

СОЗДАНИЕ СИНТЕТИЧЕСКОЙ ПОПУЛЯЦИИ С ХРОНИЧЕСКИМИ ЗАБОЛЕВАНИЯМИ НА ОСНОВЕ НЕСКОЛЬКИХ НАБОРОВ ДАННЫХ

Киселевская-Бабинина В.Я., Санникова Т.Е., Романюха А.А.

Институт вычислительной математики им. Г.И. Марчука РАН, Москва, Россия
silvaze@yandex.ru, te_san@vahoo.com, eburg101@mail.ru

Аннотация. В данной работе предлагается метод создания синтетической популяции для моделирования эпидемии. Метод основан на создании и объединении основной и вспомогательной байесовских сетей. При помощи метода была построена синтетическая популяция и для неё получены оценки эпидемии инфекционного заболевания.

Ключевые слова: синтетическая популяция, байесовские сети, математическое моделирование эпидемии.

Введение

При моделировании эпидемий инфекционных заболеваний неоднородность параметров людей играет важную роль. Динамика эпидемии зависит от распределений индивидуальных характеристик сопротивляемости, поведении, образе жизни. Например, в эпидемиологии туберкулёза учёт неоднородности параметров в популяции помогает в определении групп риска, источника заражения и необходимых мер по борьбе с туберкулёзом [1]. Моделирование распространения эпидемии без учёта неоднородности параметров приводит к недооценке числа заражённых индивидов [2]. Один из методов моделирования неоднородности – создание синтетической популяции, то есть искусственного набора индивидов с заданными характеристиками, распределение которых совпадает или близко к распределению в реальной популяции.

Существует ряд методов создания синтетических популяций, основанных либо на случайной генерации из набора реальных данных людей, либо на заполнении пустого набора индивидов характеристиками, опираясь на оценки параметров распределения этих характеристик [3]. На выбор метода создания популяции влияет то, какой тип данных есть у авторов [4]. Данные, используемые для создания популяции, подразделяются на оценки распределения характеристик по популяции, и индивидуальные данные людей, представляющие собой записи о конкретном человеке со всеми его характеристиками. Проблема в том, что использование оценок распределений отдельных характеристик не воссоздаёт корреляции между характеристиками, а многомерные индивидуальные данные малочисленны и чаще всего являются выборками из какой-то определённой группы популяции, то есть смещены [4].

В данной работе мы предлагаем способ создания синтетической популяции, основанный на объединении нескольких байесовских сетей, построенных по разным наборам индивидуальных данных. Байесовские сети являются гибким и надёжным инструментом создания синтетической популяции, которые лучше других методов воссоздают популяцию на малых выборках до 40% от всей популяции [5], могут работать со смещенными типами параметров [6], а также объединены с другими методами [7, 8]. Наш метод предполагает отдельную оценку двух байесовских сетей: основной – для выборки, отражающей большую часть популяции, но имеющей мало характеристик – и вспомогательной – для выборки, отражающей малую и смещённую часть популяции, но содержащей все необходимые характеристики. При помощи данного метода мы построили байесовскую сеть для создания синтетической популяции, где характеристиками индивидов являлись хронические заболевания. На основе сети была создана синтетическая популяция, распределения хронических заболеваний и их корреляции были сравнены с первоначальными данными. Мы промоделировали инфекцию COVID-19 в получившейся синтетической популяции с использованием индекса предрасположенности к тяжёлому течению COVID-19 и марковской модели течения COVID-19.

1. Материалы и методы

1.1. Источники данных

Для построения байесовской сети популяции были использованы данные системы наблюдения за поведенческим фактором риска (Behavioral Risk Factor Surveillance System (BRFSS)) за 2021 год, собранные центром контроля и профилактики заболеваний (Centers for Disease Control and Prevention (CDC)) [10]. Данные представляют собой результаты телефонных опросов жителей США о наличии у них установленных лечащими врачами хронических заболеваний, их поведении, питании, условиях

проживания, а также информацию о штате проживания, поле, возрасте, дате опроса и прочую информацию. Из всех данных за 2021 год мы составили выборку из ответов респондентов по полу, возрастной группе, наличию высокого давления (гипертонии), инфаркта миокарда, ишемической болезни сердца, инсульта, рака, хронической обструктивной болезни лёгких, хронических болезней почек, диабета и ожирения. Данная выборка использовалась для построения основной байесовской сети, поскольку содержала несмещённые данные, но в ней не было наблюдений по характеристикам, необходимым для моделирования COVID-19.

Вторая выборка состояла из данных пациентов, пролеченных с COVID-19 в инфекционном отделении стационара Москвы с марта 2020 года по июнь 2021 года [11]. Данные содержали пол, возраст пациента, дату госпитализации, длительность лечения, применение респираторной поддержки за каждый день лечения, исход и имеющиеся хронические заболевания. Мы отобрали заболевания, которые необходимы для оценки индекса предрасположенности к тяжёлому течению COVID-19 [12]. Также мы включили данные о типе оказанной пациенту дыхательной поддержки на момент госпитализации – была ли ему оказана инвазивная (ИВЛ) или неинвазивная (НИВЛ) искусственная вентиляция лёгких, или пациент сохранял самостоятельное дыхание. Эта выборка использовалась для создания вспомогательной байесовской сети, и содержала данные о всех необходимых хронических заболеваниях. Мы провели гармонизацию данных этой выборки: добавили данные о заболеваниях, присутствующих в BRFSS опроснике (диабет и ожирение), и разбили возраста пациентов на группы согласно группам BRFSS опросника.

Распространённость хронических заболеваний была сравнена между выборками и созданной синтетической популяцией при помощи критерия Фишера. Были вычислены ϕ -коэффициенты корреляции между заболеваниями, сравнение достоверности различий между коэффициентами корреляций проводилось при помощи вычисления доверительного интервала методом Zou [9]. Для всех критериев был выбран уровень достоверности 0,05.

1.2. Создание синтетической популяции

Для создания синтетической популяции мы использовали метод, основанный байесовской сети. Байесовская сеть – это направленная графовая модель, в которой узлами V_i графа являются случайные величины, а рёбрами r_i графа – имеющиеся условные зависимости этих величин друг от друга. В каждом узле без родителей определяется распределение вероятности значений случайной величины этого узла: $P(V_k = p)$. В каждом дочернем узле определяются условные вероятности значений в зависимости от значения родительского узла: $P(V_i = v | parents(V_i) = p)$. Структура сети предполагает, что любой дочерний узел зависит только от непосредственных родителей, и не зависит от всех прочих узлов, что называется локальным марковским свойством. При этом в любом дочернем узле можно определить безусловную вероятность его значений через произведение всех вероятностей родителей и родителей родителей этого узла по правилу Байеса:

$$P(V_1, \dots, V_n) = \prod_{i=1}^n P(V_i | parents(V_i)) \quad (1)$$

При создании синтетической популяции в качестве узлов графа выбираются необходимые характеристики индивидов и задаются возможные их значения с распределениями вероятностей этих значений. При этом можно как прямо задать вид графа и все вероятности на основе экспертных оценок, так и оптимизировать граф и вероятности на основе данных наблюдений.

В данной работе мы построили две байесовские сети по данным наблюдений. Основная сеть была создана определением оптимального графа и вероятностей значений характеристик по данным опроса жителей США; вспомогательная сеть состояла из оптимального графа и вероятностей, определённых по данным пациентов московского стационара. В основную сеть мы включили данные по полу, возрасту и 7 хроническим заболеваниям. Во вспомогательную сеть были включены пол, возраст, 13 хронических заболеваний, а также тип дыхательной поддержки, который пациент получил в момент госпитализации в стационар.

Оптимизация структуры графов по данным наблюдений была проведена при помощи максимизации логарифмической функции правдоподобия со штрафом на число рёбер (аналог критерия Акаике):

$$AIC(G|D) = \ln(P(D|G, \theta)) - d \xrightarrow{G, \theta} \min \quad (2)$$

Где G – структура графа, D – данные наблюдений, θ – оценённые параметры для выбранной структуры, d – число рёбер графа.

Поиск оптимальной структуры был проведён методом восхождения к вершине. Случайным образом были созданы 10 графов, из каждого из 10 графов были случайно получены 10 "соседних" графов-

кандидатов малочисленным удалением или добавлением нескольких рёбер. Для каждого из кандидатов были получены оценки параметров, вычислено значение логарифмической функции наибольшего правдоподобия, и был выбран наилучший кандидат с наибольшим значением логарифмической функции правдоподобия. Оценки параметров графа были проведены при помощи метода наибольшего правдоподобия по данным наблюдений.

Из-за локального марковского свойства байесовской сети распределение любой случайной величины в графе не зависит от всей структуры сети, если известны родительские узлы. Это свойство позволило нам объединить полученные две байесовские сети в одну итоговую сеть, которая использовалась для синтеза популяции. Для этого в основную сеть мы добавили отсутствующие узлы, перенеся их из вспомогательной сети с унаследованием рёбер к родительским узлам и условных вероятностей случайных величин. На Рисунке 1 показано объединение двух графов в итоговую сеть. Синтетическая популяция была создана путём случайного присвоения индивидам характеристик согласно распределениям вероятностей в порядке от родительских узлов к дочерним.

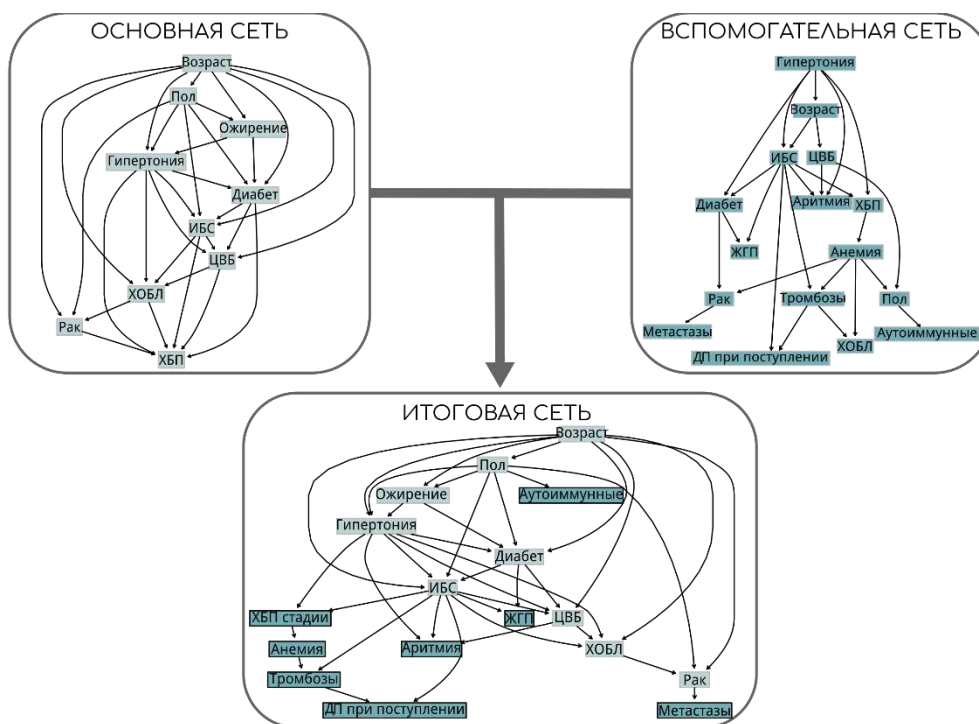


Рис. 1. Схема объединения графов основной и вспомогательной байесовских сетей в граф итоговой сети. ИБС – ишемическая болезнь сердца, ЦВБ – цереброваскулярная болезнь сердца, ХОБЛ – хроническая обструктивная болезнь лёгких, ХБП – хроническая болезнь почки, ЖГП – жировой гепатоз печени, РП – респираторная поддержка

1.3. Моделирование инфекции в синтетической популяции

Для каждого индивида синтетической популяции в зависимости от имеющихся у него хронических заболеваний было вычислено значение индекса предрасположенности к тяжёлому течению COVID-19 [12]. С вероятностью, полученной из доли заражений относительно всей численности города Москвы за 2020 год, были определены заболевшие индивиды, эта вероятность составила 0,295. Все заболевшие считались госпитализированными, и течение заболевания у каждого заражённого индивида было промоделировано при помощи марковской модели течения COVID-19 [13]. Начальное состояние индивида в модели было определено из типа дыхательной поддержки в момент госпитализации, присвоенного каждому индивиду на этапе создания популяции. С шагом моделирования в один день случайным образом состояние изменялось на основе переходных вероятностей, оцененных в зависимости от значения индекса предрасположенности [12]. Моделирование проводилось на 30 дней для каждого индивида. После 30 дней моделирования было подсчитано число умерших индивидов, число мужчин, средний возраст, сумма дней, в которые индивиды пребывали в разных состояниях модели. Описанный алгоритм был повторён 500 симуляций, полученные параметры всех симуляций были усреднены.

2. Результаты

Данные телефонного опроса жителей США состояли из закодированных ответов 438693 человек, что составило 0,13% всех жителей США. Мы сочли ответ "да" на вопрос о хроническом заболевании как наличие хронического заболевания, а прочие ответы – как отсутствие хронического заболевания. Данные пациентов московского стационара состояли из историй болезни 3212 человек, госпитализированных с COVID-19. Сравнительный анализ показал, что данные по пациентам смещены относительно данных опроса: люди с COVID-19 в среднем были старше на 5 лет, чаще мужчины, с повышенной вероятностью гипертонии, ИБС, ЦВБ, ХБП и диабета, и пониженной вероятностью рака и ХОБЛ. Сравнение параметров приведено в Таблице 1.

Таблица 1. Сравнение данных телефонного опроса жителей США, данных историй болезни пациентов с COVID-19, и характеристик синтетической популяции по полу, возрасту, наличию хронических заболеваний и оказанной дыхательной поддержке при госпитализации.

Характеристика	Данные опроса	Данные пациентов	Синтетическая популяция
Всего	438693	3212	100000
Возраст, среднее (дисперсия)	54,43 (19,96)	59,95 (15,84)*	54,5 (19,95)†
Мужчин	203760 (46,45%)	1707 (53,14%)*	46643 (46,64%)†
Гипертония	172133 (39,24%)	2126 (66,19%)*	39341 (39,34%)†
Ишемическая болезнь сердца	35323 (8,05%)	832 (25,90%)*	7879 (7,88%)†
Цереброваскулярные заболевания	17213 (3,92%)	534 (16,63%)*	3864 (3,86%)†
Рак	42349 (9,65%)	274 (8,53%)*	9762 (9,76%)†
Хроническая обструктивная болезнь лёгких	34169 (7,79%)	163 (5,07%)*	7915 (7,92%)†
Хроническая болезнь почек	17101 (3,90%)	948 (29,51%)*	20987 (20,99%)†
Диабет	57616 (13,13%)	644 (20,05%)*	13022 (13,02%)†
Ожирение	131305 (29,93%)	-	29991 (29,99%)
Аутоиммунные заболевания	-	165 (5,14%)	5522 (5,52%)
Анемия	-	463 (14,41%)	13162 (13,16%)†
Аритмия	-	471 (14,66%)	6283 (6,28%)†
Тромбозы	-	459 (14,29%)	12026 (12,03%)†
Жировой гепатоз печени	-	154 (4,79%)	5041 (5,04%)
Метастазы	-	58 (1,81%)	2054 (2,05%)
Искусственная вентиляция лёгких при госпитализации	-	155 (4,83%)	3771 (3,77%)†

* – достоверные отличия данных пациентов от данных опроса,

† – достоверные отличия синтетической популяции от данных пациентов.

Имелись достоверные корреляции между заболеваниями в выборке по данным опроса, наибольшая не превышала 0,27 (между гипертонией и диабетом). Абсолютные значения корреляций по данным пациентов в половине пар характеристик были выше, чем по данным опроса. Карта значений корреляций между заболеваниями в обоих выборках приведена на Рисунке 2.

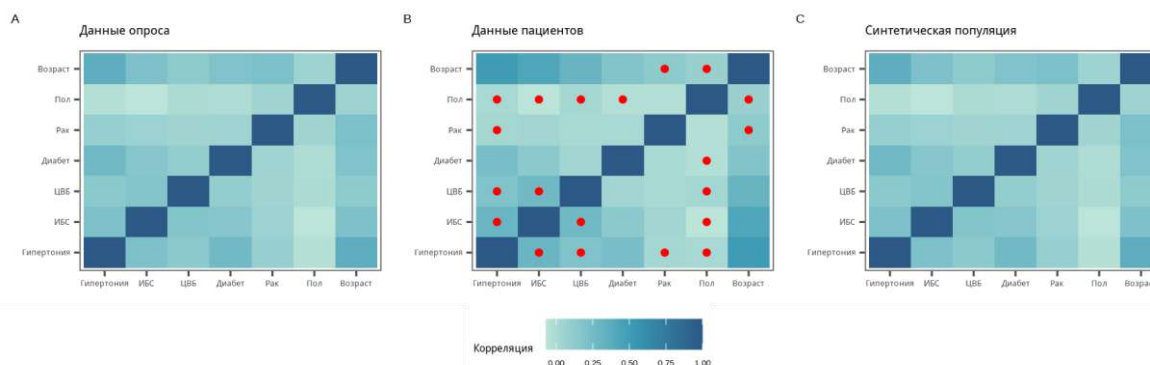


Рис. 2. Корреляции между хроническими заболеваниями: А – по данным опроса жителей США, В – по данным пациентов московского стационара, С – в синтетической популяции. Красными точками отмечены достоверные отличия между корреляциями по данным опроса и по данным пациентов

Найденные оптимальные графы основной и вспомогательной байесовских сетей, а также итоговая байесовская сеть представлены на Рисунке 1. Мы перенесли узлы, соответствующие аутоиммунным заболеваниям, жировому гепатозу печени, метастазам, аритмии, стадиям хронической болезни почки, анемии, тромбозам и типу дыхательной поддержки при госпитализации, из вспомогательной сети в основную с сохранением рёбер к родительским узлам и условных вероятностей. В процессе объединения были потеряны некоторые связи: у узлов "Анемия" и "Тромбозы" во вспомогательной сети был дочерний узел "Хроническая обструктивная болезнь лёгких", тогда как в основной сети узел "Хроническая обструктивная болезнь лёгких" имел родительские узлы "Цереброваскулярные заболевания", "Ишемическая болезнь сердца" и "Возраст". Мы приоритизировали основную сеть и опустили те рёбра, которые были исключительно во вспомогательной сети.

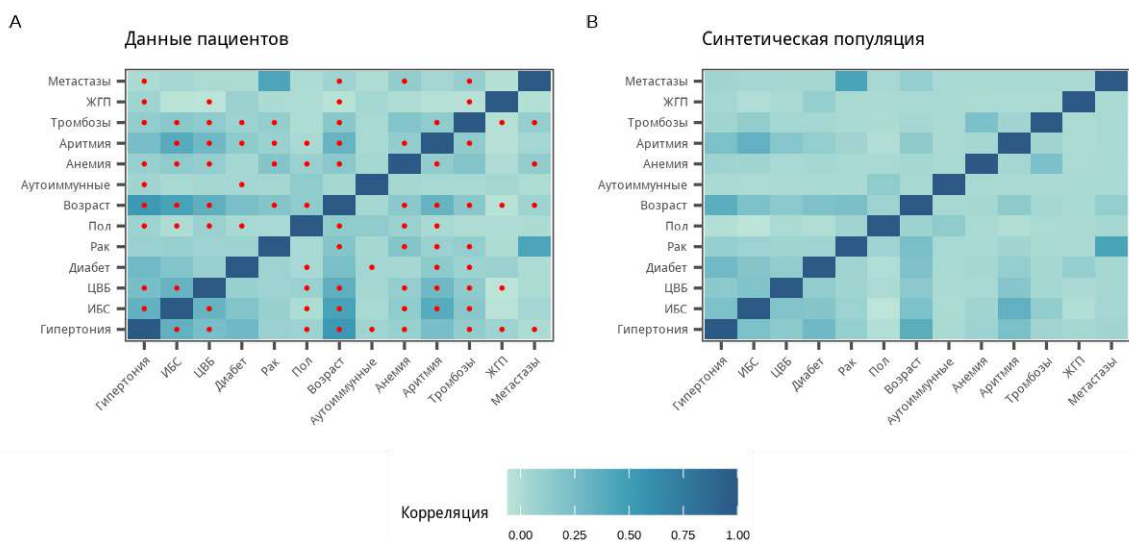


Рис. 3. Корреляции между хроническими заболеваниями: А – по данным пациентов московского стационара, В – в синтетической популяции. Красными точками отмечены достоверные отличия между корреляциями по данным пациентов и в синтетической популяции

На основе итоговой байесовской сети была создана синтетическая популяция из 100000 индивидов, характеристики созданной синтетической популяции представлены в Таблице 1, корреляции между заболеваниями в популяции представлены на Рисунках 2 и 3. Различия между синтетической популяцией и данными опроса в распределении хронических заболеваний и значениях их корреляций оказались недостоверны. Сравнение синтетической популяции и данных пациентов показало, что распределения аутоиммунных заболеваний, жирового гепатоза печени и метастатического рака достоверно совпали; распределения анемии и тромбозов, а также распределение применения ИВЛ в первый день госпитализации оказались близки по значениям, но достоверность не подтвердилась критерием Фишера; распределение аритмии в синтетической популяции оказалось достоверно ниже более чем в два раза. Значения корреляций заболеваний по данным пациентов и в синтетической популяции также оказались близки относительно, но в выборке корреляции оказались выше по абсолютным значениям почти в половине случаев (Рисунок 3).

Результаты моделирования течения заболевания COVID-19 у индивидов синтетической популяции представлены в Таблице 2. Общая летальность от COVID-19 в синтетической популяции составила 0,034, летальность госпитализированных индивидов – 0,132. К сравнению, в Москве за 2020-2021 годы было заражено 2023256 человек, умерло 36561 человек, общая летальность составила 0,018, в 2022 году госпитализировано 101519 человек, умерло 9405 человек, летальность госпитализированных составила 0,093 [14]. Выжившие индивиды были моложе на 3 года, находились на лечении на 2 дня дольше, и почти не требовали применения инвазивной вентиляции лёгких при лечении.

Летальность пролеченных на ИВЛ составила 0,734858, на НИВЛ – 0,1245113.

Таблица 2. Параметры умерших и выживших зараженных индивидов синтетической популяции.

Характеристика	Выжившие	Умершие
Всего, чел.	25625,83	3375,35
Мужчин, чел.	11835,94 (46,19%)	1634,03 (48,41%)

Характеристика	Выжившие	Умершие
Средний возраст, лет	54,05 (1,27)	57,55 (3,27)
Средняя длительность без поддержки, дн.	8,74 (2,61)	4,15 (3,01)
Средняя длительность НИВЛ, дн.	0,88 (0,19)	1,29 (0,19)
Средняя длительность ИВЛ, дн.	0,14 (0,06)	3,26 (0,28)
Средняя длительность лечения, дн.	9,77 (3,03)	7,07 (3,95)
Пролечено на ИВЛ, чел.	640,11 (2,5%)	2089,38 (61,90%)
Пролечено на НИВЛ, чел.	4450,35 (17,37%)	503,32 (14,91%)

3. Обсуждение

Представленный в данной работе метод создания синтетической популяции при помощи объединения основной и вспомогательной байесовских сетей помог создать популяцию, содержащую все требуемые нам характеристики. Основная сеть была создана при помощи данных из более широкой выборки реальной популяции, вспомогательная сеть – при помощи более чем в 100 раз меньшей выборки, смещённость данных которой была доказана сравнительным статистическим анализом. Сохранение всех узлов и рёбер основной сети позволило создать популяцию индивидов, у которых распределение характеристик и корреляции между этими характеристиками достоверно не отличаются от широкой выборки. Добавление узлов из вспомогательной сети, позволило воссоздать распространённости отсутствующих характеристик. При этом нам удалось избежать смещённости в итоговой синтетической популяции, поскольку в синтетической популяции значения распространённости и корреляций между характеристиками оказались в большинстве случаев достоверно ниже, чем в данных меньшей выборки. Статистическое совпадение остальных характеристик может быть обусловлено их редкостью, поскольку присутствие каждой из них не превышает 6% размера всей выборки.

Смещённость данных малой выборки вызвана тем, что она содержала информацию о госпитализированных пациентах с COVID-19, которые заведомо имели высокую тяжесть заболевания, в том числе и из-за большего, чем в среднем по населению, числа хронических заболеваний [15]. Это нашло отражение в наших данных: и распространённость, и корреляции хронических заболеваний у госпитализированных пациентов оказались выше по сравнению с населением США.

Создание байесовских сетей также помогло воссоздать логические взаимосвязи между характеристиками. В верхних родительских узлах оказались возраст и пол. Следующими в иерархии (то есть причинами остальных заболеваний), оказались такие распространённые заболевания, как гипертония и ожирение, что согласуется с наблюдениями врачей [16, 17]. Подобный подход к анализу и представлению хронических заболеваний, как к причинно-следственной байесовской сети между ними, актуален и активно развивается в последние годы [18, 19].

Включение в генерацию синтетической популяции не только необходимых для предсказания тяжести инфекции хронических заболеваний, но и вероятность первого состояния марковской модели течения COVID-19 позволило нам получить точные оценки параметров эпидемии. Аналогично в статье 2025 года [20] показано, что метод внедрения поведенческих и прочих факторов в процесс создания синтетической популяции позволяет добиваться большей точности, чем методы их случайной оценки.

Полученные оценки параметров течения COVID-19 у заражённых индивидов синтетической популяции в целом оказались близки к наблюдаемым. То, что среди умерших преобладали мужчины, индивиды старших возрастных групп, и им требовалось более частое применение инвазивной искусственной вентиляции лёгких согласуется с данными по COVID-19 в мире и в России [21-23]. Несовпадение летальности в синтетической популяции с данными динамики COVID-19 в Москве может быть вызвано разными периодами наблюдения.

Одним из недостатков данной работы заключается в том, что в качестве выборки для создания основной сети мы использовали данные телефонного опроса жителей США, у которых распределение хронических заболеваний или их диагностирование может отличаться от таковых у населения России [24]. Это было связано с тем, что мы не нашли в свободном доступе индивидуальных данных по хроническим заболеваниям жителей России, поэтому нам пришлось взять наиболее близкую по структуре популяцию другой страны. Кроме того, для оценки достоверности мы сравнили структуру синтетической популяции со структурами имеющихся у нас выборок. Нами не были найдены данные о распространённости у жителей России хронических заболеваний, присутствующих только в данных пациентов, из-за их редкости, что сделало невозможным оценку достоверности их распространённости в синтетической популяции.

Представленный нами метод создания синтетической популяции может быть использован не только для воссоздания распределения и взаимосвязей хронических заболеваний, но также и для других возможных характеристик, представляющих интерес для изучения и моделирования. Построение синтетической популяции можно провести по региональным данным России, чтобы учесть региональные различия в распространённости хронических заболеваний и сравнить параметры эпидемии инфекционного заболевания популяциях с разными структурами.

4. Заключение

В данной работе мы предложили метод синтеза популяции, основанный на разработке и объединении двух байесовских сетей, каждая из которых была создана на отдельном наборе данных. Метод помог воссоздать популяцию, в которой сохранились распределения и взаимосвязи характеристик из большой несмещённой выборки, а также имелись дополнительные характеристики из малой смещённой. В полученной синтетической популяции была промоделирована эпидемия инфекционного заболевания и получены оценки параметров течения инфекции у заражённых индивидов, оценки совпали с данными клинических наблюдений.

Литература

1. Trauer J. M. et al. The importance of heterogeneity to the epidemiology of tuberculosis // *Clinical infectious diseases*. – 2019. – Vol. 69, № 1. – P. 159–166.
2. Zagnutt F.J., Schoenbaum M.A., Hill A.E. The impact of population, contact, and spatial heterogeneity on epidemic model predictions // *Risk Analysis*. – 2016. – Vol. 36, № 5. – P. 939–953.
3. Müller K., Axhausen K.W., Population synthesis for microsimulation: State of the art // *Arbeitsberichte Verkehrs-und Raumplanung*. – 2010. – Т. 638.
4. Chapuis K., Taillandier P., Drogoul A. Generation of Synthetic Populations in Social Simulations: A Review of Methods and Practices // *JASSS*. – 2022. – Vol. 25, № 2. – P. 6.
5. Sun L., Erath A. A Bayesian network approach for population synthesis // *Transportation Research Part C: Emerging Technologies*. – 2015. – Vol. 61. – P. 49–62.
6. Sokolova E. et al. Causal discovery from databases with discrete and continuous variables // *European Workshop on Probabilistic Graphical Models*. – Cham: Springer International Publishing, 2014. – P. 442–457.
7. Rahman Md.N., Fatmi M.R. Population Synthesis Accommodating Heterogeneity: A Bayesian Network and Generalized Raking Technique // *Transportation Research Record: Journal of the Transportation Research Board*. – 2023. – Vol. 2677, № 6. – P. 41–57.
8. Ilahi A., Axhausen K.W. Implementing Bayesian network and generalized raking multilevel IPF for constructing population synthesis in megacities // *18th Swiss Transport Research Conference (STRC 2018)*. – STRC, 2018.
9. Zou G.Y. Toward using confidence intervals to compare correlations. // *Psychological Methods*. – 2007. – Vol. 12, № 4. – P. 399–413.
10. Behavioral Risk Factor Surveillance System. <https://www.cdc.gov/brfss>. (дата обращения 26.05.2025).
11. Киселевская-Бабинина В.Я., Понугаев К.А., Молодов В.А., Киселевская-Бабинина И.В. Использование ресурсов инфекционного коечного фонда в период эпидемии COVID-19 в зависимости от характеристик пациентов // *Журнал им. Н.В. Склифосовского «Неотложная медицинская помощь»*. – 2023. – Т. 12, № 3. – С. 481–488.
12. Киселевская-Бабинина В.Я., Санникова Т.Е. Разработка индекса коморбидности CoRS-COVID и его валидация при помощи марковской модели // *Управление развитием крупномасштабных систем (MLSD'2023): труды Шестнадцатой международной конференции*. – М.: ИПУ РАН, 2023. – № 2. – С. 1544–1552.
13. Киселевская-Бабинина В.Я. и др. Математическая модель течения COVID-19 и прогноз тяжести инфекции // *Математическое моделирование*. – 2023. – Т. 35, № 5. – С. 31–46.
14. Новости о коронавирусе в России | Стопкоронавирус. <https://объясняем.рф/stopkoronavirus/>. (дата обращения 26.05.2025).
15. Ko J.Y. et al. Risk Factors for Coronavirus Disease 2019 (COVID-19)–Associated Hospitalization: COVID-19–Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System // *Clinical Infectious Diseases*. – 2021. – Vol. 72, № 11. – P. e695–e703.
16. Kumar J. Epidemiology of hypertension // *Clinical Queries: Nephrology*. – 2013. – Vol. 2, № 2. – P. 56–61.
17. Jiang S.-Z. et al. Obesity and hypertension // *Experimental and Therapeutic Medicine*. – 2016. – Vol. 12, № 4. – P. 2395–2399.
18. McNally R.J. et al. Co-morbid obsessive–compulsive disorder and depression: a Bayesian network approach // *Psychological Medicine*. – 2017. – Vol. 47, № 7. – P. 1204–1214.
19. Jones P.J., Ma R., McNally R.J. Bridge Centrality: A Network Approach to Understanding Comorbidity // *Multivariate Behavioral Research*. – 2021. – Vol. 56, № 2. – P. 353–367.
20. Von Hoene E. et al. Synthetic population generation with public health characteristics for spatial agent-based models // *PLoS Comput Biol*. – 2025. – Vol. 21, № 3. – P. e1012439.

21. *Huang C. et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China // *The Lancet*. – 2020. – Vol. 395, № 10223. – P. 497–506.
22. *Клыпа Т.В. и др.* Клиническая характеристика пациентов с COVID-19, поступающих в отделение интенсивной терапии. Предикторы тяжелого течения // *Клиническая практика*. – 2020. – Т. 11, № 2. – С. 6–20.
23. *Арутюнов Г.П. и др.* Сравнительный анализ данных эхокардиографии и электрокардиографии выживших и умерших пациентов с COVID-19 (субанализ международного регистра “Анализ динамики Коморбидных заболеваний у пациентов, перенесших инфицирование SARS-CoV-2”) // *Российский кардиологический журнал*. – 2022. – Т. 27, № 3. – С. 9–17.
24. *Самородская И.В. и др.* Показатели смертности населения старше 50 лет от цереброваскулярных болезней за 15-летний период в России и США // *Неврология, нейропсихиатрия, психосоматика*. – 2017. – Т. 9, № 2. – С. 15–24.