



Kiselev K.V.<sup>1</sup>, Potekhina A.V.<sup>2</sup>, Osyayeva M.K.<sup>2</sup>, Noeva E.A.<sup>2</sup>, Vyborov O.N.<sup>2</sup>, Zorin A.V.<sup>2</sup>,  
Shvyrev S.L.<sup>1</sup>, Martynyuk T.V.<sup>2</sup>, Zarubina T.V.<sup>1</sup>, Chazova I.Ye.<sup>2</sup>

## DEVELOPMENT OF CONCEPTS NOMENCLATURE FOR CLINICAL DECISION SUPPORT SYSTEM IN DIAGNOSTICS OF ANGINA PECTORIS

<sup>1</sup>*Pirogov Russian National Research Medical University,*

<sup>2</sup>*Russian Cardiology Research and Production Complex,  
Moscow, 121552, Russian Federation*

### SUMMARY

**Introduction.** Since 2000<sup>th</sup> CHD morbidity in Russia is growing constantly, especially, angina pectoris morbidity. Increasing workload on physicians due to high number of patients makes us find the ways to optimize human resource applying. One of the possible ways is development of clinical decision support systems, particularly, ontology-based systems. These ontologies must adopt semantic interoperability for the concepts included, to be clearly interpreted by physicians of different medical schools.

**Research target.** Develop the concepts nomenclature for building of angina pectoris ontology.

**Methods.** For primary concepts extraction we used state clinical recommendations “Stable coronary heart disease” issued in 2016. Development and filling of the nomenclature were performed in MS Excel and MS Excel Online. The work through the primarily extracted concepts was performed with expert-cardiologists by cross-validation method and common voting on final inclusion of the concept in the nomenclature. The final nomenclature validation was performed by semi-automatic method on anonymized 610 electronic health records of the patients.

**Results.** After primary extraction of the concepts without experts, we could extract 336 basic concepts and 144 synonyms, keeping in mind the aspects of this nomenclature development.

The shared work through this nomenclature with the expert-cardiologists let us increase the number of the basic concepts up to 409 and the number of the synonyms up to 619. The nomenclature validation on electronic health records showed compliance of the concepts in aspects of diagnostics methods and medicine prescriptions and 92% compliance in diagnostic signs, which doctors evaluate in diagnostics results. Percentages of primarily extracted concepts in draft nomenclature and after being worked through by the experts and also validation results based on electronic health records, let us make a conclusion that clinical recommendations can be used as a basic data source for concepts extraction, but the draft result must be evaluated by the experts and also filled up with the usage of additional knowledge sources about angina pectoris, and after that the nomenclature can be used for angina pectoris ontology development.

**Conclusion.** The usage of clinical recommendations is proved to be the basic data source for concepts of subject area extraction. However, the draft nomenclature must be evaluated and filled up with additional synonyms and validated on electronic health records. After that, the extracted concepts and their synonyms can be used for ontology development.

**Keywords:** *angina, clinical decision support systems, CDSS, ontology, clinical guidelines.*

✉ [kirillkiselev@pm.me](mailto:kirillkiselev@pm.me)

**For citation:** Киселев К.В., Потехина А.В., Осяева М.К., Ноева Е.А., Выборов О.Н., Зорин А.В., Швырев С.Л., Мартынюк Т.В., Зарубина Т.В., Чазова И.Е. Разработка номенклатуры понятий для системы поддержки принятия врачебных решений в области диагностики стенокардии I-IV функциональных классов. Евразийский кардиологический журнал. 2018, Сентябрь 20; 3:14-19 [Trans. into Eng. ed.: Kiselev K.V., Potekhina A.V., Osyayeva M.K., Noeva E.A., Vyborov O.N., Zorin A.V., Shvyrev S.L., Martynyuk T.V., Zarubina T.V., Chazova I.Ye. Development of concepts nomenclature for clinical decision support system in diagnostics of angina pectoris. Eurasian cardiological journal. 2018, September 20; 3:20-24]

## Information about authors:

<b>Alexandra Potekhina</b>	Ph. D., Junior researcher of the Department of pulmonary hypertension and heart diseases of the Russian Cardiology Research and Production Complex
<b>Maria Osyayeva</b>	laboratory researcher of the Department of pulmonary hypertension and heart diseases of the Russian Cardiology Research and Production Complex
<b>Elena Noeva</b>	Ph. D., senior researcher of the Department of pulmonary hypertension and heart diseases of the Russian Cardiology Research and Production Complex
<b>Oleg Vyborov</b>	Ph. D., researcher of the Department of pulmonary hypertension and heart diseases of the Russian Cardiology Research and Production Complex
<b>Andrey Zorin</b>	Ph. D., researcher of the Department of pulmonary hypertension and heart diseases of the Russian Cardiology Research and Production Complex
<b>Sergey Shvyrev</b>	Ph. D., associate professor of medical cybernetics and informatics department of the Pirogov Russian National Research Medical University
<b>Tamila Martynyuk</b>	MD, head of the Department of pulmonary hypertension and heart diseases of the Russian Cardiology Research and Production Complex
<b>Irina Chazova</b>	academician of Russian Academy of Sciences, M.D., professor, director of Russian Cardiology Research and Production Complex, 3rd Cherepkovskaya st., 15a, Moscow, 121552, Russian Federation
<b>Tatiana Zarubina</b>	M.D., professor, deputy director on health informatics of Federal research institute for health organization and informatics of ministry of health of Russian Federation, head of the Department of medical Cybernetics and Informatics of the Pirogov Russian National Research Medical University
<b>Corresponding author: Kirill Kiselev</b>	Ph. D. student of the Department of medical Cybernetics and Informatics of the Pirogov Russian National Research Medical University, tel.+7-964-515-68-86; e-mail: kirillkiselev@pm.me, Ostrovityanova st., 1, Moscow, 117997, Russian Federation

### Abbreviations

CSD – circulatory system diseases  
CHD – coronary heart disease  
CG – clinical guidelines  
AP – angina pectoris

### Definitions

Knowledge base – information object containing medical data and its interpretation

Validation – act of model data compliance approve with real-world data

Ontology – formalized description of subject area, containing information about formalized concepts connections

Production model – set of conditional transitions “if-then”, which let to make transitions from one medical concept to another when making diagnostics and explaining diagnostic and treatment procedures appointment

CDSS – computer programs used by doctors to get advice when making clinical decisions

### INTRODUCTION

In the structure of morbidity of the Russian population, the share of diseases of the circulatory system (CSD) has been growing steadily since 2000 and in 2016 amounted to 4 649 thousand number of people or 31.7 per 1000 population [1]. One of the main diseases in the structure of CSD is coronary heart disease (CHD), including angina pectoris (AP). The increasing number of patients from year to year means a growing burden on therapists and cardiologists who are directly involved in AP diagnosis and treatment, which can lead to a decrease in the total amount of time

spent on one patient and a drop in the quality of care. One way to improve this situation is to use clinical decision support systems (CDSS), which are designed to advise the doctor and are designed to facilitate the diagnosis and treatment of a given pathology [2].

As a knowledge base of the CDSS, on the basis of which the program offers the doctor possible solutions, in recent years have increasingly been used not simple production models, but ontology of selected clinical area. In particular, a group of scientists confirmed the effectiveness of this approach to build a knowledge base of CDSS on the example of the development of such a system for the diagnosis and antibiotic therapy of infectious diseases [3].

In the context of high AP morbidity and modern approaches to support clinical decision-making, we have set a goal to develop a CDSS for the diagnosis and treatment of angina based on AP ontology. But, in order to ensure the semantic interoperability of the developed ontology, it was necessary to carry out work on the primary content of the ontological model and taking into account the possible options for naming those concepts that are key to it. In this regard, we have set the following goals and objectives.

### GOALS AND OBJECTIVES

Goal: to develop a nomenclature of concepts for building an ontological model of angina

Objectives:

1. Identify the main source of knowledge about the AP for the primary allocation of concepts.
2. To make an initial selection of concepts taking into account possible synonyms.
3. Clarify the resulting nomenclature with the help of experts.
4. To draw a conclusion about the possibility of using the resulting nomenclature as the basis for filling the ontology of angina.

## MATERIAL AND METHODS

As the main source of knowledge about AP, we used the Federal clinical guidelines (CG) "Stable coronary heart disease" from 2016 [4]. As a definition of "concept" the following was used: it is a set of features that are all together sufficient and each separately necessary to separate this class of objects from the other [5]. on this definition the rules of the concept separation from the CG were based.

The concepts extraction and maintenance of the nomenclature were carried out using the program MS Excel and MS Excel Online. The team of experts-cardiologists, with the help of which the specification of the nomenclature was carried out, included 6 people representing the Russian cardiology research and production center. The method of verification and addition of the primary concepts by the experts was to assign to a specific expert a set of values from the primary nomenclature, which the expert checked and supplemented with the options of names, if necessary. At general meetings, individual work was checked by cross-validation, after which the final decision was made to introduce the concept and its synonyms into the developed nomenclature by general voting. If it was necessary to add new concepts to the nomenclature, which were not selected initially, experts made a corresponding proposal with options for the names of the concepts, followed by a standard procedure of cross-validation and decision on the inclusion of a new item in the nomenclature.

To check the validity of the nomenclature and to make a decision on the possibility of its use as a source of filling the AP ontology in semi-automatic mode, the authors compared the selected concepts and synonyms with the concepts used in the description of clinical cases of 610 patients treated in the second medical Department of the Russian Cardiology Research and Production Complex in 2017. Electronic health records were depersonalized and translated into MS Word format for further processing. The patient model for the selection of medical records included:

- Gender – male and female
- Age – from 35 years
- ICD code of the underlying disease – I20.8, I20.9, I20.1
- The group of comorbidities included: type 2 diabetes, hypertension, myocardial infarction.

## RESULTS

After analyzing the possibility of using different types of literature as primary content: textbooks, monographs, articles, data of electronic health records of patients, - we, realizing that the choice should be determined not only by semantic completeness, but also considering synonymy, chose the Federal clinical recommendations for stable coronary heart disease from 2016. In this case, semantic completeness should be provided by the requirements for the content of the CG (Annex to the letter to the chief specialists from the first Deputy Minister of health of the Russian Federation №17-4/10/1-4939 from 1 September 2016). And the solution to the problem of synonymy is the elaboration of the concepts used in the text of the CG by the community of authors-developers of clinical guidelines, which are experts of the Russian cardiological society, the National society for the study of atherosclerosis and the National society for atherothrombosis. Thus, the CG is currently the most "legitimate" source of information for filling the nomenclature of concepts and development of AP ontology, despite the fact that the nature of their use by doctors in clinical practice remains recommendatory and informational.

Clinical guidelines for stable angina contain both sections aimed at reporting general information about the disease (etiology, risk

factors, pathogenesis), and sections that determine diagnostic measures and therapeutic purposes. In connection with this feature of filling the CG, we had to develop a system of rules by which a particular concept falls into the range. Considering the definition of "concept", as well as the development of ontology, we have adopted the following principles:

1. The nomenclature includes only concepts related to the clinical component of the AP description and are not related to the description of auxiliary and administrative reporting aspects. For example, "myocardial infarction", "hypercholesterolemia", "coronary angiography" – are included in the nomenclature. "ICD 10", "Diagnostics", "level of persuasiveness of recommendations" – are not included in the nomenclature;

2. The nomenclature includes the concepts from the sections of the CG describing instrumental diagnostics, laboratory diagnostics and treatment. Moreover, the names of drugs are not allocated to the nomenclature, but only the names of groups of these drugs. The restriction on the sections is determined by the scope of the CDSS – support for decision-making of the clinician at the stage of clarifying the primary diagnostic hypothesis using additional methods of instrumental and laboratory diagnosis, as well as the appointment of treatment. It means that the sections describing complaints, pathogenesis, etiology, risk factors, physical examination were not worked out at the primary stages of the concept identification;

3. The concepts are distinguished on the principles of the absence of background knowledge in cardiology: if a specialist who distinguishes concepts knows about synonyms for this concept from other sources of information, he does not include them in the nomenclature. The nomenclature gets only the information, the source of which are the CG;

4. On the basis of the previous paragraph, another rule was formulated: if there is no explicit indication in the text of the CG that the next concept is synonymous with an already selected concept, then this concept is included in the nomenclature as a separate one;

5. As synonyms of the concepts are considered not only linguistic synonyms, for example, "angina pectoris" and "thoracic toad", but also the generally accepted versions of the concepts, for example: "coronary heart disease" and "CHD". In this case, the full name without abbreviations was used as the basic concept, and all abbreviations were considered synonyms.

Thus, from the text CG has been extracted 336 root concepts and 144 synonymous to them. Highlighted concepts were made in MS Excel, where was made a markup item, which consists of three attributes:

1. Concept ID
2. Concept
3. Concept synonym

The resulting file was used for further development of the nomenclature in cooperation with experts-cardiologists.

To do this, the file was placed in the cloud service MS Excel Online, to provide parallel access to the editing of the required values of the nomenclature of experts and avoid the need for information merge of several files in one, if the editing was carried out by each expert in offline mode.

The structure of the nomenclature has also undergone changes. To ensure control over the changes of certain items, additional attributes were introduced, after which the structure of the nomenclature acquired the following form:

1. Concept ID
2. Concept
3. Concept synonym

**Table 1. Example of filling in the nomenclature of synonyms and concepts for the ontological model of angina**

Concept ID	Concept	Concept synonym	Expert commentary	Concept/synonym status	Status date change	Expert name	Concept/synonym approved mark
187	Atherosclerosis			Reviewed	15.05.2017	Vyborov	Approved
188		Atherosclerotic plaque		Added	28.05.2017	Vyborov	Approved
189		Atherosclerotic lesion		Added	28.05.2017	Vyborov	Approved
190		ATHP		Reviewed	28.05.2017	Vyborov	Approved
191	Atherosclerotic plaque		Synonym of previous root concept	Deleted	28.05.2017	Vyborov	Not approved

4. Expert commentary
5. Concept/synonym status
6. Status date change
7. Expert name
8. Concept/synonym approved mark

In the "Expert commentary" column, doctors provided guidance to the expert on the extraction of concepts regarding the modification/deletion/addition of a concept or its synonym with an explanation of the reason. Also, in this field was made a note if the concept or synonym was required to consider at the general meeting, because the expert doubted the need to include this item in the final nomenclature. The "concept/synonym status" attribute specified one of the following values:

1. Added – new item in the nomenclature. Added by an expert based on his own experience in the need to use this concept or synonym in clinical practice;
2. Changed – the concept with the given status had spelling changes, or has been moved from its synonyms in the root of the concept and vice versa;
3. Reviewed – the expert has reviewed this concept or synonym and has no additional edits or suggestions;
4. Deleted – the expert considered that this concept or synonym is redundant in the description of the AP ontology. The concept or synonym refers mostly to related pathology, or is not used in clinical practice. Also, this status was marked by duplicates of root concepts or synonyms

Date of the status change is recorded to keep track of time when nomenclature items were changed. In the column "Expert name" mentioned executive specialist, who worked out this concept or synonym. "Concept/synonym approved mark" had two possible text values: "Approved" if the concept was cross - validated and included in the final version of the nomenclature, and "Not approved" if the concept was not included in the final nomenclature. For concepts for which cross-validation was not performed, this field was left blank. An example of filling the nomenclature is presented in Table 1.

The summary of nomenclature working out with the experts is shown in the Table 2.

Comparison of these indicators with the primary content of the nomenclature shows that the number of root concepts that were included in the final nomenclature increased from 336 to 401 (42 deleted concepts are not considered from the total number in 443 concepts), but the number of synonyms of these concepts increased much more: from 144 to 619 (the number of synonyms without deleted positions). The percentages for root concepts indicate that the primary work of the CG to extract information on them allows to build a solid primary framework for the ontology being developed. But the percentage of synonyms indicates that CG do not possess sufficient knowledge concerning the distribution

of the root concepts of the possible spellings, and names. After filling in the nomenclature with primary information from the CG, additional study of the information obtained with the help of experts and involvement of additional literary sources of knowledge about the AP is required.

After working through the nomenclature of concepts with experts, we started to validate the selected concepts and their synonyms on the basis of their verification in semi-automatic mode with the concepts used in the electronic health records of 610 patients. The results of the verification showed that the nomenclature concepts, meaning instrumental studies, 100% cover the needs of doctors in the description of appointments to patients; the concepts of diagnostic signs that are isolated from the patient by means of additional instrumental diagnostics have coincided by 92% - the discrepancies are caused by a large number of variable indicators in such studies as echocardiography, coronary angiography; the concepts used to describe laboratory studies have coincided by 100%. No discrepancies in the concepts of nomenclature and medical records were found in the groups of prescribed drugs. Thus, the developed nomenclature, considering the refinement of the mismatched indicators of diagnostic studies, 100% corresponded to the concepts used by doctors in the electronic health records of patients. This fact allows us to make a conclusion about the semantic completeness of the developed nomenclature and the possibility of its use as a source of content of the AP ontology.

## SUMMARY

The increasing number of cases of coronary heart disease and angina makes us look for ways to use medical resources more effectively. One such way is to develop specialized systems to

**Table 2. Quantitative indicators of the result of working out the primary filling of the nomenclature of concepts by experts**

Parameter	Value	% of all
Root concepts	443	100.0
In particular:		
Unchanged	290	65.5
Changed	45	10.2
Added	66	14.9
Deleted	42	9.5
Synonyms	665	100.0
In particular:		
Unchanged	122	18.3
Changed	13	2.0
Added	484	72.8
Deleted	46	6.9

support clinical decision-making. Recently, special attention has been paid to CDSS, which operate on ontologies of selected clinical areas. To create such an ontology, it is necessary to provide the possibility of semantic interoperability of concepts that fill the ontology and are used in the modules of explaining the proposed solutions to the doctor.

Semantic interoperability is possible due to the development of a specialized nomenclature of concepts about AP considering the synonyms of the selected concepts.

As a primary source of filling the nomenclature of concepts, it is possible to use federal clinical guidelines for stable coronary heart disease, as the CG is the most complete official guide about this pathology.

Primary work out of CG has allowed to create the nomenclature with 336 root concepts and 144 synonyms. However, further elaboration of this list with experts-cardiologists allowed to identify 66 additional root concepts and significantly expand the number of synonyms to 619.

This indicates that even a source of information about the AP, such as federal clinical guidelines, does not have a comprehensive coverage of knowledge on a given pathology. The CG has proved to be a good source of knowledge for the primary filling of the nomenclature of concepts. But what is obtained in the initial study of the issue, cannot be immediately used as the filling of the framework of the ontology of the AP, the selected concepts require comprehensive study in joint with the experts and the involvement of additional sources of literary knowledge about this pathology.

The validation of the developed nomenclature made it possible to understand that the differences in the concepts used by doctors in electronic health records are insignificant and relate only to the variable parameters of some diagnostic procedures. After completion of the nomenclature, it corresponded to the concepts used by doctors by 100%, which allowed to use it to develop an ontological model of angina.

It is particularly worth noting the importance of the result for the medical community both in practical application and in scientific way. Synonymy is a significant problem in the description of clinical information. Doctors can use different terminology in the formation of diagnostic conclusions, medical recommendations, interpretation of certain clinical indicators, use more or less commonly used abbreviations of terms, introduce neologisms.

Emerging verbal confusion may mislead doctors with allied specialties, experts who assess the quality of care, developers of health information systems who must provide medical information for the doctors in proper way. The creation of nomenclatures of concepts of the subject area considering the possible variations of the names of the selected concepts allows to solve the problem of synonymy and significantly reduce the cases of misinterpretation of terms by doctors and other specialists with different levels of training. In addition, the nomenclature of concepts, which is the basis of the CDSS, will allow the clinical decision support systems to function considering possible variations in the use of concepts in the performance of the doctor's work on the formation of diagnostic hypotheses, appointments of additional diagnostics and treatment, etc. This makes possible the prospects of fully automatic analysis of data from electronic health records, medical assistance in the management of electronic health records, as well as the search for new hidden relationships between different clinical phenomena, which could escape the attention of researchers, which were described in similar words, for which synonymous connections were not identified. The result of full compliance of the developed nomenclature with the concepts used in the test electronic health

records, demonstrated in this work, confirms that such a solution to the problem of synonymy, allowing doctors to continue to operate with the terms familiar to them, has sufficient grounds and can be expanded to create nomenclatures of concepts in other areas of clinical knowledge.

## CONCLUSIONS

1. As the main source of knowledge about AP, it is possible to use federal clinical guidelines on stable coronary heart disease. This source is an official publication of the Ministry of health, meeting the requirements for filling and considering the solution of the problem of synonyms;

2. The study of clinical guidelines allowed to identify 336 root concepts and 144 synonyms, which allowed to create the basis of the developed nomenclature and to begin joint revision of the list of concepts with experts;

3. The study of the primary nomenclature with experts allowed to significantly expand the number of synonyms of root concepts, and the percentage ratio of the number of primary items of the nomenclature and added by experts showed that clinical guidelines have a sufficient set of information about the AP to create a framework of the developed ontology in terms of root concepts;

4. The resulting nomenclature, based on the extraction of concepts from the federal clinical guidelines, concerning the revision by experts and additional changes after validation using electronic health records of patients, can be used to fill the ontology of angina and subsequent use in the clinical decision support systems.

## REFERENCES

1. *Federal statistics service. Russian statistical yearbook [Internet]. 2017 [cited 2018 Mar 31]. Available from: [http://www.gks.ru/bgd/regl/b17\\_13/IssWWW.exe/Stg/08-12-23.doc](http://www.gks.ru/bgd/regl/b17_13/IssWWW.exe/Stg/08-12-23.doc). [In Russian]*
2. Gusev A.V., Zarubina T.V. *Clinical decision support in medical information systems. Physician and information technologies. 2017;2:60–72. [In Russian]*
3. Shen Y, Yuan K, Chen D, Colloc J, Yang M, Li Y, Lei K. *An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. Artif Intell Med. 2018 r.;86:20–32.*
4. *Ministry of health of Russian Federation/ Stable coronary heart disease: clinical guidelines [Internet]. 2016 [cited 2018 Apr 3]. Available from: <http://cr.rosminzdrav.ru/#!/recomend/133>. [In Russian]*
5. Stozhok EV. *Concept, semantics and meaning. ONV [Internet]. 2011 [cited 2018 Apr 3];(95). Available from: <https://cyberleninka.ru/article/n/termin-ponyatie-i-znachenie>. [In Russian]*

Accepted for publication: 27.07.2018