

Л. Ю. КАРАХАЛИС<sup>1</sup>, Н. С. ПАРОВА<sup>1</sup>, А. А. ХАЛАФЯН<sup>2</sup>, В. А. АКИНЬШИНА<sup>2</sup>

## ПРОГНОЗИРОВАНИЕ СТАДИИ ЗАБОЛЕВАНИЯ У БОЛЬНЫХ АДЕНОМИОЗОМ ПРИ ПОМОЩИ ДЕРЕВЬЕВ КЛАССИФИКАЦИИ

<sup>1</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный медицинский университет» Министерства здравоохранения Российской Федерации, ул. Седина, 4, Краснодар, Россия, 350063

<sup>2</sup>Факультет компьютерных технологий и прикладной математики ФГБОУ ВО КГУ, ул. Ставропольская, 149, г. Краснодар, Россия, 350040.

### АННОТАЦИЯ

**Цель.** Разработать математическую модель прогнозирования стадий распространения аденомиоза по клиническим результатам обследования больных статистическим методом *деревья классификации*.

**Материалы и методы.** Проведена выборка из 84 больных с аденомиозом, посредством непараметрического корреляционного анализа выявлены показатели, взаимосвязанные со стадиями заболевания, прогнозированием по клиническим результатам обследования больных статистическим методом *деревья классификации*.

**Результаты.** Методом ветвления удалось построить приемлемое дерево классификации, в котором достигнут компромисс между сложностью дерева и количеством ошибочных классификаций. Метод позволяет при построении дерева классификации определить роль (значимость) предикторов в модели классификации.

**Заключение.** Создание программных приложений автоматизирует процедуру классификации и делает возможным их использование медицинским персоналом, не имеющим специализированной подготовки в области анализа данных.

**Ключевые слова:** аденомиоз, прогнозирование, математическая модель

**Для цитирования:** Карахалис Л.Ю., Папова Н.С., Халафян А.А., Акиншина В.А. Прогнозирование стадии заболевания у больных аденомиозом при помощи деревьев классификации. *Кубанский научный медицинский вестник*. 2018; 25(4): 37-42. DOI: 10.25207 / 1608-6228-2018-25-4-37-42

**For citation:** Karakhalis L.Yu., Papova N.S., Khalafyan A.A., Akinshina V.A. Prediction of the stage of a disease in patients with adenomyosis using classification trees. *Kubanskij nauchnyj medicinskij vestnik*. 2018; 25(4): 37-42. (In Russ., English abstract). DOI: 10.25207 / 1608-6228-2018-25-4-37-42

L. YU. KARAKHALIS<sup>1</sup>, N. S. PAPOVA<sup>1</sup>, A. A. KHALAFYAN<sup>2</sup>, V. A. AKINSHINA<sup>2</sup>

PREDICTION OF DISEASE STAGE IN PATIENTS WITH ADENOMYOSIS USING CLASSIFICATION TREE

<sup>1</sup>Federal State Budgetary Educational Institution of Higher Education Kuban State Medical University of the Ministry of Healthcare of the Russian Federation, Sedina str., 4, Krasnodar, Russia, 350063.

<sup>2</sup>Kuban State University, Stavropolskaya str., 149, Krasnodar, Russia, 350040.

### ABSTRACT

**Aim.** This study was designed to develop the mathematical prediction model of adenomyosis spread stages according to the results of clinical examination using the *classification tree* statistical method.

**Materials and methods.** During this study we conducted the sampling of 84 patients with adenomyosis. By means of nonparametric correlation analysis we identified the indicators which were interconnected with the disease stage and prediction according to the results of clinical examination of the patients by means of the *classification tree* statistical method.

**Results.** We managed to build a suitable classification tree that helped to reach the compromise between the tree complexity and the amount of false classifications. This method allows us to define to role (significance) of the predictors in the classification model.

**Conclusion.** The creation of software applications automatizes the classification procedure and makes it possible for medical staff who don't have specialized training in data analysis sphere to use it.

**Keywords:** adenomyosis, prediction, mathematical model

### Введение

В недалеком прошлом математическая сторона медицинских исследований исчерпывалась применением критерия Стьюдента, или коэффи-

циента корреляции, сегодня же современный уровень развития компьютерных технологий позволяет использовать самый широкий спектр методов многомерного статистического анализа, в том

числе и классификационного. В работе показана возможность применения деревьев классификации для прогнозирования стадий аденомиоза по результатам клинических исследований.

Аденомиоз – это одно из самых распространенных гинекологических заболеваний, из-за трудностей диагностики его частота находится в достаточно широких пределах: 5-70%, при этом сопутствуя бесплодию [1-7]. У пациентки с аденомиозом имеются, как правило, разнохарактерные проявления клиники заболевания, которые представлены тяжелыми менструальными кровотечениями, дисменореей. При этом у части пациенток клинические проявления могут отсутствовать или быть бессимптомными. Сегодня нет стандартных диагностических критериев визуализации, а, следовательно, выбор тактики ведения пациентки и ее лечения является сложной задачей [8, 9]. Так сложилось, что исторически диагноз аденомиоза у женщин репродуктивного и перименопаузального возрастов устанавливался после гистерэктомии. Однако применение дооперационной визуализации (ультразвуковое исследование, магнитно-резонансная томография, лапароскопия, гистероскопия) позволило выяснить, что аденомиоз встречается в возрасте ранней репродукции и даже у подростков [10].

**Цель исследования:** разработать математическую модель прогнозирования стадий распространения аденомиоза по клиническим результатам обследования больных статистическим методом *деревья классификации*.

### Материалы и методы

Нами проведена выборка из 84 больных, страдающих аденомиозом, посредством непараметрического корреляционного анализа выявлены показатели, взаимосвязанные со стадиями заболевания: дисменорея (болезненные менструации), длительность менструального цикла (МЦ) в днях, длина и толщина тела матки в мм, число фолликулов в яичнике, тазовая боль, возникающая вне менструации, индекс резистентности левой маточной артерии (ЛМА), видимые при проведении гистероскопии, расположенные в полости матки кровоточащие эндометриоидные гетеротопии, некровоточащие эндометриоидные гетеротопии, дефекты эндометрия и выбухания эндометрия. Первые 7 показателей являются количественными, остальные 4 – качественные, которые принимают значения *да, нет*. Больные в группах по степени выраженности тяжести заболевания были представлены следующим образом: 1-я стадия распространения аденомиоза – 24 пациентки; 2-я стадия – 35 пациенток и 3-я – 25 пациенток. Стадии тяжести распространения заболевания были установлены при помощи проведенного в послеоперационном периоде гистологического исследования. Наличие взаимосвязей позволило решить задачу прогнозирования стадий заболевания по клиническим результатам обследования больных

статистическим методом *деревья классификации*.

### Результаты и обсуждение

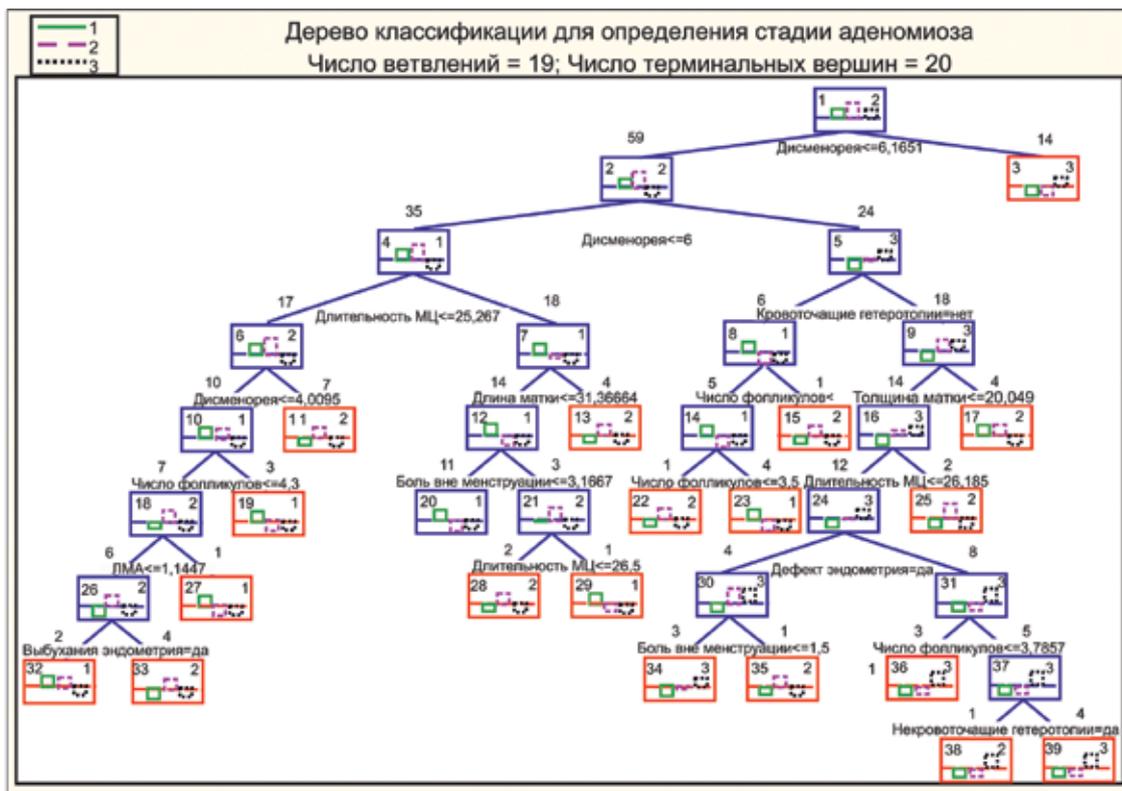
В пакете *STATISTICA* [11] реализованы методы построения бинарных (двоичных) деревьев классификации, предполагающие ветвление по одному показателю (предиктору). Бинарное дерево предполагает ветвление только по двум возможным направлениям каждого предиктора. Важным достоинством деревьев классификации является возможность графического представления результатов и простота интерпретации, допускающие абсолютную прозрачность процедуры классификации. Структура метода такова, что пользователь имеет возможность по управляемым параметрам строить деревья приемлемой сложности, добиваясь минимальных ошибок классификации.

Методом ветвления – *дисперсионное одномерное ветвление*; при правиле остановки – *прямая остановка*; доли неверно классифицированных объектов – 0,02; критерия согласия *Джини*; при пользовательских ценах ошибок классификации, заданных в таблице 1, удалось построить приемлемое дерево классификации, в котором достигнут компромисс между сложностью дерева и количеством ошибочных классификаций. Смысл цен ошибочных классификаций достаточно прост – цена выше, если больные ошибочно классифицируются по 2 удаленным классам 1 и 3; цена – ниже, если больные ошибочно классифицируются по двум соседним классам 1 и 2, 2 и 3.

Граф дерева классификации представлен на рисунке 1. Вершины (узлы) дерева изображены в виде прямоугольников. Номер узла написан в левом верхнем углу прямоугольника, над которым указано количество больных, отнесенных методом к этому узлу. Внутри каждой вершины графа дерева изображены столбчатые диаграммы, представляющие классы (группы) больных с соответствующей стадией заболевания. Терминальные вершины дерева, из которых дальнейшее ветвление невозможно отмечены красным цветом.

Для оценки адекватности модели из обучающей выборки случайным образом была выделена тестовая выборка, состоящая из трех больных стадии 1, трех больных стадии 3 и пяти больных стадии 2. Об адекватности построенной модели классификации можно судить по количеству ошибочно классифицированных больных обучающей (табл. 2) и тестовой (табл. 3) выборок деревом. Из Таблицы 2 следует, что все больные стадии 1 (21 женщина) и 22 больные стадии 3 классифицированы деревом верно; из 30 больных стадии 2 неверно классифицированы как больные стадии 1 и стадии 3 соответственно 1 и 2 человека.

Из таблицы 3 следует, что одна больная из 3 тестовой выборки стадии 1 ошибочно отнесена к стадии 2; четыре больные стадии 2 ошибочно отнесены к стадии 3, но, нет ошибочных классификаций больных 3 стадии.



**Рис. 1.** Дерево классификации стадий заболевания.  
**Fig. 1.** Classification tree of the disease stage.

Метод позволяет при построении дерева классификации определить роль (значимость) предикторов в модели классификации. В таблице 4 отображены ранги значимости в баллах в порядке их убывания. Из таблицы 4 видно, что наиболее важным рангом является *дисменорея* (100 баллов), далее идут по значимости *кровотокащие гетеротопии* (85 баллов) и т.д., замыкает рейтинг предикторов *выбухание эндометрия* (24 балла).

По дереву, изображенному на рисунке 1, легко составить алгоритм классификации.

### Алгоритм

**Шаг 1.** Если *дисменорея*  $\leq 6,165145$  баллов, переходим к шагу 2; в противном случае у больного стадия 3, переходим к шагу 33;

**Шаг 2.** Если *дисменорея*  $\leq 6$  баллов, переходим к шагу 3; в противном случае переходим к шагу 10;

**Шаг 3.** Если *длительность МЦ*  $\leq 25,26704$  дней, переходим к шагу 4; в противном случае переходим к шагу 7;

**Шаг 4.** Если *дисменорея*  $\leq 4,009462$  баллов, переходим к шагу 5; в противном случае у больного стадия 2;

**Шаг 5.** Если *число фолликулов*  $\leq 4,3$ , переходим к шагу 6; в противном случае у больного стадия 1;

**Шаг 6.** Если *ЛМА*  $\leq 1,44674$ , переходим к шагу 7, в противном случае у больного стадия 1;

**Шаг 7.** Если *выбухание эндометрия* – да, то у больного стадия 1; в противном случае у больного стадия 2, в обоих случаях переходим к шагу 33;

**Шаг 8.** Если *длина матки*  $\leq 31,36664$  мм, переходим к шагу 9; в противном случае у больного стадия 2, переходим к шагу 33;

**Шаг 9.** Если *боль вне менструации*  $\leq 3,166667$  балла, то у больного стадия 1; в противном случае переходим к шагу 10;

**Шаг 10.** Если *длительность МЦ*  $\leq 26,5$  дней, то у больного стадия 2, в противном случае – ста-

Таблица 2 / Table 2

### Ошибки классификации на обучающей выборке

The classification errors on the training sample

Таблица 1 / Table 1  
**Матрица цен ошибок классификации**  
 Price matrix of classification errors

Класс	Предсказанные (строки) и наблюдаемые (столбцы)		
	Класс 1	Класс 2	Класс 3
1	–	10,000	15,000
2	10,000	–	10,000
3	15,000	10,000	–

Класс	Предсказанные (строки) и наблюдаемые (столбцы), Объем обучающей. выборки 73		
	Стадия 1	Стадия 2	Стадия 3
1	–	1	0
2	0	–	0
3	0	2	–

Таблица 3 / Table 3

### Ошибки классификации на тестовой выборке

The classification errors on the test sample

Класс	Предсказанные (строки) и наблюдаемые (столбцы), Объем обучающей. выборки 73		
	Стадия 1	Стадия 2	Стадия 3
1	–	4	0
2	1	–	0
3	0	2	–

дия 1, переходим к шагу 33;

**Шаг 10.** Если *кровоточащие гетеротопии* – нет (отсутствуют), то переходим к шагу 12, в противном случае переходим к шагу 14;

**Шаг 12.** Если *число фолликулов*  $\leq 4$ , переходим к шагу 13, в противном случае у больного стадия 2, переходим к шагу 33;

**Шаг 13.** Если *длительность МЦ*  $\leq 26,5$  дней, то у больного стадия 2, в противном случае у больного стадия 1, переходим к шагу 33;

**Шаг 14.** Если *кровоточащие гетеротопии* – нет (отсутствуют), то переходим к шагу 15, в противном случае переходим к шагу 17;

**Шаг 15.** Если *число фолликулов*  $\leq 4$ , переходим к шагу 16, в противном случае у больного стадия 2, переходим к шагу 33;

**Шаг 16.** Если *число фолликулов*  $\leq 3,5$ , то у больного стадия 2, в противном случае у больного стадия 1, переходим к шагу 33;

**Шаг 17.** Если *толщина матки*  $\leq 20,049$  мм, то переходим к шагу 18, в противном случае у больного стадия 2, переходим к шагу 33;

**Шаг 18.** Если *длительность МЦ*  $\leq 26,18475$  дней, то переходим к шагу 19, в противном случае у больного стадия 2, переходим к шагу 33;

**Шаг 19.** Если *дефекты эндометрия* – да (присутствуют), то переходим к шагу 20, в противном случае переходим к шагу 21;

Таблица 4 / Table 4

### Ранги значимости предикторов модели

Grades of significance of the model predictors

Переменная	0 = низкая значимость; 100 = высокая
	Ранг
<i>дисменорея</i>	100
<i>кровоточащие гетеротопии</i>	85
<i>длина матки</i>	78
<i>некровоточащие гетеротопии</i>	74
<i>длительность МЦ</i>	61
<i>дефекты эндометрия</i>	49
<i>толщина матки</i>	41
<i>число фолликулов</i>	38
<i>боль вне менструации</i>	35
<i>ЛМА</i>	35
<i>выбухание эндометрия</i>	24

**Шаг 20.** Если *боль вне менструации*  $\leq 1,5$  баллов, то у больного стадия 3, в противном случае у больного стадия 2, переходим к шагу 33;

**Шаг 21.** Если *число фолликулов*  $\leq 3,785714$ , то у больного стадия 3, в противном случае переходим к шагу 22;

**Шаг 22.** Если *некровоточащие эндометриоидные гетеротопии* – да (присутствуют), то у больного стадия 2, в противном случае – стадия 3, переходим к шагу 33;

**Шаг 33.** Завершение работы алгоритма.

Из алгоритма становится объяснимым столь высокий рейтинг у *дисменореи* – если дисменорея  $> 6,165$ , то больных следует отнести к стадии 3, и таких больных оказалось 14 из 25 – более 50%.

Ввод данных

Введите данные о пациентке

Дисменорея

Длительность МЦ

Количество фолликулов

Боль вне менструации

Длина матки

Толщина матки

ЛМА

Выбухания эндометрия

Дефекты эндометрия

Кровоточащие гетеротопии

Некровоточащие гетеротопии

Продолжить

Рис. 2. Окно ввода данных.

Fig. 2. The data entry window.

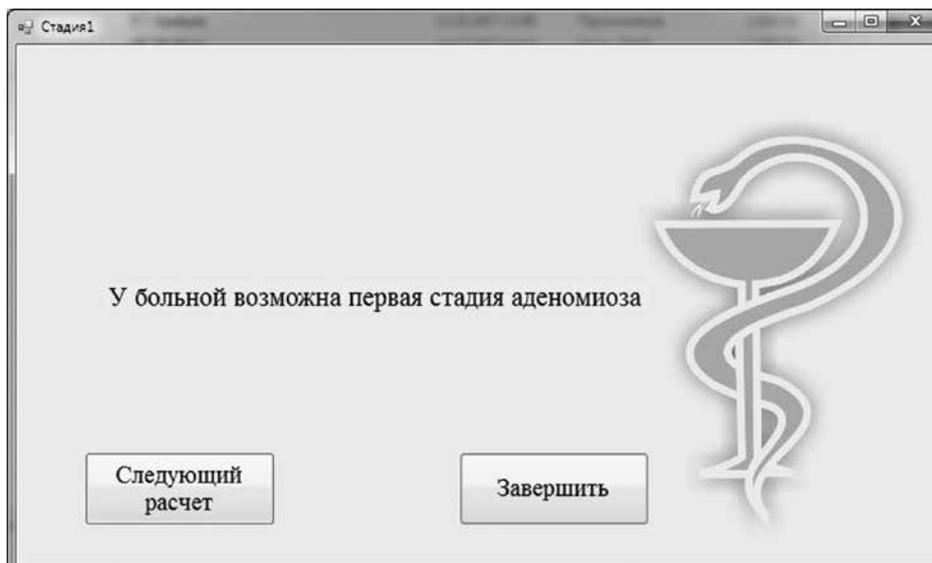


Рис. 3. Окно результатов прогнозирования.  
Fig. 3. The window of prediction results.

Работу алгоритма можно проиллюстрировать на примере. Предположим, что у больной А. дисменорея равна 5 баллам, длительность МЦ составляет 27 дней, количество фолликулов – 4, индекс резистентности в ЛМА – 0,87, боль вне менструации – 4 балла, длина матки – 25 мм, толщина матки – 20 мм, есть выбухания эндометрия, кровоточащие эндометриоидные гетеротопии. Сделаем прогноз стадии при помощи алгоритма.

**Шаг 1.** Так как дисменорея  $\leq 6,165145$  баллов, переходим к шагу 2;

**Шаг 2.** Так как дисменорея  $\leq 6$  баллам, переходим к шагу 3;

**Шаг 3.** Так как длительность МЦ  $> 25,26704$  дней, переходим к шагу 7;

**Шаг 7.** Так как выбухание эндометрия – да (имеются), то у больной стадия 1, переходим к шагу 33;

**Шаг 33.** Завершение работы алгоритма.

Для того, чтобы автоматизировать работу алгоритма на языке *Visual Basic (VB)* была написана программа с простым интерфейсом, не требующим от пользователя знаний компьютерных технологий анализа данных. Окно программы представлено на рисунке 2. Достаточно ввести данные больной в соответствующие поля окна.

Если щелкнуть по кнопке *Продолжить*, появится окно (рис. 3) с результатами прогнозирования возможной стадии заболевания.

### Заключение

Несомненно, что применение современных средств анализа данных, реализованных в виде статистических пакетов, открывает самые широкие возможности решения задач предсказания принадлежности больных к определенным классам по совокупности клинических показателей. Создание по разработанным алгоритмам программных приложений автоматизирует процедуру классификации и делает возможным их исполь-

зование медицинским персоналом, не имеющим специализированной подготовки в области анализа данных. Подобные приложения могут лечь в основу различных систем поддержки принятия врачебных решений.

### ЛИТЕРАТУРА / REFERENCES

1. Арутюнян А.Ф., Гайдуков С.Н., Кустаров В.Н. Современные аспекты патогенетически обоснованной терапии аденомиоза. *Педиатр.* 2016; 7(3): 92-97. doi: 10.17816/PED7392-97. [Arutunyan A.F., Gajdukov S.N., Kustarov V.N. Actual issues of pathogenetically substantiated therapy of adenomyosis. *Pediatr.* 2016; 7(6): 92-97. (In Russ., English abstract)].
2. Дамиров М.М., Олейникова О.Н., Майорова О.В. Генитальный эндометриоз: взгляд практикующего врача. М.: БИНОМ. 2013. 152 с. [Damirov M.M., Oleinikova O.N., Majorova O.V. Genital'nyj endometrioz: vzglyad praktikuyuschego vracha. M.: BINOM. 2013. 152 p. (In Russ.)].
3. Стрижаков А.Н., Давыдов А.И., Пашков В.М., Лебедев В.А. Доброкачественные заболевания матки /2-е изд. перераб. и доп. М.: ГЭОТАР-Медиа. 2014. 312 с. [Strijakov A.N., Davydov A.I., Pashkov V.M., Lebedev V.A. Dobrokachestvennye zabolevaniya матки. 2-e izd. pererab. i dop. M.: GEOTAR-Media. 2014. 312 p. (In Russ.)].
4. Адамян Л.В., Сонова М.М., Логинова О.Н., Арсланян К.Н. Роль ароматазы в развитии генитального эндометриоза. *Акушерство, гинекология, репродукция.* 2016; 10(4): 39-48. DOI: 10.17749/2313-7347.2016.10.4.039-048. [Adamyany L.V., Sonova M.M., Loginova O.N., Arslanyan K.N. The role of aromatase in the development of endometriosis. 2016; 10(4): 39-48. (In Russ., English abstract)].
5. Адамян Л.В. Эндометриоз: диагностика, лечение и реабилитация. Клинические рекомендации. Москва. 2013. 86 с. [Adamyany L.V. Endometrioz: diagnostika, lechenie i reabilitaciya. Klinicheskie rekomendacii. Moskva. 2013. 86 p. (in Russ.)].
6. Garavaglia E., Audrey S., Annalisa I., Stefano F., Iacopo T., Laura C., Massimo C. Adenomyosis and its impact on women fertility. *Iran J Reprod Med.* 2015; 13(6): 327-336.
7. Vannuccini S., Tosti C., Carmona F., Huang S.J., Chapron C., Guo S.W., Petraglia F. Pathogenesis of adenomyosis: an up-

date on molecular mechanisms. *Reprod Biomed Online*. 2017; 35: 592-601. DOI: 10.1016/j.rbmo.2017.06.016.

8. Struble J., Reid S., Bedaiwy M.A. Adenomyosis: a clinical review of a challenging gynecologic condition. *J Minim Invasive Gynecol*. 2016; 23(2): 164-85. DOI: 10.1016/j.mig.2015.09.018.

9. Карахалис Л.Ю., Пенжоян Г.А., Червонная И.Ю., Васина И.Б. Обоснование выбора препаратов при аденомиозе для восстановления репродуктивного потенциала. *Врач*. 2010; 10: 57-62. [Karakhalis L.Yu., Penjhojan G.A., Chervonnaya I.Yu., Vasina I.B. Obosnovanie vybora preparatov pri adenomyoze dlya vosstanovleniya reproduktivnogo potentsiala. *Vrach*. 2010; 10: 57-62. (In Russ.)].

10. Ryan G.L., Stolpen A., Van Voorhis B.J. An unusual cause of adolescent dysmenorrhea. *Obstetrics and Gynecology*. 2006; 108(4): 1017-1022. DOI: 10.1097/01.AOG.0000237163.98010.b3.

11. Халафян А.А. СТАТИСТИКА 6. Математическая статистика с элементами теории вероятностей. М.: Бином. 2010. 491 с. [Kalafyan A.A. STATISTIKA 6. Matematicheskaya statistika s elementami teorii veroyatnostej. M.: BINOM. 2010. 491 p. (In Russ.)].

*Поступила / Received 05.02.2018*

*Принята в печать / Accepted 27.04.2018*

*Авторы заявили об отсутствии конфликта интересов / The authors declare no conflict of interest*

**Контактная информация:** Карахалис Людмила Юрьевна; тел.: +7 (988) 244-40-44; e-mail: lomela@mail.ru; Россия, 350063, г. Краснодар, ул. Седина, д. 4.

**Corresponding author:** Lyudmila Y. Karakhalis; tel.: +7 (988) 244-40-44; e-mail: lomela@mail.ru; 4, Sedina str., Krasnodar, Russia, 350063.