

УДК 543

ПРИМЕНЕНИЕ МЕТОДОВ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ ДЛЯ УСТАНОВЛЕНИЯ КРИТЕРИЕВ ДЕГРАДАЦИИ ПРОБ МОЧИ В ЦЕЛЯХ ДОПИНГОВОГО КОНТРОЛЯ

© 2021 г. И. И. Подольский^{a, *}, Е. С. Мочалова^b, А. З. Темердашев^c, Э. М. Гашимова^c

^aООО “Брукер”

ул. Пятницкая, 50/2, стр. 1, Москва, 119017 Россия

^bНациональная антидопинговая лаборатория (институт) Московского государственного университета имени М.В. Ломоносова

Елизаветинский пер., 10, стр. 1, Москва, 105005 Россия

^cКубанский государственный университет

ул. Ставропольская, 149, Краснодар, 350040 Россия

*e-mail: podolskiy.i.i@gmail.com

Поступила в редакцию 19.01.2021 г.

После доработки 22.01.2021 г.

Принята к публикации 25.01.2021 г.

Рассмотрены результаты применения различных статистических методов анализа к обработке большого массива данных, полученных с помощью валидированного метода газохроматографического анализа с тандемным масс-спектрометрическим детектированием. Полученные данные являются неотъемлемой частью стероидного профиля человека и используются в качестве характеристичных признаков в лабораториях допинг-контроля. Рассмотрена возможность применения статистических методов анализа для установления факта деградации проб мочи в процессе хранения на основе хеометрических подходов, выявлены факторы, подвергающиеся наибольшему влиянию в ходе деградации образцов мочи, продемонстрированы возможности и ограничения различных статистических методов при работе с большими массивами данных.

Ключевые слова: стероиды, допинг, деградация мочи, ГХ–МС, хеометрика, дискриминантный анализ, нейронные сети.

DOI: 10.31857/S0044450221060074

Присутствие микроорганизмов и протекание протеолитических ферментативных процессов в образцах мочи неизбежно и способно существенно влиять на результаты некоторых анализов, проводящихся в рамках допингового контроля. Ускорение данных процессов происходит при несоблюдении условий хранения и транспортировки образцов, описанных в руководстве Всемирного антидопингового агентства (ВАДА) [1]. Основными факторами, влияющими на деградацию мочи, являются температура хранения и длительность хранения при определенной температуре, что особенно актуально в летний период в южных регионах. К сожалению, на сегодняшний день нет универсального консерванта, позволяющего остановить все виды микробиологической и ферментной активности, однако отдельные процессы деградации можно остановить путем ввода ряда веществ и реагентов. Так, например, азид натрия и хлорид ртути эффективно ингибируют бактериальную трансформацию глюкуроида анд-

ростерона в 5 α -андростан-3,17-дион в моче [2–4]. Ингибировать активность протеаз можно с использованием хлорида бестатина, ЭДТА, аprotинина и т.д. [2, 5, 6]. Однако независимо от используемого метода консервации необходимо учитывать возможность влияния реагентов на процессы пробоподготовки и анализа, которые могут существенно отличаться для разных антидопинговых лабораторий ввиду отсутствия унифицированных методик анализа и наличия оригинальных процедур [7–10]. Кроме того, наличие в емкостях для хранения до введения образцов каких-либо жидкостей (консервирующих растворов) или порошков может быть использовано для оспаривания полученных в ходе исследования результатов. В качестве аргументов ссылаются на то, что запрещенные соединения могли быть добавлены в саму емкость, что могло тем самым привести к возможности оспорить или скрыть ряд положительных проб, выдав их за ложноположительный результат.

Таблица 1. Стандартные образцы стероидных гормонов

Название	Чистота, %	Производитель
Тестостерон (Т)	98	Cerilliant, США
Эпитестостерон (Эпит)	98	Cerilliant, США
Андростерон (Андро)	98	Riedel-de-Haen, США
Этиохоланолон (Этио)	99	Sigma Aldrich, США
5 α -Андростан-3 α ,17 β -диол (5α-диол)	99	NMI, Австралия
5 β -Андростан-3 α ,17 β -диол (5β-диол)	99	NMI, Австралия
Форместан (Ф)	98	Toronto Research Chemical (Канада)
Дегидроэпиандростерон (ДГЭА)	99	LGC Standards, Великобритания
7 β -ОН-дегидроэпиандростерон (7β-ОН-ДГЭА)	99	LGC Standards, Великобритания
5 β -Андростан-3,17-дион (5β-дион)	99	TRC, Канада
D3-Тестостерон (D3-Т)	90	NMI, Австралия
D3-Эпитестостерон (D3-Эпит)	94	NMI, Австралия
D3-5 α -Андростан-3 α ,17 β -диол (D3-5α-диол)	98	ArtMolecule, Франция
D5-Эпиандростерон (D5-ЭпиА)	90	ArtMolecule, Франция
D5-Андростерон глюкуронид (D5-А)	90	ArtMolecule, Франция
Метилтестостерон (МТ)	99	Toronto Research Chemical (Канада)

Совокупность данных факторов привела к тому, что в пробы мочи на сегодняшний день не добавляют консерванты. Это стало причиной появления многоэтапного контроля, включающего как оценку возможности фальсификации образца, так и оценку его деградации [11–17].

В лабораторной практике антидопинговых лабораторий первыми этапами анализа является определение pH образцов мочи и их относительной плотности. Физиологически нормальными значениями является диапазон pH от 4.5 до 8.0. Более высокие значения pH могут говорить о возможной микробиологической деградации образца, в то время как значения pH < 4.5 могут быть следствием подмены образца или его разбавления, например, яблочным соком (pH ~ 3.0).

Более сложными маркерами деградации является оценка наличия и соотношения 5 α -андростан-3,17-диона (**5 α -дион**) и 5 β -андростан-3,17-диона (**5 β -дион**) в моче.

На сегодняшний день известен ряд работ, посвященных не только определению эндогенных стероидов в моче [18–23], распространенности их конъюгатов [24–29], но и изучению активности различных микроорганизмов, обуславливающих деградацию стероидов, при хранении проб мочи [2–6, 11–17]. Учитывая отличающуюся скорость деградации различных групп стероидов, отсутствие контроля за протеканием данного процесса может существенно исказить получаемые результаты.

Среди методов контроля деградации и фальсификации проб одним из наиболее многообещающих (помимо подхода, основанного на определе-

нии содержания эндогенных стероидов) является метод, основанный на применении полимеразной цепной реакции (**ПЦР**) [16, 17], где при деградации наблюдается только изменение соотношений сигналов по выбранным локусам (при длительном хранении возрастает погрешность исследования). Подобный метод полностью исключает возможность фальсификации за счет подмены образца на этапе сдачи пробы или даже ее хранения.

Целью данной работы является изучение возможности применения различных алгоритмов статистического анализа в целях установления факта деградации образцов мочи, поступивших на анализ, как с учетом действующих критериев ВАДА и оценки их эффективности на больших массивах данных, так и путем определения содержания андрогенных стероидов, используемых при оценке стероидного профиля, для установления факта деградации проб мочи в дополнение к уже известным маркерам деградации.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Реагенты и реактивы. Стандартные образцы стероидных гормонов перечислены в табл. 1. Использовали дигидрофосфат калия (99%, Sigma-Aldrich, США), дитиотреитол (**ДТТ**) (98%, Sigma-Aldrich, США), гидрокарбонат калия (99.7%, Sigma-Aldrich, США), иодид аммония (99%, Sigma-Aldrich, США), диэтиловый эфир (99%, Fisher Chemical, США), β -глюкуронидазу *E. coli* K12 (Roche, Германия), N-метил-N-триметилсилилтрифторацетамид (**МСТФА**) (99%, Macharey-Na-

Таблица 2. Соответствие аналитов их дейтерированным внутренним стандартам (ВС) и концентрации внутренних стандартов в пересчете на 3 мл мочи

Соединение	Внутренний стандарт	Концентрация ВС, нг/мл
Тестостерон, ДГЭА, 7β-ОН-ДГЭА	D3-Тестостерон	50
Эпитестостерон	D3-Эпитестостерон	50
Андростерон, этиохоланолон, 5β-Андростандион	D5-Андростерон глюкуронид*	1000
5α-Диол, 5β-диол	D3-5α-Диол	50

* В пересчете на агликон.

gel, Германия), безводный сульфат натрия (99%, Sigma-Aldrich, США), гелий 6.0 (Линде газ Рус, Россия), аргон 6.0 (Линде газ Рус, Россия).

Оборудование. Использовали газовый хроматограф Thermo Trace-1310, оснащенный капиллярной колонкой Agilent HP Ultra-1 (17 м × 0.2 мм, толщина пленки неподвижной жидкой фазы 0.11 мкм) со следующей температурной программой: начальная температура термостата колонок 180°C с последующим нагревом до 235°C со скоростью 4°C/мин с последующей интенсификацией скорости нагрева термостата до 20°C/мин до достижения температуры 310°C. Общая продолжительность анализа составила 22 мин. Масс-спектрометрическое детектирование осуществляли с применением тандемного масс-спектрометрического детектора Thermo TSQ Quantum XLS. Задержка на время элюирования растворителя – 4.25 мин. Температура инжектора – 300°C, температура переходной линии – 300°C, газ-носитель – гелий, скорость потока газ-носителя через колонку – 0.6 мл/мин, температура источника ионизации – 250°C, давление газа-мишени (аргона) в ячейке соударений – 1.5 мТорр. Данные обрабатывали с использованием программного пакета STATISTICA 10 (StatSoft).

Биологические образцы. В исследование были включены 10915 образцов мочи (4402 женских и 6513 мужских) профессиональных спортсменов в рамках программы допинг-контроля 2014 и 2015 гг. В изучаемой выборке 140 женских и 95 мужских проб мочи имели явные признаки деградации в соответствии с действующими критериями ВАДА. Пробы анализировали непосредственно после их доставки в лабораторию организацией, занимающейся пробоотбором, в соответствии с регламентом пробоотбора и транспортировки ВАДА.

Подготовка проб к анализу соответствовала рутинной пробоподготовке, используемой в практике антидопинговых лабораторий [7]. Непосредственно после получения образцов измеряли значения pH и плотности. Пробоподготовка для хромато-масс-спектрометрического исследования состояла из трех основных стадий: гидролиза, экстракции, дериватизации. К 3 мл мочи добавляли 50 мкл раствора дейтерированных внутренних стандартов (табл. 2), 1 мл фосфатного буферного

раствора (pH 6.4), содержащего 30 мкл β-глюкуронидазы *E. coli* K12 и дополнительный внутренний стандарт (ДВС) для оценки воспроизводимости и работоспособности оборудования (метилтестостерон с концентрацией 150 нг/мл) и инкубировали при 57°C. Экстракцию проводили 5 мл диэтилового эфира после добавления 1 мл карбонатного буферного раствора для придания раствору достаточно высокого pH. Для высаливания и более четкого разделения фаз вода–эфир в пробу добавляли 1.5 г сульфата натрия. После тщательного перемешивания в течение 10 мин и центрифугирования (5 мин при 4000 об/мин) пробу охлаждали в криостатической бане при –30°C, верхний слой (эфирный) переливали в чистые пробирки и упаривали досуха в твердотельном нагревателе при 70°C. После охлаждения в пробирку добавляли 50 мкл дериватирующего раствора (20 мг ДТТ и 15 мг NH₄I на 10 мл МСТФА) и инкубировали в течение 20 мин при 70°C. Реакционную смесь охлаждали и переносили в вials с силианизированной вставкой объемом 100 мкл.

Линейность диапазона определяемых концентраций позволяет воспользоваться методом ограничивающих растворов. В нижнем калибровочном диапазоне использовали калибратор “низкий”, соответствующий пределу количественного обнаружения (ПКО), в верхнем – “высокий” (табл. 3). Верификацию калибровки проводили с помощью образца контроля качества (КК), содержание аналитов в котором соответствовало рассчитанной на предварительной выборке (100 образцов) медиане маркеров стероидного профиля. Растворы калибраторов готовили на матрице мочи, очищенной от стероидов методом твердофазной экстракции. Допустимым считалось отклонение от приписанных значений КК до 3%. Рассчитанная стандартная неопределенность измерений (*u_c*) на уровне ПКО и 5-кратном ПКО приведена в табл. 3.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Поскольку процесс биodeградации образцов поливариантен и непредсказуем, при представлении данных целесообразно не только выделение доли подвергшихся деградации образцов относи-

Таблица 3. Калибраторы и стандартная неопределенность для маркеров стероидного профиля

Соединение	“Низкий” (ПКО), нг/мл	“Высокий”, нг/мл	КК, нг/мл	u_c ПКО, %	$u_c \times 5 \times$ ПКО, %
T	1	200	10	6.1	10.2
ЭпиT	1	200	10	15.2	11.7
5 α -Диол	2	250	30	27.5	19.2
5 β -Диол	5	800	80	12.6	11.2
Андро	200	10000	2200	23.6	13.5
Этио	200	10000	1800	24.7	11.4
ДГЭА	10	150	30	15.2	12.1
7 β -ОН-ДГЭА	10	150	25	14.3	11.3
5 β -Дион	1	50	2	13.3	11.4

Примечание: u_c – рассчитанная стандартная неопределенность измерений.

тельно общего числа проб, но и указание показателей в соответствии с критериями ВАДА, на основании которых они могут быть признаны деградировавшими (рис. 1).

Несмотря на то, что в рассматриваемой выборке доля деградировавших образцов невысока, общее количество проб велико и выборка является репрезентативной.

Очевидно, что объективная оценка процесса деградации образцов в отсутствие возможности введения ингибиторов деятельности микроорганизмов или иных компонентов, повышающих стабильность образцов, возможна исключительно с использованием эндогенных соединений, в частности, других андрогенных стероидов и их соотношений, присутствующих в образцах. К таким соединениям относятся 5 α -андростандион, тестостерон, андростерон и эпитестостерон, определение которых является обязательным для описания стероидного профиля человека, а также форместан.

На сегодняшний день известно большое количество различных статистических методов анализа, однако при их апробации на больших массивах данных могут возникать ложные корреляции и взаимосвязи между данными, что будет приводить к увеличению числа ложноположительных и ложноотрицательных результатов. Таким образом, интересно рассмотрение наиболее распространенных методов статистического анализа данных с целью выявления их ограничений и возможностей при работе с большими выборками данных.

В целях выявления отличий между образцами, подвергшимися деградации, и недеградированными пробами построили диаграммы рассеяния, представленные на рис. 2. Из этого рисунка невозможно сделать однозначный вывод о наличии корреляции между использованными показателями, однако на первый взгляд может показаться, что существует тренд, свидетельствующий об

уменьшении концентраций 5 α -диол и этиохола-нолона в ходе деградации. При детальном рассмотрении диапазона, соответствующего области, в которой лежит большая часть результатов анализа деградировавших проб, не удалось выявить явные взаимосвязи, поэтому приняли решение использовать статистические методы анализа.

Статистический анализ данных применяли для исследования влияния процесса деградации на профиль основных параметров, контролируемых ВАДА, выявления дополнительных признаков, по которым можно бы было судить о деградации проб, помимо уже установленных, и для оценки эффективности установления факта деградации образцов мочи при сочетании комбинаций различных параметров. Для выявления показателей, наиболее подверженных изменениям в процессе деградации, целесообразным представлялось применение метода главных компонент (МГК). Решение задачи установления факта деградации возможно с применением различных классификационных алгоритмов. Наиболее интересным представляется дискриминантный анализ, преимуществом которого является возможность “визуализации” эффективности разделения проб на группы посредством канонического анализа, предусматривающего разделение исследуемых групп в двумерном пространстве функций (канонических корней). Одним из условий применения предлагаемых методов статистической обработки данных является нормальность распределения параметров исследуемой выборки с допустимыми умеренными отклонениями. Поэтому на первоначальном этапе оценивали нормальность распределения исследуемых показателей, для чего использовали гистограммы и тест Колмогорова–Смирнова. Распределение основного массива показателей было близко к нормальному, однако имело положительную асимметрию. В случае показателей T, и отношений

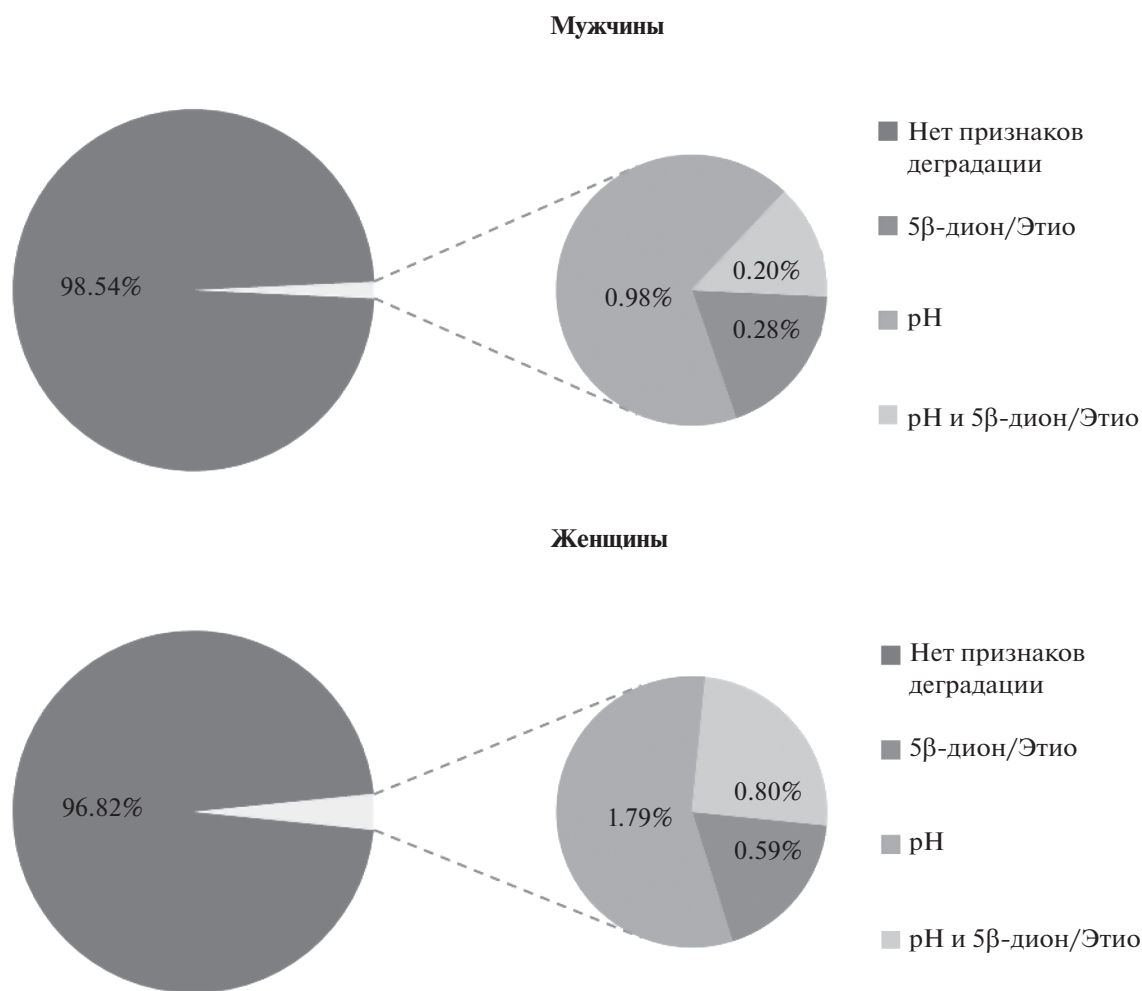


Рис. 1. Соотношение общего числа проанализированных образцов к количеству продеградировавших проб с указанием несоответствующего норме показателя.

Т/ЭпиТ и Андро/Т наблюдалось бинормальное распределение. Для нормализации данных предприняли попытку применить степенное преобразование Бокса–Кокса. Распределение большинства переменных удалось привести к нормальному, однако применение данного приема привело к искажению данных, существенному изменению корреляционных связей между показателями и значительному ухудшению производительности построенных моделей, вследствие чего массив данных подвергали статистической обработке без предварительной нормализации. Учитывая возможные ограничения, связанные с использованием описанных выше статических методов, для анализа исследуемого массива данных целесообразным представлялось применение более сложных методов, использование которых не предусматривает требований к распределению данных и ориентированных на обработку больших массивов данных, в частности, нейронных сетей.

МГК предусматривает факторный анализ, позволяющий оценить значимость факторов и их вклад в объяснение общей вариации посредством графиков “каменистой осыпи” (рис. 3). Как видно, основной вклад в объяснение общей вариации вносят первые четыре фактора, на долю которых приходится 68–69% общей вариации.

Наибольшие корреляции показателей с факторами наблюдались в случае следующих групп показателей: 1 группу составил ряд стероидов, обычно не используемых для выявления деградации; 2 группу – соотношения этих показателей; в 3 группу в мужской популяции вошло одно соотношение 5α-диол/ЭпиТ, в женской – 5α-диол/ЭпиТ и Т/ЭпиТ; в 4 группу вошли основные показатели, по которым на сегодняшний день оценивают деградацию пробы: rH и 5α-андростан-3,17-дион (табл. 4).

Для оценки эффективности выявления факта деградации строили классификационные модели

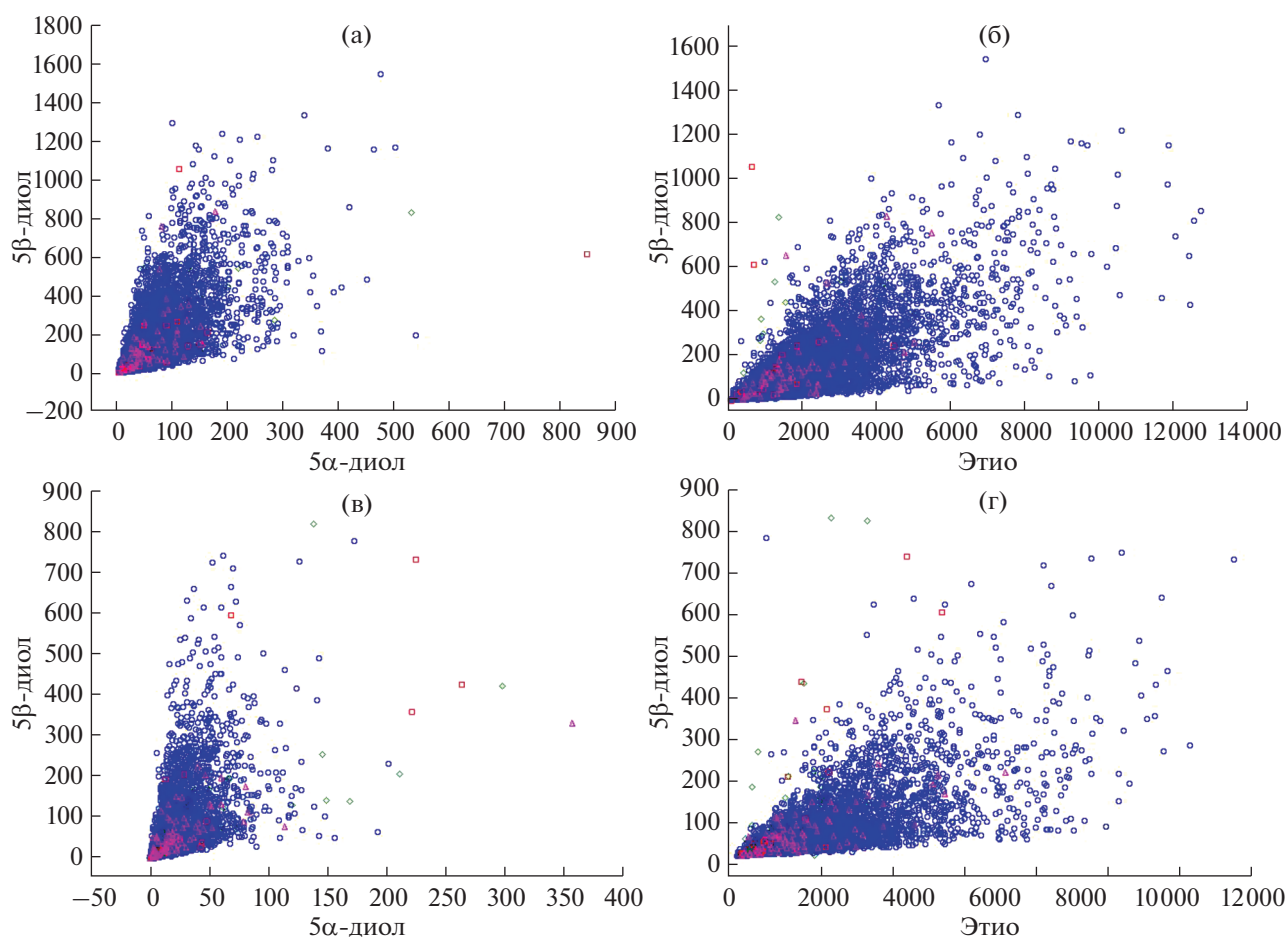


Рис. 2. Диаграммы рассеяния для мужских (а), (б) и женских (в), (г) проб: (○) – образцы, в которых не было найдено запрещенных ВАДА соединений, (□) – признаки деградации по отношению 5β-дион/этиохоланолон, (△) – признаки деградации по рН, (◇) – признаки деградации по рН и отношению 5β-дион/этиохоланолон.

с применением дискриминантного анализа. Модели были построены в три этапа: на первом этапе оценивали эффективность выявления факта деградации с применением стандартных параметров, по которым судят о деградации (рН и 5β-дион). Затем оценивали способность совокупности остальных показателей, мониторинг которых проводится в рамках контроля стероидного профиля, и их соотношений выявлять продеградировавшие пробы. На завершающем этапе строили модели, сочетающие комбинацию как основных, так и остальных показателей.

На рис. 4а, 4б представлена визуализация распределения проб в зависимости от наличия деградации и ее характера, построенная исключительно на основе параметров, по которым судят о деградации (рН и 5β-дион). Значения лямбды Уилкса позволяет оценить качество построенных моделей. Значения данного показателя изменяются в пределах от 0 до 1 и чем ближе значения к нулю, тем лучше качество модели. В случае мужчин значение лямбды Уилкса составило 0.298,

женщин – 0.482, что свидетельствует об относительно неплохом качестве модели. Если судить исключительно по правильности отнесения нормальных проб и деградированных вне зависимости от причин деградации, то в мужской популяции 3 пробы из 6418 нормальных были отнесены к деградированным по рН, и 4 из 18 проб деградированных по 5β-дион были отнесены к нормальным; в женской популяции из 4262 нормальных проб 12 было отнесено к деградированным по рН, а 14 из 32 проб, деградированных по 5β-дион, были отнесены к нормальным. Остальные ошибки классификации связаны с неправильным указанием причины деградации, что не столь существенно для решения общей поставленной задачи.

При рассмотрении всех параметров, за исключением основных (рН и 5β-дион), значение лямбды Уилкса для мужчин составило 0.904, а для женщин – 0.883, что свидетельствует о значительном ухудшении качества модели. Таким образом, можно сделать вывод, что выявление деградации на основании показателей стероидного профиля,

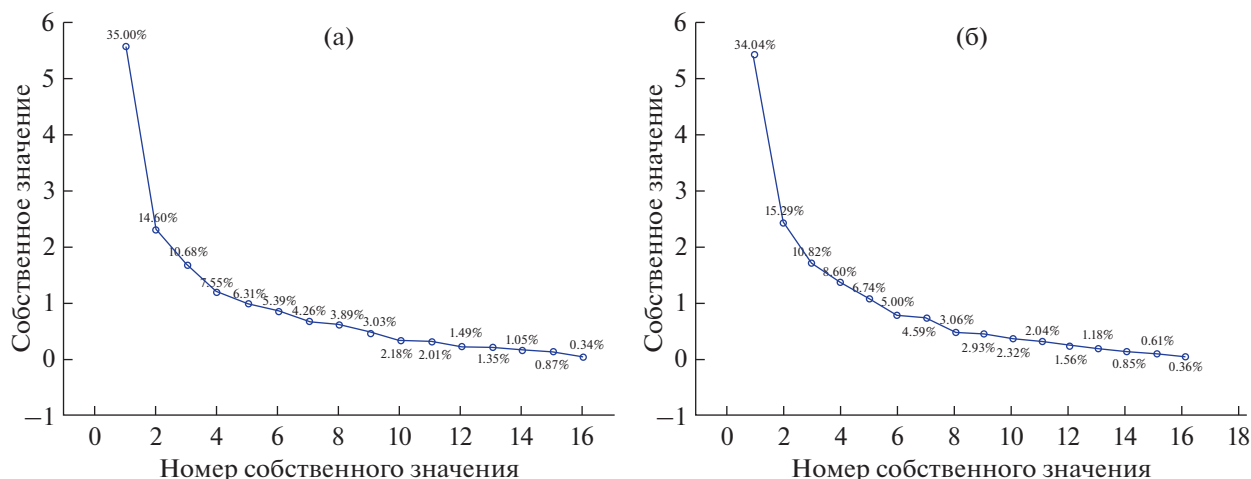


Рис. 3. Графики "каменистой осыпи": (а) – мужчины, (б) – женщины.

контролируемых ВАДА, без основных показателей (рН и 5β-дион) не представляется возможным, и деградация незначительно отражается на основных показателях (рис. 4в, 4г). Анализ значений параметров *p*-критерия, лямбды Уилкса и частной лямбды в модели позволил выявить компоненты, вклад которых в классификацию наиболее значимый (табл. 5).

На основании полученных результатов в модели, построенные на рН и 5β-дион, добавили значимые параметры, выявленные на предыдущем

этапе для популяций мужчин и женщин. Наблюдали уменьшение значения параметра лямбды Уилкса по сравнению с моделью, построенной исключительно на рН и 5β-дион: оно составило 0.284 для мужчин и 0.467 для женщин, однако улучшения оказались незначительными (рис. 4д, 4е). При оценке значимости факторов их вклад может быть неочевиден, поэтому из исходной модели исключили заведомо весомые показатели, что позволило выявить наиболее чувствительные к процессу деградации показатели. Однако для по-

Таблица 4. Корреляции исследуемых параметров с факторами

Вещество	Мужчины				Женщины			
	фактор 1	фактор 2	фактор 3	фактор 4	фактор 1	фактор 2	фактор 3	фактор 4
Андро	-0.874080*	0.277442	0.001408	0.075332	0.883480	0.256011	-0.026340	-0.078250
Этио	-0.849720	-0.116440	-0.222740	-0.184620	0.846070	-0.130510	0.229919	0.034853
Т	-0.736210	-0.266050	-0.025060	0.339401	0.772550	-0.307950	-0.065520	-0.114790
ЭпиТ	-0.70760	0.246876	-0.236140	-0.186880	0.667930	0.104432	0.466914	0.087107
5α-Диол	-0.719460	-0.000250	0.571458	-0.116210	0.816258	0.139766	-0.306050	0.166610
5β-Диол	-0.663680	-0.440420	0.103634	-0.260590	0.771130	-0.387840	-0.054300	0.164345
Ф	-0.761980	0.099992	-0.200990	-0.031220	0.742509	0.179485	0.095076	-0.020910
ДГЭА	-0.897110	0.124312	-0.045740	0.127471	0.848224	0.236028	-0.039510	-0.092550
7β-ОН-ДГЭА	-0.711090	0.203893	-0.041800	0.117745	0.655278	0.327430	-0.051580	-0.094420
Андро/Т	0.074840	0.621756	-0.029620	-0.352930	-0.157000	0.619519	0.145629	0.226926
Андро/Этио	0.001160	0.662727	0.370324	0.417243	-0.066510	0.633306	-0.398570	-0.177820
5α-Диол/5β-диол	0.065076	0.825474	0.340687	0.174148	-0.165970	0.882408	-0.114390	0.060737
5α-Диол/ЭпиТ	-0.163250	-0.224110	0.827473	-0.019810	0.121551	0.036511	-0.841480	0.082932
рН	0.136145	-0.126800	0.348335	-0.447980	-0.102710	-0.138090	-0.076370	0.611784
5β-Дион	-0.104060	-0.073510	0.356089	-0.466440	0.101780	-0.079890	-0.233500	0.736761
Т/ЭпиТ	-0.055800	-0.620230	0.237106	0.570536	0.080658	-0.437760	-0.601540	-0.310330

* Жирным шрифтом выделены наиболее значимые корреляции.

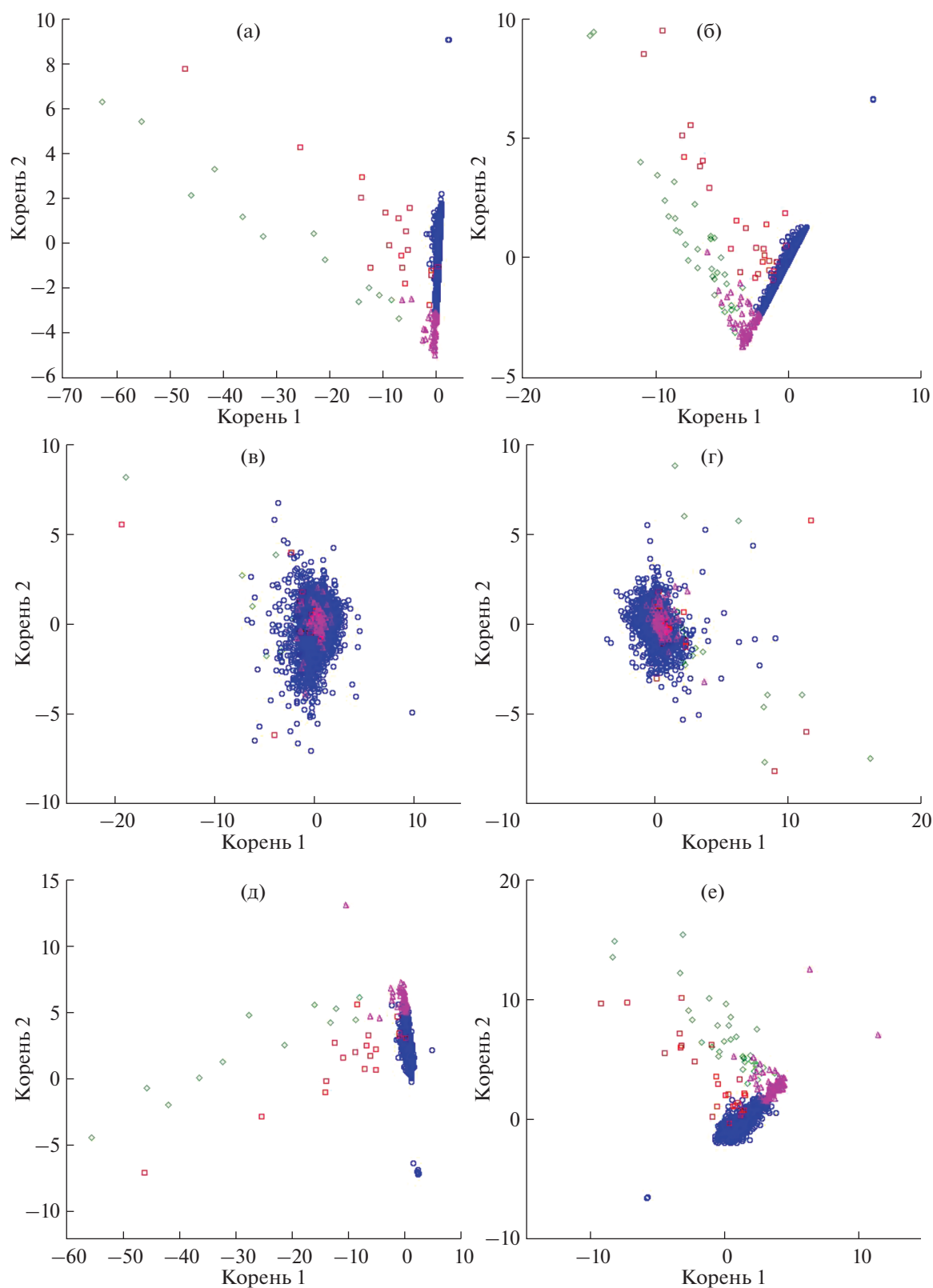


Рис. 4. Диаграммы рассеяния канонических значений для образцов мочи в зависимости от наличия деградации и ее характера на основании различных параметров: 1) рН и 5β-диол у мужчин (а) и женщин (б); 2) все параметры, кроме рН и 5β-диол у мужчин (в) и женщин (г); 3) все параметры у мужчин (д) и женщин (е): (○) – образцы, в которых не было найдено запрещенных ВАДА соединений, (□) – признаки деградации по отношению 5β-диол/этиохоланолон, (Δ) – признаки деградации по рН, (◇) – признаки деградации по рН и отношению 5β-диол/этиохоланолон.

Таблица 5. Уровни значимости переменных, входящих в классификационные модели

Показатель	Мужчины				Показатель	Женщины			
	лямбда Вилкса	частная лямбда	F-remove	p-критерий		лямбда Вилкса	частная лямбда	F-remove	p-критерий
Андро	0.284	0.999	1.7	0.162	Андро	0.462	0.991	13	0.000
T	0.284	0.999	1.3	0.273	T	0.460	0.996	6.5	0.000
ЭпиT*	0.285	0.997	7.4	0.000	5α-диол	0.462	0.990	15	0.000
5α-Диол	0.284	0.999	1.5	0.221	5β-диол	0.461	0.993	9.8	0.000
5β-Диол	0.286	0.993	14	0.000	7β-ОН-ДГЭА	0.459	0.997	3.7	0.012
T/ЭпиT	0.285	0.997	6.2	0.000	T/ЭпиT	0.458	0.998	2.9	0.032
5α-Диол/ЭпиT	0.292	0.973	59	0.000	Андро/Этио	0.460	0.995	6.7	0.000
pH	0.329	0.864	341	0.000	5α-диол/ЭпиT	0.459	0.996	5.8	0.001
5β-Дион	0.773	0.368	3728	0.000	pH	0.629	0.727	547	0.000
					5β-дион	0.623	0.735	527	0.000

* Жирным шрифтом выделены наиболее значимые показатели и переменные в классификационной модели.

лучения комплексной оценки к вновь выявленным критериям добавили уже известные ранее pH и 5β-диол с последующим построением итоговой модели (табл. 5). Интересно, что в мужской популяции параметры Андро, T и 5β-диол теряют свою значимость на фоне pH и 5β-дион, а также значимость всех остальных параметров на фоне pH и 5β-дион значительно снижается.

На основании полученных из дискриминантного анализа данных построили по две нейросетевые модели типа многослойный перцептрон с одним скрытым слоем для каждой из исследуемых популяций: включающие только основные показатели pH и 5β-дион и включающие основные показатели в совокупности с наиболее значимыми показателями для мужской и женской популяций. Для обучения нейронных сетей применяли алгоритм Бройдена–Флетчера–Гольдфарба–Шанно. Выборку предварительно разделили на три части: обучающую, контрольную и тестовую (60, 10 и 30%). Оптимизация параметров модели осуществляется при обучении, где для проверки точности обучения используется контрольная выборка. После завершения процесса обучения сеть тестируется на независимом тестовом наборе данных. Исследованы различные топологии нейронных сетей и выбраны параметры моделей, обеспечивающие наилучшую производительность: для женской популяции наилучшими оказались сети с логистической функцией ак-

тивации нейронов скрытого слоя и softmax (обобщение логистической функции для многомерного случая) – выходного слоя, для мужской популяции – гиперболическая функция активации нейронов скрытого слоя и softmax – выходного. Количество нейронов на выходном слое соответствовало количеству исследуемых групп (4). Количество нейронов в скрытом слое варьировалось для каждой модели и составило для основных параметров в мужской популяции – 4, в женской – 3, для основных параметров с добавлением наиболее значимых показателей в мужской популяции – 9, в женской – 7.

В табл. 6 приведены матрицы классификаций моделей, построенных с применением дискриминантного анализа и нейронных сетей на всем массиве данных. По горизонтали отображено количество случаев, задействованных в построении модели, по вертикали – количество случаев, отнесенных в группу соответствующим классификационным алгоритмом.

Задача верного установления факта деградации является основной, поэтому в первую очередь обращали внимание на ошибки, связанные с отнесением продеградированных проб к чистым и наоборот. С данной точки зрения добавление параметров в модель приводит к увеличению таких ошибок при использовании любого метода классификации, что негативно отражается на решении основной задачи, однако в женской попу-

Таблица 6. Матрицы классификаций моделей, построенных с использованием дискриминантного анализа и нейронных сетей

Переменные, вошедшие в модель	Группы (подвергшиеся деградации или пригодные для анализа)	Результаты классификации проб							
		мужчины				женщины			
		пригодные для анализа	5 β -диол	5 β -диол + рН	рН	пригодные для анализа	5 β -диол	5 β -диол + рН	рН
Дискриминантный анализ									
рН + 5 β -диол	Пригодные для анализа	6415	0	0	3	4250	0	0	12
	5 β -диол	4	11	2	1	14	3	8	1
	5 β -диол + рН	0	5	8	0	0	5	16	14
	рН	0	1	0	63	0	0	1	78
Все значимые переменные	Пригодные для анализа	6410	0	0	8	4253	0	0	9
	5 β -диол	5	11	2	0	14	5	7	0
	5 β -диол + рН	0	5	8	0	0	2	18	15
	рН	7	1	0	56	5	1	1	72
Нейронные сети									
рН + 5 β -диол	Пригодные для анализа	6418	0	0	0	4262	0	0	0
	5 β -диол	4	13	1	0	4	22	0	0
	5 β -диол + рН	0	0	12	1	0	0	33	2
	рН	0	1	1	62	1	0	10	68
Все значимые переменные	Пригодные для анализа	6415	2	0	1	4262	0	0	0
	5 β -диол	4	11	3	0	6	20	0	0
	5 β -диол + рН	0	0	13	0	0	1	34	0
	рН	1	0	1	62	2	0	0	77

ляции можно наблюдать улучшение качества классификации относительно причины деградации. При сравнении используемых алгоритмов классификации видно, что нейронные сети позволяют правильно классифицировать большее число наблюдений, а также наблюдается большая устойчивость модели при добавлении других факторов (например: количество неправильно классифицированных чистых и продеградированных проб увеличивается с 7 до 20 в случае дискриминантного анализа, а при использовании нейронных сетей – с 4 до 8), однако дискриминантный анализ позволяет визуализировать качество дискриминации, что удобно для представления данных. Использование комбинации статистических методов позволило наиболее полно проанализировать данные с точки зрения группи-

ровки исследуемых показателей, классификации данных на основе разных показателей и визуализации полученных результатов.

* * *

Таким образом, деградация пробы сказывается не только на рН и концентрации 5 β -диона, но и других компонентах, представленных в табл. 5, однако использовать их в качестве маркеров деградации пробы не представляется возможным, так как их изменение является недостаточно значимым, чтобы надежно описывать протекающий процесс как у мужчин, так и у женщин по сравнению с основными весомыми показателями. Следует отметить, что исключение их из модели приводит к ухудшению ее прогностической способ-

ности. Значимость соотношения Андро/Этио в табл. 5 может быть обусловлена высоким уровнем значимости в модели самого Андро. Представленные модели также позволяют визуализировать и эффективно классифицировать результаты, особенно с применением дискриминантного анализа, в то время как применение нейросетей может способствовать автоматизации и повышению надежности процесса потокового анализа.

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 21-13-00240) с использованием оборудования ЦКП “Эколого-аналитический центр” ФГБОУ ВО КубГУ.

СПИСОК ЛИТЕРАТУРЫ

- World Anti-Doping Agency. Urine Sample Collection Guidelines. https://www.wada-ama.org/sites/default/files/resources/files/wada_guidelines_urine_sample_collection_2014_v1.0_en.pdf (2014) (28.12.2020).
- Tsivou M., Livadara D., Georgakopoulos D.G., Koupparis M.A., Atta-Politou J., Georgakopoulos C.G. Stabilization of human urine doping control samples: II. Microbial degradation of steroids // *Anal. Biochem.* 2009. V. 388. № 1. P. 146.
- Bon-chu C., Meng-Chun H. Androgen biosynthesis and degradation // *Androgens and Androgen Receptor.* 2002. P. 1.
- Tsivou M., Georgakopoulos D.G., Dimopoulou H.A., Koupparis, M.A., Atta-Politou J., Georgakopoulos C.G. Stabilization of human urine doping control samples: A current opinion // *Anal. Bioanal. Chem.* 2011. V. 401. № 2. P. 553.
- Tsivou M., Giannadaki E., Hooghe F., Roels K., Van Gansbeke W., Garribba F., Lyris E., Deventer K., Mazzarino M., Donati F., Georgakopoulos D.G., Van Eenoo P., Georgakopoulos C.G., de la Torre X., Botrè F. Doping control container for urine stabilization: a pilot study // *Drug Test. Anal.* 2017. V. 9. № 5. P. 699.
- Chiang Y.R., Wei S.T., Wang P.H., Wu P.H., Yu C.P. Microbial degradation of steroid sex hormones: implications for environmental and ecological studies // *Microb. Biotechnol.* 2020. V. 13. № 4. P. 926.
- Sobolevsky T., Krotov G., Dikunets M., Nikitina M., Mochalova E., Rodchenkov G. Anti-doping analyses at the Sochi Olympic and Paralympic Games 2014 // *Drug Test. Anal.* 2014. V. 6. № 11–12. P. 1087.
- Georgakopoulos C. Detecting – the 2004 Athens Olympic Games // *Drug Test. Anal.* 2020. V. 12. № 5. P. 641.
- Baltazar-Martins G., Del Mar Plata M., Muñoz-Guerra J., Argaya J., Del Coso J. Chronicle of an unintentional MDMA intoxication of a field hockey team to prevent its qualification for the Beijing 2008 Olympics: A case report // *Drug Test. Anal.* 2020. V. 12. № 5. P. 647.
- Pereira H.M.G., Sardela V.F., Padilha M.C., Mirotti L., Casilli A., de Oliveira F.A., de Albuquerque Cavalcanti G., Rodrigues L.M.L., de Araujo A.L.D., Levy R.S., Teixeira P.A.C., de Oliveira F.A.G., Duarte A.C.G., Carneiro A.C.D., Evaristo J.A.M., Dos Santos G.R.C., da Costa G.C.V., de Lima Castro F., Nogueira F.C.S., Scalco F.B., Pizzatti L., de Aquino Neto F.R. Doping control analysis at the Rio 2016 Olympic and Paralympic Games // *Drug Test. Anal.* 2017. V. 9. № 11–12. P. 1658.
- Bredenhöft M., Baginski R., Parr M.K., Thevis M., Schänzer W. Investigations of the microbial transformation of cortisol to prednisolone in urine samples // *J. Steroid Biochem. Mol. Biol.* V. 129. 2012. № 1–2. P. 54.
- Thevis M., Geyer H., Mareck U., Sigmund G., Henke J., Henke L., Schänzer W. Detection of manipulation in doping control urine sample collection: A multidisciplinary approach to determine identical urine samples // *Anal. Bioanal. Chem.* 2007. V. 388. № 7. P. 1539.
- Mazzarino M., Abate M.G., Alocci R., Rossi F., Stinchelli R., Molaioni F., de la Torre X., Botrè F. Urine stability and steroid profile: Towards a screening index of urine sample degradation for anti-doping purpose // *Anal. Chim. Acta.* 2011. V. 683. № 2. P. 221.
- De Clercq N., Julie V.B., Croubels S., Delahaut P., Vanhaecke L. A validated analytical method to study the long-term stability of natural and synthetic glucocorticoids in livestock urine using ultra-high performance liquid chromatography coupled to orbitrap-high resolution mass spectrometry // *J. Chromatogr. A.* 2013. V. 1301. P. 111.
- Lippi G., Mattiuzzi C., Banfi G. Controlling sources of preanalytical variability in doping samples: Challenges and solutions // *Bioanalysis.* 2013. V. 5. № 12. P. 1571.
- Aoki K., Tanaka H., Ueki M. DNA typing for personal identification of urine after long-term preservation for testing in doping control // *Drug Test. Anal.* 2017. V. 9. № 8. P. 1116.
- Pires N.M.M., Dong T., Yang Z., Santos S.M.B. A Fluorescence sensing method with reduced DNA typing and low-cost instrumentation for detection of sample tampering cases in urinalysis // *Ann. Biomed. Eng.* 2020. V. 48. № 2. P. 644.
- Kuوران T., Saugy M., Baume N. Confounding factors and genetic polymorphism in the evaluation of individual steroid profiling // *Br. J. Sports Med.* 2014. V. 48. № 10. P. 848.
- De Wilde L., Roels K., Van Renterghem P., Van Eenoo P., Deventer K. Steroid profiling in urine of intact glucuronidated and sulfated steroids using liquid chromatography-mass spectrometry // *J. Chromatogr. A.* 2020. V. 1624. Article 461231.
- Martinez-Brito D., Notarianni M.L., Iannone M., de la Torre X., Botrè F. Validation of steroid sulfates deconjugation for metabolic studies. Application to human urine samples // *J. Pharmacol. Toxicol. Methods.* 2020. V. 106. Article 106938.
- Naldi A. P., Fayad P.B., Prévost M., Sauvé S. Analysis of steroid hormones and their conjugated forms in water and urine by on-line solid-phase extraction coupled to liquid chromatography tandem mass spectrometry // *Chem. Central J.* V. 10. 2016. № 1. P. 1.
- Cha E., Kim S., Kim H.J., Lee K.M., Kim K.H., Kwon O.S., Lee J. Sensitivity of GC-EI/MS, GC-EI/MS/MS, LC-ESI/MS/MS, LC-Ag + CIS/MS/MS, and GC-ESI/MS/MS for analysis of anabolic steroids in doping control // *Drug Test. Anal.* 2015. V. 7. № 11–12. P. 1040.

23. *Andersen D. W., Linnet K.* Screening for anabolic steroids in urine of forensic cases using fully automated solid phase extraction and LC-MS-MS // *J. Anal. Toxicol.* 2014. V. 38. № 9. P. 637.
24. *Wozniak B., Matraszek-Zuchowska I., Zmudzki J.* LC-MS/MS fast analysis of androgenic steroids in urine // *Anal. Bioanal. Chem.* 2012. V. 403. № 10. P. 2965.
25. *Fabregat A., Pozo O.J., Marcos J., Segura J., Ventura R.* Use of LC-MS/MS for the open detection of steroid metabolites conjugated with glucuronic acid // *Anal. Chem.* 2013. V. 85. № 10. P. 5005.
26. *Badoud F., Grata E., Boccard J., Guillaume D., Veuthey J.L., Rudaz S., Saugy M.* Quantification of glucuronidated and sulfated steroids in human urine by ultra-high pressure liquid chromatography quadrupole time-of-flight mass spectrometry // *Anal. Bioanal. Chem.* 2011. V. 400. № 2. P. 503.
27. *Matysik S., Schmitz G.* Determination of steroid hormones in human plasma by GC-triple quadrupole MS // *Steroids.* 2015. V. 99 B. P. 151.
28. *Forsdahl G., Zanitzer K., Erceg D., Gmeiner G.* Quantification of endogenous steroid sulfates and glucuronides in human urine after intramuscular administration of testosterone esters // *Steroids.* 2020. V. 157. Article 108614.
29. *McDonald J. G., Matthew S., Auchus R. J.* Steroid profiling by gas chromatography-mass spectrometry and high-performance liquid chromatography-mass spectrometry for adrenal diseases // *Horm. Cancer.* 2011. V. 2. № 6. P. 324.