

УДК 519.234.3

DOI: [10.26102/2310-6018/2020.28.1.027](https://doi.org/10.26102/2310-6018/2020.28.1.027)

Применение корреляционного анализа для выявления факторов из анамнеза женщины, влияющих на исход беременности, полученной с помощью ЭКО

С.Л. Синотова¹, О.В. Лимановская², А.Н. Плаксина³, В.А. Макутина⁴

^{1,2}ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина», Екатеринбург, Российская Федерация

³ФГБОУ ВО «Уральский государственный медицинский университет Министерства здравоохранения Российской Федерации», Екатеринбург, Российская Федерация

⁴АО «Центр семейной медицины», Екатеринбург, Российская Федерация

Резюме: В статье рассматривается поиск статистической взаимосвязи между заболеваниями мочеполовой системы, хроническими заболеваниями, проведенными хирургическими вмешательствами, и другими данными анамнеза жизни, наследственного анамнеза женщины и исходом беременности, полученной с помощью вспомогательных репродуктивных технологий (ЭКО). В качестве исхода выбраны, имеющиеся у рассматриваемой группы пациенток – 338 женщин, сроки и способы родоразрешения, а также диагнозы их 402 детей на этапе родильного дома. В статье описан универсальный алгоритм приведения текстовых данных в пригодный для проведения анализа вид при помощи морфологического анализатора rumorphy2 и библиотеки Scikit-Learn Python 3.6. Для выявления значимых факторов использовался корреляционный анализ, выполненный с применением точного критерия Фишера, критерия согласия Хи-квадрат и интервальных оценок долей, Z-критерия для разности двух долей. Сила взаимосвязи факторов и исходов оценена критерием Крамера. Результатом анализа является выделение 56 значимых факторов и 34 значимых корреляций, которые будут учтены при разработке математической модели предсказания исхода беременности и оценки здоровья ребенка, зачатого при помощи ВРТ (ЭКО), на этапе планирования.

Ключевые слова: обработка текста, корреляционный анализ, достигаемый уровень значимости, анализ малых данных, вспомогательные репродуктивные технологии, V Крамера, критерий согласия хи-квадрат, точный критерий Фишера, математическая статистика.

Для цитирования: Синотова С.Л., Лимановская О.В., Плаксина А.Н., Макутина В.А. Применение корреляционного анализа для выявления факторов из анамнеза женщины, влияющих на исход беременности, полученной с помощью ЭКО. *Моделирование, оптимизация и информационные технологии*. 2020;8(1). Доступно по: https://moit.vivt.ru/wp-content/uploads/2020/02/SinotovaSoavtors_1_20_1.pdf DOI: 10.26102/2310-6018/2020.28.1.027

Application of correlation analysis to identify factors from a woman's anamnesis influencing the results of pregnancy obtained by INF

S.L. Sinotova¹, O.V. Limanovskaja², A.N. Plaksina³, V.A. Makutina⁴

^{1,2} FSAEI HE «UrFU named after the first President of Russia B.N. Yeltsin», Ekaterinburg, Russian Federation

³ FSBEI HE «USMU of the Ministry of Health of the Russian Federation», Ekaterinburg, Russian Federation

⁴The Family Medicine Centre, Ekaterinburg, Russian Federation

Abstract: This article discusses the search for a statistical relationship between diseases of the genitourinary system, chronic diseases, surgical interventions and other data on the anamnesis, a woman's heredity and pregnancy outcome obtained using assisted reproductive technologies (IVF). The study is conducted with the aim of developing a mathematical model for predicting pregnancy and assessing the health of a child conceived using ART (IVF) at the stage of planning. The conclusions are based on data on 338 women and the diagnoses of their 402 children at the stage of the maternity hospital. A research was made for the effect of 56 binary signs on the outcome of pregnancy, described by 38 characteristics. To identify significant factors, a correlation analysis was performed using Fisher's exact test, Chi-square test, and using interval estimates of the shares, and the Z-criterion for the difference of two shares. As the outcomes, the terms and methods of delivery available to the patient group under consideration, as well as the diagnoses of children at the stage of the maternity hospital were selected. To assess the strength of the relationship, Cramer's V is applied. The result of the analysis is the identification of 56 significant factors and 35 significant correlations, which will be taken into account in the future for the development of the regression model.

Keywords: correlation analysis, p-value, analysis of small data, assisted reproductive technologies, Cramer's V, Chi-square test, Fisher's exact test, mathematical statistics.

For citation: Sinotova S.L., Limanovskaja O.V., Plaksina A.N., Makutina V.A. Application of correlation analysis to identify factors from a woman's anamnesis influencing the results of pregnancy obtained by IVF. *Modeling, Optimization and Information Technology*. 2020;8(1). Available from: https://moit.vivt.ru/wp-content/uploads/2020/02/SinotovaSoavtors_1_20_1.pdf DOI: 10.26102/2310-6018/2020.28.1.027 (In Russ).

Введение

Более сорока лет на помощь бездетным парам приходят вспомогательные репродуктивные технологии (ВРТ). Большое число влияющих факторов, наличие различных методик ВРТ и медикаментозного сопровождения, отсутствие систематических данных об исходах беременностей и рожденных младенцах создают сложность в понимании влияния на здоровье детей в дальнейшем.

Много ученых создают модели, предсказывающим как наступление беременности, так и ее исход, но единой принятой модели, как и единых значимых для ее построения факторов, не существует. Их поиск ведется до сих пор.

Есть ряд факторов, которые выделяются исследователями и применяются ими в моделях [1,2,3,4,5,6,7]. Самыми часто встречающимися факторами из анамнеза женщины являются: возраст [1,3,4,5,6,7], ИМТ [2,4], причина бесплодия [4,5,7], некоторые данные акушерского анамнеза [1,3,4,5]. Реже исследуют влияние следующих факторов: тип бесплодия [1,3], курение [2], этническая принадлежность [5], данные первого протокола [5,7] и число неудачных попыток ВРТ [3,6,7]. В некоторых работах рассматривается влияние социальных факторов и образа жизни пациентки. Так, авторы [2] подтверждают, что успех ВРТ можно прогнозировать с учетом факторов образа жизни: пациенты с нормальным весом имели значительно более высокую вероятность живорождения / беременности, чем с избыточным, некурящие более склонны к успешности программы, чем курильщики. В то же время в диссертации [8] установлено, что на формирование нарушений здоровья у детей, рожденных с помощью ВРТ, больше всего влияют факторы, обусловленные экстрагенитальной патологией, акушерско-гинекологическим анамнезом, не физиологическим течением беременности и психологическими характеристиками матери, недоношенностью ребенка; а социальные факторы и факторы, связанные с процедурой ЭКО, не значимы.

Таким образом, в настоящее время, несмотря на интенсивные исследования в этой области, не выделен набор факторов из анамнеза женщины, которые точно необходимо

учитывать при построении прогноза успешности процедуры ВРТ (ЭКО).

Данная работа посвящена выявлению корреляции факторов из наследственного анамнеза, данных о заболеваниях мочеполовой системы и хирургических вмешательствах, хронических заболеваний, анамнеза жизни и здоровья (в том числе впервые ли была попытка ЭКО) женщины с исходом программы ЭКО.

Материалы и методы

Ранее была в Автоматизированной системе «Региональный акушерский мониторинг (АС «РАМ»)) (свидетельство о государственной программе для ЭВМ № 2017619189 от 16.08.2017) получена база данных из 821 оригинальной записи о женщинах и рожденных ими 836 детях. База содержит информацию об анамнезе женщины, протоколе ВРТ, результате полученной беременности, и здоровье детей при рождении, путем объединения баз медицинских организаций за 2014 – 2018 гг. ГАУЗ СО «ОДКБ» и АО «Центр семейной медицины» с сохранением персональных данных. После того, как были убраны строки с неизвестным исходом и выбраны только те, для которых имеется дата вступления в протокол ВРТ, в выборке осталось 566 исходов для 564 женщин. У двух пациенток имелось по два исхода: прерывание и роды. Оставлен только один поздний исход – роды, как результат последней операции. Фильтр по программе ЭКО оставил в Таблице 338 женщин и их 402 детей, которые были использованы для данного анализа.

Объединение данных, разрозненных по разным базам было осуществлено с помощью уникального идентификатора пациента. Приведение данных в удобный вид, обработка и анализ выполнялись с использованием Anaconda — дистрибутива Python 3.6 [9]. Для кластеризации данных применен метод k-средних (KMeans), предварительная обработка для кластеризации выполнена с применением морфологического анализатора rymorphy2 [10] и классов CountVectorizer, TfidfVectorizer библиотеки Scikit-Learn [11].

Алгоритм приведения текстовых данных в вид, пригодный для корреляционного анализа, и последующего сопоставления данных с исходами беременностей:

1. Создание Таблицы (Pandas DataFrame), содержащей один исходный столбец с текстом.
2. Если текст представлен в произвольной форме и визуально понятно, что в каждой ячейке содержатся данные для одной группы, но число групп неизвестно, то выполняется кластеризация данных. Кластеризация данных представляет собой автоматическое выделение групп (кластеров) похожих по смыслу строк. Это преобразование выполнялась методом k-средних (k-means) при помощи соответствующей функции python, которая принимает на вход число кластеров (найденно визуально или заданием большего числа групп, при небольшом их количестве) и матрицу частот слов. До применения K-means необходимо подготовить данные и провести следующие операции:
 - a. Из текста убрать любые цифры, знаки препинания, символы, имеющиеся в тексте и обособленные пробелом, привести все слова в начальную форму при помощи библиотеки rymorphy2.
 - b. Создать матрицу частот слов (с помощью класса TfidfVectorizer, библиотеки sklearn), встречающихся в тексте: столбцы – все слова из текста (словарь), строки – исходная Таблица с текстовой информацией. На пересечении строки и столбца – число, показывающее, значимость каждого слова в тексте.
 - c. Выполнить корректировку словаря – отбросить ненужные для понимания слова.

Результатом применения k-means являются группы (классы), представленные

средствами python Таблицей, в которой столбцы – это номера классов, а строки – это исходный текст. Для дальнейшего использования данных необходимо заменить номер кластера его названием, и, при возникновении ошибок группировки данных, выполнить их корректировку вручную.

3. Если текст получен с помощью SQL-запросов к базе данных и визуально понятно, что в каждой ячейке содержатся данные для одной группы, но число групп неизвестно, то выполняется кластеризация данных, при этом обработкой при помощи библиотеки `rummy2` и удалением «мусора» можно пренебречь, поскольку текст не содержит грамматических ошибок и различных форм слов.
4. Если в тексте содержатся данные, характеризующие пациента с разных сторон, например: «роды, в стационаре, Срочные, оперативные, кесарево сечение, в нижнем сегменте, плановое», то кластеризация не применима. В таких случаях производятся следующие операции:
 - a. Предварительно из текста убираются любые цифры, знаки препинания, символы, имеющиеся в тексте и обособленные пробелом, приведение всех слов в начальную форму при помощи библиотеки `rummy2`.
 - b. Создается матрица частот слов (с помощью класса `CountVectorizer`, библиотеки `sklearn`), встречающихся в тексте: столбцы – все слова из текста (словарь), строки – исходная Таблица с текстовой информацией. На пересечении строки и столбца – число, показывающее, сколько раз слово встретилось в строке.
 - c. Выполняется корректировка словаря – отбрасывание не нужных для понимания слов.
 - d. Строится матрица частот слов еще раз с новым словарем.
5. Для того, чтобы совмещать Таблицы с разным количеством строк, нужно сгруппировать данные с помощью функции `groupby` по принципу «одна строка – один ID» с возможным добавлением строк для тех пациентов, которых нет в некоторых характеристиках.
6. Выполнить слияние данных анамнеза с Таблицей «Исходы» при помощи функции `merge` по ID пациента.

Данный алгоритм универсален, позволяет обрабатывать данные любых больших объемов.

Корреляционный анализ выполнен средствами Python 3.6 и включает в себя поиск зависимостей при помощи следующих методов:

1. Корреляционный анализ с применением точного критерия Фишера и критерия согласия Хи-квадрат. Проверка нулевой гипотезы выполнялась по алгоритму, представленному схемой на Рисунке 1. Уровень значимости во всех случаях принимался равным 0,05.
2. Корреляционный анализ с применением интервальных оценок долей. Для каждой пары «характеристика - исход» выборка делилась на две части: первую составляют исходы, принадлежащие женщинам, у которых была данная характеристика, вторую – если характеристика отсутствовала.

Первичная выборка содержит 338 записей, поэтому объем двух бинарных выборок получался небольшой. Во всех выборках наблюдалось малое число единиц и исключительно малое при исходах – диагнозах при рождении, следовательно, для построения интервальной оценки доли использовался метод Уилсона [12,13].



Рисунок 1 - Алгоритм корреляционного анализа с применением точного критерия Фишера и критерия согласия Хи-квадрат

Figure 1 – Correlation analysis algorithm using Fisher's exact test and Chi-square test

1. Корреляционный анализ с применением Z-критерия для разности двух долей. Нулевые гипотезы были построены по следующему принципу: выборка, в которой у женщины есть определенный признак, не отличается от выборки, в которой у женщин этого признака нет. Уровень значимости принимался равным 0,05.

Результаты

Выявление всех факторов. Первоначально осуществлялась обработка текстовых данных по исходам беременностей и приведение этих данных в числовой вид, пригодный для анализа. В данных каждого текстового столбца были выделены группы, и каждый пациент соответственно получил бинарный признак, относится ли он к этой группе (есть ли у него такая характеристика) или нет.

Таблица 1 содержит информацию о том, как выглядели данные до обработки и после применения алгоритма для обработки текстовых данных, описанного в разделе Материалы и методы.

Вся информация об исходе беременности и здоровье ребенка на этапе родильного дома описывается с помощью 38 дихотомических характеристик. Для корреляционного анализа использовались данные анамнеза, которые были представлены с помощью 56 бинарных признаков. Информация содержалась в 4 Таблицах. Первая Таблица отображала данные по наличию заболеваний мочеполовой системы у женщин, вторая - данные о хирургических вмешательствах, третья - данные по хроническим заболеваниям женщин и четвертая Таблица - прочие данные по анамнезу жизни, здоровья и наследственности.

Таблица 1 – предобработка данных об исходах беременности

Table 1 - pretreatment of pregnancy outcomes

№	Изначальный столбец	Вид данных	№	Полученная дихотомическая характеристика, описывающая исход беременности						
1	ID пациента	номинальные	1	ID пациента (без изменений)						
2	Исход: роды или прерывание беременности	текст, полученный с помощью SQL-запросов к базе	2	прерывание						
3	Информация об исходе: срок родов/ прерывания, способ родоразрешения	текст, содержащий различные характеристики для анализа в одной ячейке, полученный с помощью SQL-запросов к базе	3	плановое кесарево сечение (кс)						
			4	экстренное кс						
			5	неотложное кс						
			6	самостоятельные роды						
			7	вакуум-экстракция						
			8	преждевременные роды						
			9	сверхранные роды						
			10	ранние роды						
			11	срочные роды						
4	Количество рожденных детей	дискретные	12	одноплодная беременность						
5	Диагноз при рождении	текст, полученный с помощью SQL-запросов к базе	13	P00.2	22	P24.0	31	P92.1		
			14	P03.4	23	P25.3	32	Q21.0		
			15	P05.0(1,2)	24	P28.9	33	Q24.9		
			16	P07.3	25	P54.8	34	Q79.0		
			17	P08.0(1)	26	P55	35	Z20.6		
			18	P12.0	27	P59.8	36	Z37.0(2)		
			19	P20.0(1,9)	28	P61.9	37	Z38.0(3)		
			20	P21.0(1)	29	P70.0	38	Z39.2		
			21	P22.0(1)	30	P91.0	39	Z76.2		

Данные первых трех Таблиц, состоящих из двух столбцов (ID – характеристика) и их обработка представлены в Таблице 2. Из трех характеристик было выделено 28 признаков для анализа. Выделенные хронические заболевания являются самыми популярными (n>10).

Таблица 2 – предобработка данных о заболеваниях мочеполовой системы, хронических заболеваниях, проведенных хирургических вмешательствах
Table 2 – data preprocessing on diseases of the genitourinary system, chronic diseases, performed surgical interventions

№	Изначальный столбец	Вид данных	№	Полученная характеристика, описывающая дихотомическая анамнез женщины
1	Заболевания мочеполовой системы	текст, полученный с помощью SQL-запросов к базе	1	хронический аднексит
			2	аденомиоз
			3	наружный генитальный эндометриоз
			4	миома матки
			5	синдром поликистоза яичников
			6	цистит или пиелонефрит
		дихотомические, добавлены	7	нет заболеваний
			8	одно заболевание
			9	два заболевания
			10	три заболевания
2	Хирургические вмешательства	произвольный текст, содержащий различные характеристики для анализа в одной ячейке	11	вмешательства на молочных железах
			12	вмешательства на щитовидной железе
			13	выскабливание полости матки
			14	гистерорезектоскопия
			15	гистероскопия
			16	лапароскопия
			17	лапаротомия
			18	резекция в мочеполовой системе
			19	эмбриоскопия
		дихотомические, добавлены	20	не было вмешательств
3	Хронические заболевания	произвольный текст, содержащий различные характеристики для анализа в одной ячейке	21	аутоиммунный тиреоидит
			22	анемия
			23	гастрит
			24	гиперпролактинемия
			25	гипотиреоз
			26	диффузная фиброзно-кистозная мастопатия
			27	инфекционные заболевания
			28	миопия
			29	нейроциркуляторная и/или вегетососудистая дистония
			30	тромбофилия и/или варикозное расширение вен
			31	холецистит и/или панкреатит
			32	эутиреоз
		дихотомические, добавлены	33	нет хр. заболеваний

Анамнез жизни, здоровья, наследственности был представлен в виде столбцов, содержащих, как произвольную текстовую, так и числовую информацию. Для приведения данных к виду, пригодному для анализа, была произведена замена информации на бинарную: при наличии определенной характеристики присваивалась 1, при отсутствии – 0. В результате из 17 характеристик было выделено 25 дихотомических признака, представленных в Таблице 3.

Таблица 3 – признаки, полученные из анамнеза жизни, здоровья и наследственных данных
Таблица 3 – features got from an anamnesis, health estimate and hereditary data

1	18-25 лет	13	ИМТ норма
2	26-28 лет	14	ИМТ ожирение
3	29-30 лет	15	курение
4	31-35 лет	16	нормальная продолжительность цикла
5	36-40 лет	17	онкология у родственников
6	старше 40 лет [14]	18	особенности профессии (вред факторы)
7	аллергия	19	первая попытка ВРТ
8	гипертония, ишемия, инсульт, тромбозы у родственников	20	применялась ли гормональная контрацепция или ВМС
9	гормональное лечение в настоящее время	21	регулярные менструации без гормональных препаратов
10	диспансерный учет	22	черепно-мозговые травмы
11	ИМТ избыточный вес	23	язвы, ожирение, сахарный диабет у родственников
12	ИМТ недостаточный вес		

Некоторые характеристики не были разбиты на более конкретные из-за малой частоты наблюдений ($n < 5$), неприменимой для дальнейшего анализа.

Корреляционный анализ.

Для выявления факторов, влияющих на исход беременностей, был проведен корреляционный анализ, согласно описанному выше алгоритму, в котором проверялось наличие статистически значимой связи между каждым признаком ($n=56$) и исходом ($n=37$) без разделения на группы по признаку многоплодной или одноплодной беременности. Значимые корреляции, имеющие значение $p\text{-value} < 0,05$, и обнаруженными всеми тремя подходами, приведены в Таблице 4.

Таблица 4 – Значимые корреляции ($p\text{-value} < 0.05$)

Таблица 4 – Significant correlations ($p\text{-value} < 0.05$)

№	значимые корреляции	непараметрические критерии ($p\text{-value}$ и критерий)	интервальная оценка долей	Z-критерий для разности двух долей ($p\text{-value}$)	Критерий Крамера
заболевания мочеполовой системы					
1	два заболевания - самостоятельные роды	0.0050 ¹	(0.0044 ; 0.0565) (0.0616 ; 0.1406)	0.0051	0.14

2	наружный генитальный эндометриоз - диагнозы P20.1, P20.0, P20.9	0.0329 ¹	(0.1176 ; 0.7693) (0.0368 ; 0.0874)	0.0016	0.1207
3	нет заболеваний - самостоятельные роды	0.0003 ³	(0.8022 ; 0.9197) (0.6349 ; 0.7571)	0.0003	0.1887
4	нет заболеваний - экстренное кс	0.0041 ³	(0.4779 ; 0.7296) (0.7413 ; 0.8354)	0.0041	0.1466
5	пиелонефрит и/или цистит – диагноз P91.0	0.0462 ¹	(0.2065 ; 1.0) (0.034 ; 0.0828)	4.0682e-05	0.105
хирургические вмешательства					
6	выскабливание полости матки - диагноз P03.4	0.0048 ¹	(0.3424 ; 1.0) (0.081 ; 0.1481)	0.0001	0.1533
7	выскабливание полости матки - диагнозы P20.1, P20.0, P20.9	0.0099 ¹	(0.2307 ; 0.8824) (0.0791 ; 0.146)	0.0006	0.1475
8	гистероскопия - самостоятельные роды	0.0049 ¹	(0.5856 ; 0.7481) (0.7749 ; 0.8753)	0.0008	0.1757
9	гистероскопия - срочные роды	0.0008 ¹	(0.6688 ; 0.7795) (0.8114 ; 0.9399)	0.0014	0.1658
10	лапаротомия - диагноз P55	0.0054 ¹	(0.3424 ; 1.0) (0.1629 ; 0.2486)	0.0055	0.1034
хронические заболевания					
11	анемия - диагноз P54.8	0.0296 ¹	(0.2065 ; 1.0) (0.0141 ; 0.05)	9.7157e-09	0.1512
12	гиперпролактинемия - диагноз P54.8	0.0533 ¹	(0.2065 ; 1.0) (0.0317 ; 0.0793)	2.4148e-05	0.1084
13	гиперпролактинемия - неотложное кс	0.0366 ³	(0.0532 ; 0.2554) (0.0258 ; 0.0734)	0.0366	0.0935
14	гиперпролактинемия – ранние роды	0.0386 ¹	(0.0705 ; 0.4519) (0.0283 ; 0.0702)	0.0096	0.1088
15	нет хронических заболеваний - диагноз P08.0, P08.1	0.0106 ¹	(0.4365 ; 0.9699) (0.2459 ; 0.3432)	0.0042	0.1313
16	нет хронических заболеваний - неотложное кс	0.0075 ³	(0.0532 ; 0.2554) (0.2758 ; 0.3819)	0.0075	0.1357
17	нет хронических заболеваний - срочные роды	0.0007 ³	(0.2966 ; 0.4152) (0.1014 ; 0.2517)	0.0007	0.1776
18	тромбофилия и/или варикозное расширение вен - диагноз P70.0	0.0318 ¹	(0.0454 ; 0.3213) (0.0129 ; 0.0433)	0.0061	0.116
19	холецистит и/или панкреатит - диагноз P07.3	0.0201 ¹	(0.0715 ; 0.5907) (0.0123 ; 0.0471)	0.0002	0.1451
20	эутиреоз - диагноз P54.8	0.0385 ¹	(0.2065 ; 1.0) (0.0205 ; 0.0612)	5.5167e-07	0.1307
анамнез жизни, здоровья, наследственный					
21	26-28 лет - преждевременные	0.0054 ³	(0.1226 ; 0.3557) (0.0568 ; 0.1214)	0.0054	0.1381
22	29-30 лет - диагнозы Z37.0, Z37.2	0.0006 ²	(0.268 ; 0.732) (0.0989 ; 0.1732)	0.0001	0.1875
23	гормональное лечение в настоящее время - диагноз Z39.2	0.02685 ¹	(0.3006 ; 0.9544) (0.1557 ; 0.2404)	0.0059	0.1157
24	есть нарушения цикла - диагноз Z38.0,3	0.0056 ³	(0.0212 ; 0.0963) (0.0993 ; 0.194)	0.0056	0.1408

25	ИМТ норма - диагноз P70.0	0.0112 ³	(0.2216 ; 0.5921) (0.6003 ; 0.7056)	0.0112	0.1266
26	ИМТ ожирение - прерывание	0.0094 ³	(0.1636 ; 0.5268) (0.0903 ; 0.1633)	0.0094	0.1247
27	онкология у родственников - диагноз P12.0	0.0418 ¹	(0.2077 ; 0.9385) (0.0889 ; 0.1585)	0.0042	0.1078
28	первая попытка - диагнозы Z37.0, Z37.2	0.01144 ¹	(0.0401 ; 0.3994) (0.4501 ; 0.5592)	0.008	0.153
29	первая попытка - самостоятельные роды	0.0037 ³	(0.5051 ; 0.6762) (0.3635 ; 0.4962)	0.0037	0.1303
30	применялась ли гормональная контрацепция или вмс - диагноз Z38.0,3	0.0024 ³	(0.1988 ; 0.3488) (0.0953 ; 0.1886)	0.0024	0.1572
31	применялась ли гормональная контрацепция или вмс - срочные роды	0.0014 ³	(0.1797 ; 0.284) (0.0373 ; 0.1488)	0.0014	0.1647
32	старше 40 лет - диагноз P70.0	5.96e-08 ³	(0.1255 ; 0.4647) (0.0131 ; 0.0501)	5.9629e-08	0.2675
33	ЧМТ - диагноз P28.9	0.0296 ¹	(0.2065 ; 1.0) (0.0141 ; 0.05)	9.7157e-09	0.1512
34	ЧМТ - диагноз Q79.0	0.0296 ¹	(0.2065 ; 1.0) (0.0141 ; 0.05)	9.7157e-09	0.1512

1 – точный критерий Фишера

2 – критерий согласия хи-квадрат с коррекцией Йетса

3 - критерий согласия хи-квадрат

Обсуждение

По данным анализа видно, что было выделено 24 значимых фактора и 34 значимых корреляции. Замечено, что чаще всего результаты поиска зависимостей с помощью разных методов одинаковы.

Для оценки силы влияния взаимосвязей был рассчитан коэффициент Крамера (в Таблице 4). Интерпретация значений V Крамера осуществлялась согласно рекомендациям Rea & Parker. [15] Связь средней силы обнаружена в одной корреляции: возраст матери больше 40 лет в средней степени влияет на возникновение диагноза P70.0 (Синдром новорожденного от матери с гестационным диабетом) у новорожденного. Несущественная сила связи также обнаружена в одной: гиперпролактинемия в анамнезе влияет на возникновение ситуации, когда понадобится проведение неотложного кс. Остальные все связи слабые.

Заключение

Из заболеваний мочеполовой системы, хирургических вмешательств, анамнеза жизни, здоровья и наследственности были обнаружены 34 статистически значимых корреляции с признаками, описывающими исходы беременностей, полученных с помощью ЭКО. Все использованные методы корреляционного анализа выявили примерно одинаковые зависимости. В качестве значимых корреляций были выбраны те, которые обнаружили все три метода. Сила связи, оцененная по V Крамера оказалась средней только в одном случае, несущественной также в одном, и слабой во всех остальных корреляциях. Данные зависимости будут учтены при построении регрессионной модели.

В дальнейшем, планируется исследовать влияние акушерского анамнеза, фактора бесплодия, экологического фактора (места проживания), данных протокола на исход

беременности, полученной с помощью ЭКО. К исходу планируется добавить данные о здоровье детей раннего возраста, а выборки разделить по признаку многоплодной и одноплодной беременности.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-90121.

Acknowledgments: The reported study was funded by RFBR, project number 19-37-90121.

ЛИТЕРАТУРА

1. Амирова А.А. Прогнозирование исходов ЭКО и ЭКО/ИКСИ у бесплодных супружеских пар при некоторых формах бесплодия : Дис. ... канд. мед. наук. Москва; 2011. Доступно по: <https://www.dissercat.com/content/prognozirovanie-iskhodov-eko-i-ekoiksi-u-bespolodnykh-supruzheskikh-par-pri-nekotorykh-forma> (дата обращения 17.02.2020).
2. Satvinder Purewal, Sarah C.E. Chapman, Olga B.A. van den Akker. A systematic review and meta-analysis of lifestyle and body mass index predictors of successful assisted reproductive technologies. *Journal of Psychosomatic Obstetrics & Gynecology*. 2019;40(1):2-18 DOI:10.1080/0167482X.2017.1403418
3. L.L. van Loendersloot, M. van Wely, S. Repping, P.M.M. Bossuyt, F. van der Veen. Individualized decision-making in IVF: calculating the chances of pregnancy. *Human Reproduction*. 2013;28(11):2972-2980 DOI: 10.1093/humrep/det315
4. Dhillon R.K., McLernon D.J, Smith P.P, Fishel S., Dowell K., Deeks J.J., Bhattacharya S., Coomarasamy A. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Human Reproduction*. 2016;31(1):84-92 DOI: 10.1093/humrep/dev268
5. David J McLernon, Ewout W Steyerberg, Egbert R te Velde, Amanda J Lee, Siladitya Bhattacharya. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ*. 2016;355(8082) DOI: 10.1136/bmj.i5735
6. Porcu, G., Lehert, P., Colella, C, Giorgetti C. Predicting live birth chances for women with multiple consecutive failing IVF cycles: a simple and accurate prediction for routine medical practice. *Reproductive Biology and Endocrinology*. 2013;11(1) DOI: 10.1186/1477-7827-11-1
7. Katarina Kebbon Vaegter, Tatevik Ghukasyan Lakic, Matts Olovsson, Lars Berglund, Thomas Brodin, Jan Holte Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertility and Sterility*. 2017;107(3):641-648.e2 DOI: 10.1016/j.fertnstert.2016.12.005
8. Гаджимурадова Н.Д. Состояние здоровья и прогнозирование его нарушений у детей первого года жизни, родившихся от одноплодной беременности после экстракорпорального оплодотворения: Дис. ... канд. мед. наук. Пермь; 2017. Доступно по: <http://www.dslib.net/pediatrics/sostojanie-zdorovja-i-prognozirovanie-ego-narushenij-u-detej-pervogo-goda-zhizni.html> (дата обращения 18.02.2020).
9. Платформа для обработки данных и машинного обучения Anaconda. Доступно по: <https://www.anaconda.com> (дата обращения 18.02.2020)
10. Морфологический анализатор rymorphy2. Доступно по: <https://rymorphy2.readthedocs.io/en/latest/> (дата обращения 18.02.2020).

11. Pedregosa et al. Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12:2825-2830.
12. Wilson E. B. Probable inference, the law of succession, and statistical inference. *Journal of American Statistical Association*. 1927;22:209-212.
13. Гржибовский А.М. Доверительные интервалы для частот и долей. *Экология человека*. 2008;5:57-60.
14. Колбая Т.Т. Беременность и роды у женщин различных возрастных групп: Дис. ... канд. мед. наук. Москва; 2011. Доступно по: <http://medical-diss.com/medicina/beremennost-i-rody-u-zhenschin-razlichnyh-vozzrastnyh-grupp> (дата обращения 18.02.2020)
15. Гржибовский А. М. Анализ номинальных данных (независимые наблюдения). *Экология человека*. 2008;6:58-68.
16. Гланц С. Медико-биологическая статистика. Пер. с англ. М.: Практика;1998.

REFERENCES

1. Amirova A.A. Prognozirovanie iskhodov EHKO i EHKO/IKSI u bespolodnykh supruzheskikh par pri nekotorykh formakh besplodiya [dissertation]. Moscow; 2011. (In Russ) Available at: <https://www.dissercat.com/content/prognozirovanie-iskhodov-eko-i-ekoiksi-u-bespolodnykh-supruzheskikh-par-pri-nekotorykh-forma> (accessed 17.02.2020).
2. Satvinder Purewal, Sarah C.E. Chapman, Olga B.A. van den Akker. A systematic review and meta-analysis of lifestyle and body mass index predictors of successful assisted reproductive technologies. *Journal of Psychosomatic Obstetrics & Gynecology*. 2019;40(1):2-18 DOI:10.1080/0167482X.2017.1403418
3. L.L. van Loendersloot, M. van Wely, S. Repping, P.M.M. Bossuyt, F. van der Veen. Individualized decision-making in IVF: calculating the chances of pregnancy. *Human Reproduction*. 2013;28(11):2972-2980 DOI: 10.1093/humrep/det315
4. Dhillon R.K., McLernon D.J, Smith P.P, Fishel S., Dowell K., Deeks J.J., Bhattacharya S., Coomarasamy A. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Human Reproduction*. 2016;31(1):84-92 DOI: 10.1093/humrep/dev268
5. David J McLernon, Ewout W Steyerberg, Egbert R te Velde, Amanda J Lee, Siladitya Bhattacharya. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ*. 2016;355(8082) DOI: 10.1136/bmj.i5735
6. Porcu, G., Lehert, P., Colella, C, Giorgetti C. Predicting live birth chances for women with multiple consecutive failing IVF cycles: a simple and accurate prediction for routine medical practice. *Reproductive Biology and Endocrinology*. 2013;11(1) DOI: 10.1186/1477-7827-11-1
7. Katarina Kebbon Vaegter, Tatevik Ghukasyan Lakic, Matts Olovsson, Lars Berglund, Thomas Brodin, Jan Holte Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertility and Sterility*. 2017;107(3):641-648.e2 DOI: 10.1016/j.fertnstert.2016.12.005
8. Gadzhimuradova N.D. Sostoyanie zdorov'ya i prognozirovanie ego narushenii u detei pervogo goda zhizni, rodivshikh sya ot odnoplodnoi beremennosti posle ehkstrakorporal'nogo oplodotvoreniiya [dissertation]. Perm; 2017. (In Russ) Available at: <http://www.dslib.net/pediatria/sostojanie-zdorovja-i-prognozirovanie-ego-narushenij-u-detej-pervogo-goda-zhizni.html> (accessed 18.02.2020).
9. Anaconda - solutions for Data Science Practitioners and Enterprise Machine Learning. Available at: <https://www.anaconda.com> (accessed 18.02.2020)

10. Morphological analyzer pymorphy2. Available at: <https://pymorphy2.readthedocs.io/en/latest/> (accessed 18.02.2020).
11. Pedregosa et al. Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12:2825-2830.
12. Wilson E. B. Probable inference, the law of succession, and statistical inference. *Journal of American Statistical Association*. 1927;22:209 – 212.
13. Grjibovski A.M. Confidence intervals for proportions. *Human Ecology*. 2008;5:57-60. (In Russ)
14. Kolbaya T.T. Beremennost' i rody u zhenshin razlichnykh vozrastnykh grupp [dissertation]. Moscow; 2011. (In Russ) Available at: <http://medical-diss.com/medicina/beremennost-i-rody-u-zhenshin-razlichnykh-vozrastnykh-grupp> (accessed 18.02.2020)
15. Grjibovski A.M. Analysis of nominal data (independent observations). *Human Ecology*. 2008;6:58-68. (In Russ)
16. Glantz S. Primer of biostatistics. M.: Practica;1998. (In Russ)

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Синотова Светлана Леонидовна, аспирант, младший научный сотрудник кафедры интеллектуальных информационных технологий института фундаментального образования ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина», Екатеринбург, Российская Федерация
e-mail: sveta.volkova92@mail.ru
ORCID: [0000-0002-4318-5223](https://orcid.org/0000-0002-4318-5223)

Svetlana L. Sinotova, P.G., Junior Researcher of The Department of Intellectual Information Technologies, Institute of Fundamental Education, FSAEI HE «UrFU named after the first President of Russia B.N.Yeltsin», Ekaterinburg, Russian Federation

Лимановская Оксана Викторовна, к.х.н., доцент кафедры интеллектуальных информационных технологий института фундаментального образования ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина», Екатеринбург, Российская Федерация
e-mail: o.v.limanovskaia@urfu.ru
ORCID: [0000-0002-2084-3916](https://orcid.org/0000-0002-2084-3916)

Oksana V. Limanovskaja, Candidate of Chemical Sciences, Associate Professor of The Department of Intellectual Information Technologies, Institute of Fundamental Education, FSAEI HE «UrFU named after the first President of Russia B.N.Yeltsin», Ekaterinburg, Russian Federation

Плаксина Анна Николаевна, к.м.н., ассистент кафедры физической и реабилитационной медицины ФГБОУ ВО «Уральский государственный медицинский университет Минздрава РФ», Екатеринбург, Российская Федерация
e-mail: burberry20@yandex.ru
ORCID: [0000-0002-3119-478X](https://orcid.org/0000-0002-3119-478X)

Anna N. Plaksina, Candidate of Medical Sciences, Assistant of The Department of Physical and Rehabilitation Medicine, FSBEI HE «USMU of the Ministry of Health of the Russian Federation», Ekaterinburg, Russian Federation

Макутина Валерия Андреевна, к.б.н., эмбриолог АО «Центр семейной медицины», Екатеринбург, Российская Федерация
e-mail: makutina_v@rambler.ru
ORCID: [0000-0003-1127-2792](https://orcid.org/0000-0003-1127-2792)

Valerija A. Makutina, Candidate of Biological Sciences, Embryologist of The Family Medicine Centre, Ekaterinburg, Russian Federation