling systems for predicting the development or assessing the severity of various diseases or pathological conditions, apart from its serious prospects, poses a range of significant challenges to the medical industry. To implement and use such models correctly in the healthcare

system, the following tasks need to be addressed: technical – to ensure high quality data, legal – to protect patient confidentiality, organisational – to develop standards and protocols for the use and interpretation of modelling results, educational – to ensure proper training of healthcare professionals. At the same time, the process of mathematical processing of clinical data opens wide opportunities for machine learning and artificial intelligence, which, apart from predicting the risks of disease progression, will allow for diagnosis, patient classification, determining the most effective treatment strategy and monitoring the patient's condition throughout the entire therapy process. Considering the above prospects and opportunities, mathematical modelling using the most advanced information technology systems still cannot be considered as an alternative to the work of doctors. These technologies primarily serve as a tool used by a specialist to achieve results more accurately, reliably, and quickly. Therefore, the mathematical model for predicting the progression of GORD was developed to solve the problems of untimely detection of the disease and improve the overall epidemiological situation with gastroenterological diseases.

DISCUSSION

Analysing the findings obtained, the key aspects can be highlighted that were important in developing a mathematical model for predicting the progression of GORD. The medical component was critical for the result, which lied in selecting reliable prognostic factors among all clinical indicators that could be determined during the interview and physical, laboratory, and instrumental examinations. In studying the level of influence of these indicators, it was found that the close relationship between the digestive tract organs does not allow considering a certain clinical sign as the only indicator of predicting the progression of GORD due to the lack of sufficient specificity. Therefore, to improve the prognostic ability of the developed model, a set of clinical signs of the patient was formed, which had a direct or indirect impact on the development of reflux disease and included symptoms (heartburn, frequent belching, regurgitation), damage to the internal walls of organs (mucous membranes of the oesophagus, stomach, duodenum), gastrointestinal tract dysfunction (bile reflux and other disorders), and the presence of chronic gastroenterological diseases (chronic gastroduodenitis). M.E. Hossain *et al*. [26] described an approach to finding possible predictors of disease prognosis based on the use of electronic health data in a comprehensive literature review. The researchers investigated the significance of electronic data for use in predicting the development of diseases and found high accuracy of some modelling approaches developed based on such sources of medical information. The researchers also emphasised the viability and informativeness of the developed models, which were involved in the clinical process and showed promising results during operation. Considering the researchers' findings in the context of the present study, since the search for clinical indicators for building a model for predicting the progression of GORD was also carried out based on an analysis of materials on the relevant subject from previously published scientific sources, the effectiveness of the proposed approach can be confirmed.

When developing a model that contains several independent variables (clinical signs), attention should be paid to ways to optimise the model so that duration and resource consumption do not affect the overall benefit of its operation. It was possible to optimise the model in terms of procedure time and use of diagnostic equipment due to the choice of the method of instrumental examination – oesophagogastroduodenoscopy, as this procedure allows detecting all the listed clinical signs during one session using one device. The detection of heartburn and frequent burping symptoms occurs when collecting complaints and patient history during a consultation with a doctor and does not require additional equipment. H. Xu *et al*. [27] considered a comparable principle of optimising the model for diagnosing coronary heart disease. Based on a retrospective multicentre study, researchers identified eight significant risk factors, including demographic factors, clinical signs, and laboratory markers. The model was optimised by using standard research methods for determining cardiac disorders that do not require additional time or material resources. Generally, the approach of using standardised laboratory and instrumental research methods is most often used for optimisation in the disease progression prediction models [28-30].

A major step in creating the model was to choose a method of statistical processing of the collected data. The logistic regression used as an example demonstrated the possibility of quantifying the impact of each of the prognostic factors on the risk of GORD progression. This is because, despite its versatility, this method makes it possible to encrypt medical information in such a way that the correct results are obtained during the calculation. Thus, the probability of the disease progression is determined not only by the presence of a certain number of clinical indicators (independent variables), but also by the level of their prognostic significance (the coefficient reflecting the change in the log odds of developing reflux disease when the corresponding independent variable changes by one). Thus, this method makes it possible to identify the significance of prognostic indicators and assess the level of their interaction with the occurrence of the disease. Y. Deng *et al*. [31] employed a combination of machine learning and logistic regression to create their model to predict the risk of carotid plaques in adults with fatty liver disease. Based on the analysis of data from a population-based cross-sectional study, the researchers selected 5 independent prognostic factors for the risk of carotid plaques from 27 clinical parameters. The use of a combination of statistical methods and advanced information technology systems has revealed a higher predictive efficiency of the created model compared to the use of logistic regression and machine learning algorithms separately. In the context of the objectives of the present study, the evidence-based approach to the selection of predictors of prognosis and statistical processing of the results provided a prominent level of prognostic significance of the developed model, which makes it possible to apply it in clinical practice. However, to increase productivity, it is worth considering the possibility of using machine learning or artificial intelligence in combination with the statistical method, having worked out in advance the legal, organisational, and technical support for this process.

The final stage in the development of a predictive model is to assess its adequacy through validation. Since this model was created for further use in clinical practice, the proposed method of validation included testing the statistical significance and predictive ability on an independent sample of patients. E.W. Steyerberg & Y. Vergouwe [32] evidentially confirmed this approach, having investigated the problematic issues of models for clinical prognosis. The researchers proposed steps to improve them, including the development and validation process. The researchers also proved the significance of testing the prognostic model according to two criteria: adequacy assessment, which is to determine the level of accuracy of the statistical method of processing the results, and effectiveness, which is carried out directly in clinical settings with an independent sample of patients.

Generally, when analysing the development of mathematical models for predicting the development or severity of diseases, the main criteria and vector followed by developers are evidence and personalisation. That is why research is currently underway in many areas of medicine to revise established diagnostic procedures due to limited information content, difficulty of interpretation, risk or discomfort for the patient, long waiting times, or other characteristic signs of obsolescence. Thus, the expediency of replacing some radiography methods with magnetic resonance imaging or computed tomography, liver function tests with ultrasound diagnostics, computed tomography and endoscopic retrograde cholangiopancreatography, and the Mantoux test with photofluorography was covered [33-35]. In the present study, when choosing a method for collecting clinical information, attention was also paid to the informativeness, accuracy, and speed of obtaining clinical information. Therefore, among endoscopic methods, the choice was between oesophagogastroduodenoscopy and capsule endoscopy. The advantage of capsule endoscopy was the absence of discomfort for the patient, but the disadvantages included:

- the procedure duration from 8 to 10 hours, depending on the duration of the capsule's natural movement through the digestive tract;
- limited use for patients with intestinal obstruction, perforation of the digestive system, swallowing disorders, motor disorders of the oesophagus or stomach.

The disadvantage of oesophagogastroduodenoscopy was the relative discomfort for the patient associated with the specific mechanics of the procedure, while the advantages were informativeness, speed, and the possibility of using it in cases where capsule endoscopy is contraindicated. Despite the accuracy of the developed model, which is ensured by the use of a set of clinical indicators that have an established relationship with the development of GORD, certain questions persist due to the fact that the indices studied separately have a low level of specificity for predicting this disease. Summarising the analysis of the results, it can be concluded that mathematical modelling of the risk of reflux disease progression is a reliable tool in clinical gastroenterology for the prompt detection and effective implementation of measures for the prevention and treatment of this disease.

CONCLUSIONS

The creation of a mathematical model for predicting the progression of GORD was based on the identification of clinical indicators that can be used as prognostic factors, the substantiation of the choice of clinical research method for identifying these indicators, and the choice of the optimised method for processing clinical information.

A review of the relevant literature revealed links between certain clinical signs and the progression of GORD. The analysis of these relationships helped to identify prognostic factors for the development of the disease, including heartburn, frequent belching, regurgitation, damage to the mucous membrane of the oesophagus, stomach, duodenum, chronic gastroduodenitis, gastrointestinal dysfunction, and bile reflux. Oesophagogastroduodenoscopy proved to be the most informative method of detecting these risk factors, the technical characteristics of which allow for the informativeness and speed of obtaining the necessary clinical parameters. The logistic regression method was used to identify the correlation between the patient's clinical parameters and the development of reflux disease. Its principle was to encode clinical information for the mathematical calculation of the probability of a disease occurring. According to the criteria of the logistic equation, all clinical parameters were assigned the values of the corresponding variables: the dependent variable is the probability of GORD progression, independent variables are prognostic factors for the progression of GORD coded in binary.

Recommendations for model validation include assessment of the overall model adequacy – statistical significance, and validation – predictive ability determined based on a clinical trial with an independent sample. The developed prognostic model is of practical significance for patients, as it allows them to identify the risks of developing the disease in time and take relevant preventive or therapeutic measures to thwart the consequences of the disease; for the healthcare system, increasing the diagnostic potential of gastroenterology and optimising the allocation of labour and economic resources; for further development and improvement of the diagnostic area, using the capabilities of the most advanced information technology systems – machine learning and artificial intelligence. The practical significance of the developed model allows recommending its use in the clinical practice of gastroenterologists.

One of the limitations of the present study was the low level of specificity of the identified prognostic factors separately from each other, which may slightly affect the accuracy of the developed model. Therefore, the key vector for further research in this area may be the search for indicators with a higher level of specificity, e.g., laboratory or genetic markers of GORD.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

The authors of this study declare no conflict of interest.

REFERENCES

[1] Reva T, Reva V, Trefanenko I, Shumko H, Gaidichuk V. Variants of gastroesophageal reflux disease depending on the type of reflux. Ukr J Med Biol Sport. 2022;7(1):171–76. DOI: 10.26693/jmbs07.01.171

[2] Chue KM, Goh DWX, Chua CME, Toh BC, Ong LWL, Wong WK, et al. The Hill's classification is useful to predict the development of postoperative gastroesophageal reflux disease and erosive esophagitis after laparoscopic sleeve gastrectomy. J Gastrointest Surg. 2022;26(6):1162–70. DOI: 10.1007/s11605-022-05324-x
[3] Shinozaki S, Osawa H, Hayashi Y, Miura Y, Yano T, Lefor AK, Yamamoto H. Predictors and timing for the development of symptomatic gastroesophageal reflux disease after successful *Helicobactor pylori* eradication therapy. Scand J Gastroenterol. 2021;57(1):16–21. DOI: 10.1080/00365521.2021.1975310
[4] Wessels EM, Masclee GMC, Bastiaansen BAJ, Fockens P, Bredenoord AJ. Incidence and risk factors of reflux esophagitis after peroral endoscopic myotomy. Neurogastroenterol Motil. 2024 Apr 8;36(6):e14794. DOI: 10.1111/nmo.14794
[5] Armstrong D, Hungin AP, Kahrilas PJ, Sifrim D, Sinclair P, Vaezi MF, Sharma P. Knowledge gaps in the management of refractory reflux-like symptoms: Healthcare provider survey. Neurogastroenterol Motil. 2022;34(10):e14387. DOI: 10.1111/nmo.14387
[6] Lee SP, Kae SH, Jang HJ, Koh DH, Jung ES. Inter-observer variability of experts and trainees for the diagnosis of reflux esophagitis: Comparison of linked color imaging, blue laser imaging, and white light imaging. J Dig Dis. 2021;22(7):425–32. DOI: 10.1111/1751-2980.13023
[7] Myasoyedov S, Andreieshchev S. Morpho-functional monitoring of the esophagus status as the base for the dispenser observation in the patients with achalasia of cardia after surgical esophagocardiomyotomy. Clin Surg. 2021;88(9–10):16–20. DOI: 10.26779/2522-1396.2021.9-10.16
[8] Shevchenko B, Babii O, Prolom N, Titova M, Tarabarov S, Ushchina S. The role of endoscopic ultrasound in assessment of physiological cardia insufficiency during diagnosis of hiatal hernia. Gastroenterol. 2024;58(1):50–56. DOI: 10.22141/2308-2097.58.1.2024.587
[9] Savarino E, Marabotto E, Savarino V. Recent insights on functional heartburn and reflux hypersensitivity. Cur Opin Gastroenterol. 2022;38(4):417–22. DOI: 10.1097/mog.0000000000000846
[10] Di Mario F, Crafa P, Franzoni L, Tursi A, Brandimarte G, Russo M, et al. A comparison of different symptomatic reflux esophagitis treatments: A real-world study. Adv Clin Exp Med. 2023;32(9):1075–80. DOI: 10.17219/acem/171001
[11] HajHosseini P, Takalloozadeh M. An isotropic hyperelastic model of esophagus tissue layers along with three-dimensional simulation of esophageal peristaltic behavior. J Bioeng Res. 2019;1(2):12–27. DOI: 10.22034/JBR.2019.189018.1009
[12] Miftahussurur M, Waskito LA, Syam AF, Nusi IA, Wibawa IDN, Rezkitha YAA, et al. Serum pepsinogen level as a biomarker for atrophy, reflux esophagitis, and gastric cancer screening in Indonesia. J Res Med Sci. 2022;27(1):90. DOI: 10.4103/jrms.jrms_983_21
[13] Wang M, Lou E, Xue Z. The role of bile acid in intestinal metaplasia. Front Physiol. 2023;14:1115250. DOI: 10.3389/fphys.2023.1115250
[14] Causes and symptoms of esophagitis [Internet]. [cited 2023 Dec 1]. Available from: https://medikom.ua/simptomy-ehzofagita/
[15] Yalaki S, Pulat H, Ilhan A. Localization of *Helicobacter pylori* gastritis and the relation of existing histopathological features with reflux esophagitis. Scand J Gastroenterol. 2020;55(1):27–33. DOI: 10.1080/00365521.2019.1701071
[16] Dunn CP, Wu J, Gallagher SP, Putnam LR, Bildzukewicz NA, Lipham JC. Understanding the GERD barrier. J Clin Gastroenterol. 2021;55(6):459–68. DOI: 10.1097/mcg.0000000000001547
[17] Hu ZW, Wang ZG, Wu JM, Tian SR, Zhang Y, Zhan XL, et al. Relationship between the severity of reflux esophagitis and the esophageal motility features on high resolution manometry. Chin Med J. 2017;97(42):3306–11. DOI: 10.3760/cma.j.issn.0376-2491.2017.42.007
[18] Chen L, Zhu G, She L, Ding Y, Yang C, Zhu F. Analysis of risk factors and establishment of a prediction model for endoscopic primary bile reflux: A single-center retrospective study. Front Med. 2021;8:758771. DOI: 10.3389/fmed.2021.758771
[19] Qu X, Shi Y. Bile reflux and bile acids in the progression of gastric intestinal metaplasia. Chin Med J. 2022;135(14):1664–72. DOI: 10.1097/CM9.0000000000002290
[20] Cohen J, Safdi MA, Deal SE, Baron TH, Chak A, Hoffman B, et al. Quality indicators for esophagogastroduodenoscopy. Am J Gastroenterol. 2006;101(4):886–91. DOI: 10.1111/j.1572-0241.2006.00676.x
[21] Barret M, Chaussade S, Boustière C, Canard JM, Schott AM, Ponchon T, et al. Diagnostic yield of esophagogastroduodenoscopy in France. Clin Res Hepatol Gastroenterol. 2021;45(4):101540. DOI: 10.1016/j.clinre.2020.08.015
[22] Martín-Marcos I, Fernández-Morte N, Balsategui-Martín M, Ortiz-Cantero A, Bermúdez-Ampudia C, López-Picado A, et al. Evaluation of pharyngeal lidocaine anesthesia for esophagogastroduodenoscopy: Double-blind randomized control trial. Dig Endosc. 2021;34(4):808–15. DOI: 10.1111/den.14168
[23] Suda T, Shirota Y, Takimoto H, Tsukada Y, Takishita K, Nadamura T, et al. Image quality of abdominal ultrasonography after esophagogastroduodenoscopy is preserved by using carbon dioxide insufflation: A non-inferiority test in the same subject. PLoS One. 2022;17(9):e0275257. DOI: 10.1371/journal.pone.0275257
[24] Bilder HG, Soccini C, Lasa JS, Zubiaurre I. Impact of time to esophagogastroduodenoscopy in patients with nonvariceal upper gastrointestinal bleeding: A systematic review and meta-analysis. J Gastroenterol Mexico. 2022;87(3):320–29. DOI: 10.1016/j.rgmxen.2021.11.010
[25] What are logistic models used for? [Internet]. [cited 2023 Dec 1]. Available from: https://www.houseofmath.com/uk/encyclopedia/funktsiyi/matematychni-modeli/typy-modeley/dlya-choho-vykorystovuyutsya-lohistychni-modeli

[26] Hossain ME, Khan A, Moni MA, Uddin S. Use of electronic health data for disease prediction: A comprehensive literature review. IEEE/ACM Trans Comput Biol Bioinform. 2021;18(2):745–58. DOI: 10.1109/tcbb.2019.2937862
[27] Xu H, Duan Z, Miao C, Geng S, Jin Y. Development of a diagnosis model for coronary artery disease. Ind Heart J. 2017;69(5):634–39. DOI: 10.1016/j.ihj.2017.02.022
[28] Cui Z, Liu JX, Gao YL, Zhu R, Yuan SS. LncRNA-disease associations prediction using bipartite local model with nearest profile-based association inferring. IEEE J Biomed Health Inform. 2020;24(5):1519–27. DOI: 10.1109/jbhi.2019.2937827
[29] Holcomb KM, Staples JE, Nett RJ, Beard CB, Petersen LR, Benjamin SG, et al. Multi-model prediction of West Nile virus neuroinvasive disease with machine learning for identification of important regional climatic drivers. GeoHealth. 2023;7(11):e2023GH000906. DOI: 10.1029/2023GH000906
[30] Hao M, Huang X, Liu X, Fang X, Li H, Lv L, et al. Novel model predicts diastolic cardiac dysfunction in type 2 diabetes. Ann Med. 2023;55(1):766–77. DOI: 10.1080/07853890.2023.2180154
[31] Deng Y, Ma Y, Fu J, Wang X, Yu C, Lv J, et al. Combinatorial use of machine learning and logistic regression for predicting carotid plaque risk among 5.4 million adults with fatty liver disease receiving health check-ups: Population-based cross-sectional study. JMIR Public Health Surveill. 2023;9:e47095. DOI: 10.2196/47095
[32] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925–31. DOI: 10.1093/eurheartj/ehu207
[33] Ahlawat S. Current state-of-the-art imaging techniques in the domain of whole-body MRI and its advantages over other whole-body PET in different musculoskeletal diseases. Eur Radiol. 2023;33(12):8573–75. DOI: 10.1007/s00330-023-09883-8
[34] Rastogi A. Changing role of histopathology in the diagnosis and management of hepatocellular carcinoma. World J Gastroenterol. 2018;24(35):4000–13. DOI: 10.3748/wjg.v24.i35.4000
[35] Tandon A, Tandon S. Fallacies of mantoux in the diagnosis of latent tuberculosis. J Assoc Physic India. 2022;70(12):94–95.

Розробка та валідація математичної моделі для прогнозування розвитку рефлюкс-езофагіту на основі даних відеоезофагогастродуоденоскопії

Олександр Галушко

Аспірант
Центральноукраїнський державний університет імені Володимира Винниченка
25000, вул. Шевченка, 1, м. Кропивницький, Україна
https://orcid.org/0009-0009-0275-6542

Юрій Гуртовий

Кандидат фізико-математичних наук, доцент
Центральноукраїнський державний університет імені Володимира Винниченка
25000, вул. Шевченка, 1, м. Кропивницький, Україна
https://orcid.org/0000-0002-1499-7089

Анотація. Дослідження мало на меті визначити комплекс прогностичних факторів розвитку рефлюкс-езофагіту для використання їх в розробці математичної моделі прогнозування даного захворювання на основі результатів відеоезофагогастродуоденоскопії. Під час його проведення було встановлено низку прогностичних факторів рефлюкс-езофагіту та задіяно статистичний метод для встановлення рівня їх кореляції із розвитком хвороби. За результатами дослідження було виявлено зв'язок між певними клінічними показниками та виникненням рефлюкс-езофагіту, на основі чого сформовано комплекс прогностичних факторів розвитку гастроезофагеальної рефлюксної хвороби, який включав: печію, часті відрижки, регургітацію, ушкодження слизової оболонки стравоходу, шлунку, дванадцятипалої кишки, наявність хронічного гастродуоденіту, порушення функціонування шлунково-кишкового тракту, жовчний рефлюкс. В створенні математичної моделі прогнозування було використано метод логістичної регресії, який застосовувався для виявлення кореляції між клінічними показниками пацієнта й фактом виникнення рефлюксної хвороби та визначення ймовірності її розвитку. Для приведення клінічної інформації у відповідність статистичній формулі їй надавалось значення незалежних змінних, а наявність чи відсутність певного показника кодувалось за двійковою системою числення. Для перевірки розробленої моделі приводились рекомендації із оцінки статистичної значущості незалежних змінних для визначення її адекватності, та визначення предиктивної здатності шляхом випробування на незалежній вибірці пацієнтів. Розроблена прогностична модель має важливе практичне значення для пацієнтів, галузі охорони здоров'я та подальшого розвитку напрямку, так як дозволяє вчасно виявити захворювання і вжити відповідних заходів з профілактики та лікування, підвищити діагностичний потенціал галузі, оптимізувати розподіл медичних ресурсів та задіювати можливості машинного навчання й штучного інтелекту на основі існуючої моделі

Ключові слова: печія; регургітація; слизова оболонка стравоходу; шлунково-кишковий тракт; жовч; логістична регресія