

УДК 617.711+311

DOI: 10.24884/1682-6655-2022-21-4-84-93

Н. В. КОРНЕЕВА¹, В. П. ЛЕОНОВ², К. В. ЖМЕРЕНЕЦКИЙ¹

Возможности дискриминантного анализа и логистической регрессии в изучении параметров микроциркуляции на примере биомикроскопии бульбарной конъюнктивы (лекция)

¹ Федеральное государственное бюджетное образовательное учреждение высшего образования «Дальневосточный государственный медицинский университет» Министерства здравоохранения Российской Федерации, г. Хабаровск, Россия 680000, Россия, г. Хабаровск, ул. Муравьева-Амурского, д. 35

² НЦ «БИОСТАТИСТИКА», г. Томск, Россия

634050, Россия, г. Томск, пр. Ленина, д. 36

E-mail: Gladkova1982@mail.ru

Статья поступила в редакцию 28.06.22 г.; принята к печати 05.09.22 г.

Резюме

Статья посвящена актуальной теме – корректному использованию современных статистических методик в анализе данных. Понимая сложность выбора статистического метода анализа собственных данных, его проведения и корректной трактовки, подчеркивается важность взаимодействия ученого-медика (биолога) с профессионалами по биостатистике. В статье в доступной форме расставлены акценты на правильном представлении собственных результатов в табличном и иллюстративном материале, напоминая о различиях между нормальным и ненормальным распределениями количественных признаков в совокупности. Приводятся примеры использования многомерных методов – дискриминантного анализа и логистической регрессии. На примере собственных результатов изучения микроциркуляции методом биомикроскопии бульбарной конъюнктивы, авторы показывают возможности этих сложных и нечасто используемых исследователями статистических методик. В выводах подчеркивается необходимость сотрудничества исследователя-медика (биолога) со специалистами по биомедицинской статистике для корректного применения статистических методик, максимального извлечения полезной информации из собственных первичных данных, а также помощи в интерпретации полученных результатов. Авторы приводят обширный список статистической литературы, изучение которой может помочь ученым-медикам разобраться в этой сложной науке.

Ключевые слова: биометрика, статистика, статистический анализ, биомикроскопия конъюнктивы, микроциркуляция, дискриминантный анализ, логистическая регрессия

Для цитирования: Корнеева Н. В., Леонов В. П., Жмеренецкий К. В. Возможности дискриминантного анализа и логистической регрессии в изучении параметров микроциркуляции на примере биомикроскопии бульбарной конъюнктивы (лекция). Регионарное кровообращение и микроциркуляция. 2022;21(4):84–93. Doi: 10.24884/1682-6655-2022-21-4-84-93.

UDC 617.711+311

DOI: 10.24884/1682-6655-2022-21-4-84-93

N. V. KORNEEVA¹, V. P. LEONOV², K. V. ZHMERENETSKY¹

The potential of modern statistical analysis for the study of microcirculation parameters on the example of the bulbar conjunctiva biomicroscopy (lecture)

¹ The Far Eastern State Medical University, Khabarovsk, Russia

35, Murav'eva-Amurskogo str., Khabarovsk, Russia, 680000

² Science Center Biostatistics, Tomsk, Russia

36, Lenin str., Tomsk, Russia, 634050

E-mail: Gladkova1982@mail.ru

Received 28.06.22; accepted 05.09.22

Summary

The article is devoted to the correct use of modern statistical methods in data analysis. Considering the complexity of choosing and implementation of statistical method for data analysis and correct interpretation of the results we emphasize the importance of interaction between a medical scientist (biologist) and biostatistician. The article in a clear way emphasizes the correct presentation of the results as a table and graphic form, recalls the differences between normal and other types of distribution of quantitative characteristics in the aggregate. Examples of using the multivariate methods are given in detail: discriminant analysis and logistic regression. The results of the original bulbar conjunctiva microcirculation study by the biomicroscopy were provided an example of the potential of these complex and infrequently used statistical methods. The conclusions emphasize the need for cooperation between a medical researcher (biologist) and biomedical statistician for the correct application of

statistical methods, the maximum extraction of useful information from the primary data, as well as assistance in interpreting the results obtained. The authors provide an extensive list of statistical literature that studying can help medical scientists understand this complex science.

Keywords: *biometrics, statistics, statistical analysis, biomicroscopy of the conjunctiva, microcirculation, discriminant analysis, logistic regression*

For citation: Korneeva N. V., Leonov V. P., Zhmerenetsky K. V. The potential of modern statistical analysis for the study of microcirculation parameters on the example of the bulbar conjunctiva biomicroscopy (lecture). *Regional hemodynamics and microcirculation*. 2022;21(4):84–93. Doi: 10.24884/1682-6655-2022-21-4-84-93.

Введение

Начать данную статью хочется словами Гюстава Флобера: «То, что понимают плохо, часто стараются объяснить с помощью слов, которые не понимают». Часто написание раздела, касающегося описания статистических методов анализа в статьях и диссертационных работах, представляется для исследователей наиболее сложной задачей. Многие до сих пор применяют абстрактные формулировки с упоминанием достоверности, критерия Стьюдента, анализом вариационных рядов и т. д. А между тем данный раздел позволяет понять, какую совокупность анализировали авторы, насколько полными, корректными и оригинальными являются сведения, извлеченные из первичного материала, представленные в работе, а также оценить уровень профессионализма ученого.

Материалы, представленные ниже, позволят читателю разобраться в сложностях статистического анализа и, мы надеемся, подтолкнет исследователей к внимательному ознакомлению и изучению приставленного списка литературы, где, в том числе, имеется ссылка на нашу монографию «Конъюнктивальная биомикроскопия: методология анализа» [1], в которой третья глава посвящена корректному использованию статистических методик применительно к изучению микроциркуляторных параметров. Считаем, что подобная самостоятельная работа будет полезна читателям для улучшения своих знаний в области биомедицинской статистики.

Представление данных

Первичные данные, получаемые при любых способах исследования микроциркуляции (МЦ), можно разделить на две большие группы – количественные и качественные признаки. К первым, например, относятся средний диаметр микрососудов, их число на единицу площади и т. д. Качественные характеристики, например, отражают присутствие или отсутствие какого-либо признака, степень его выраженности в баллах и т. д.

Корректное представление этих признаков в таблицах имеет свои особенности, позволяя грамотным читателям получать дополнительную информацию о представляемом явлении еще до изучения конкретных данных статистического анализа. Качественные характеристики представляются в виде абсолютных цифр, в скобках – процент.

Для корректного представления количественных признаков необходимо оценить их распределение в анализируемой совокупности. Здесь возможны два варианта: нормальное распределение (Гаусса, или Гаусса – Лапласа) и ненормальное распределение.

Исследователю-медику (биологу) следует знать, какие статистические критерии нужно использовать для проверки гипотезы нормальности распределения. Обращаем внимание читателя на то, что проверять нормальность распределения следует для каждой совокупности признаков в каждой из сравниваемых групп. Реальное число таких статистических критериев по проверке нормальности распределения, анализируемого количественного признака, гораздо более 10. Например, наиболее популярны такие критерии, как модифицированный критерий χ^2 , Локка – Спурье (Locke – Spurrer), Шапиро – Уилка (Shapiro – Wilk), Колмогорова – Смирнова (Kolmogorov – Smirnov), Крамера – Мизеса (Cramer – von Mises), Андерсона – Дарлинга (Anderson – Darling) и т. д. [2–6]. При этом в разных статистических программах заложены разные наборы критериев проверки нормальности распределения. Интерпретация проста: при значениях достигнутого уровня значимости менее 0,05 гипотеза о нормальном распределении отвергается. То есть при $p < 0,05$ распределение ненормальное.

Таким образом, важно корректно представлять данные в таблицах и тексте, что позволит читателям понять, подчиняется исследуемая совокупность признаков нормальному или ненормальному распределению.

При нормальном распределении необходимо использовать запись в виде: $M \pm SD / M \pm \sigma$, M – среднее значение, SD / σ (стандартное отклонение).

При ненормальном распределении следует записывать данные в виде: $Me (25–75 \%)$, где первая цифра – медиана (Me – представляет собой середину распределения и делит ее на две половины, или второй квартиль 50 % (Q_2)), в скобках указан интерквартильный интервал с размахом от 25 % (Q_1) до 75 % (Q_3) [7, 8]. При нормальном распределении для сравнения средних количественных показателей необходимо использовать параметрические методы анализа, при ненормальном – непараметрические.

Анализ данных

В данной статье решено представить читателям собственный опыт применения двух методов многомерного анализа, которые нечасто встречаются в статьях и диссертационных исследованиях. На примере своих данных мы покажем, как можно продуктивно использовать эти методы в анализе МЦ-параметров, полученных с помощью биомикроскопии бульбарной конъюнктивы (БМБК).

Первый – дискриминантный анализ, который позволяет провести комплексный анализ набора количественных признаков, т. е. установить, какие из

Таблица 1

Классификационная матрица

Table 1

Classification matrix

Группа наблюдения	Строки: фактически наблюдаемые классификации. Столбцы: предсказанные классификации			
	процент правильного предсказания	1-я группа (прекратили курить)	2-я группа (курят)	3-я группа (не курят)
1-я (прекратили курить)	97,1429	34	1	0
2-я (курят)	100,0000	0	16	0
3-я (не курят)	100,0000	0	0	29
Общий процент	98,7500	34	17	29

Таблица 2

Показатели стандартизованных коэффициентов признаков, включенных в дискриминантную функцию

Table 2

Standardized feature coefficients included in the discriminant function

Включенные признаки	Стандартизованные коэффициенты дискриминантной функции	
	ось x	ось y
Средние значения артериоло-венулярного коэффициента	-0,63247	-0,334143
Не курит, лет	0,31861	-0,549810
Число сигарет в день	1,63145	-0,366189
Число капилляров на конъюнктиве левого глаза	-0,31878	-0,407164
Распространенность агрегации в центральной зоне бульбарной конъюнктивы	0,34974	0,381330
Стаж курения	1,27083	-0,320127
Индекс пачка/лет	-2,00124	0,455481
Средние диаметры венул (по обоим глазам)	-0,22022	-0,391717

переменных вносят свой вклад в дискриминацию между совокупностями.

В наших исследованиях данный метод использовался в изучении МЦ и микроциркуляторного русла (МЦР) у прекративших курить, курящих и некурящих молодых лиц. Эти данные были опубликованы в третьем номере «Дальневосточного медицинского журнала» за 2016 г. в статье «Оценка состояния микроциркуляторного русла и микроциркуляции у практически здоровых людей молодого возраста, прекративших курение табака» [9]. К этому времени нами были накоплены данные о динамике состояния МЦР и МЦ у прекративших курить молодых лиц, полученные с помощью простых статистических методик сравнения средних значений количественных признаков в группах прекративших курить, курящих и некурящих лиц. Однако перед нами встали вопросы: «Являются ли наши результаты и выводы объективными? Не являются ли они результатом искусственно выделенных нами отдельных признаков и групп? Не повлияют ли на полученные результаты количественные признаки, которые ранее не были включены в анализ и не связывались нами с микроциркуляцией?». Все это послужило поводом к применению метода дискриминантного анализа, который позволил учесть в статистической процеду-

ре все количественные признаки, зафиксированные в базе данных для каждого индивидуума, включая не только параметры МЦР и МЦ, но и анамнестические и антропометрические данные. Важно было выяснить, какие количественные переменные различают три анализируемые группы (прекративших курить, курящих и некурящих), и увидеть на построенных в двух дискриминантных осях графиках расположение этих групп сравнения. Из всего перечня количественных признаков дискриминантный анализ позволяет выявить те, которые максимально связаны с различиями наблюдений, заранее разделенных на группы, и повторно их переклассифицировать. Чем ближе процент переклассификации к 100 %, тем сильнее связаны группирующий признак и количественные признаки, вошедшие в дискриминантные функции. В результате в одной группе оказываются наблюдения, наиболее близкие по своим свойствам, тогда как межгрупповое отличие максимизируется [2, 10–13].

Другими словами, дискриминантный анализ, во-первых, позволил ответить на вопрос о действительном существовании различий МЦ-параметров между искусственно сформированными группами на основании анамнеза курения; во-вторых, ответить на вопрос, какие из МЦ-признаков являются самыми

ценными, информативными и объективными для обследования с помощью БМБК бывшего курильщика, чтобы сделать вывод о том, что он отказался от курения.

В первоначальный анализ включили все 27 количественных признаков, характеризующих МЦР и МЦ, данные анамнеза и антропометрические сведения.

В результате получена дискриминантная функция, для группирующего признака «Группы наблюдения» состоящая из 8 признаков: «стаж курения», «не курит (лет)», «число сигарет в день», «индекс пачка/лет», «число капилляров на конъюнктиве левого глаза», «средние диаметры венул (по обоим глазам)», «средние значения артериоло-венулярного коэффициента» и «распространенность агрегации в центральной зоне бульбарной конъюнктивы». Остальные 19 признаков не вошли в дискриминантную функцию, поскольку имели достигнутый уровень значимости выше 5 %. В табл. 1 приведена классификационная матрица, которая показывает, что к 1-й группе из 35 наблюдений отнесены – 34, ко 2-й группе из 16 – 16, и к 3-й группе – также 100 % наблюдений. Общий процент правильной переклассификации составил 98,75 %.

Важно понимать, какие из 8 включенных в функцию количественных признаков вносят больший вклад в правильную переклассификацию, т. е. по каким признакам в наибольшей степени имеются различия между тремя изучаемыми группами. Для этого оценивали коэффициенты дискриминантных функций по двум осям ординат, данные о которых приведены в табл. 2.

Данные табл. 2 использованы для построения графика, позволяющего увидеть взаимное расположение трех анализируемых групп на основании созданной дискриминантной функции (рис. 1).

Согласно данным рис. 1, все три исследуемые группы на графике занимают свою область. 3-я группа – «некурящие» – находится поодаль от двух других по горизонтальной оси X таким образом, что между ней и 1-й и 2-й группами появляется свободное пространство, тогда как по вертикальной оси Y выше всего располагается 2-я группа – «курящие», группа 3 – «некурящие» – находится в нижней части графика, а группа «прекративших курить» отдалается по оси Y от курящих и имеет большую площадь перекрытия с некурящими. Полученные результаты, однако, не позволяли сделать вывод, что при отказе от курения происходит восстановление количественных параметров МЦР и МЦ, так как в данное пространственное распределение исследуемых групп по осям X и Y максимальные вклады вносят признаки, характеризующие процесс курения: «индекс пачка/лет», «число сигарет в день», «стаж курения», «не курит (лет)» (см. стандартизованные коэффициенты в табл. 2).

Для выявления взаимосвязи группирующего признака «группы наблюдения» и тех признаков, которые отражают только потенциальные последствия курения, а не его динамику, и интенсивность, на втором этапе из анализа исключили 4 признака, характери-

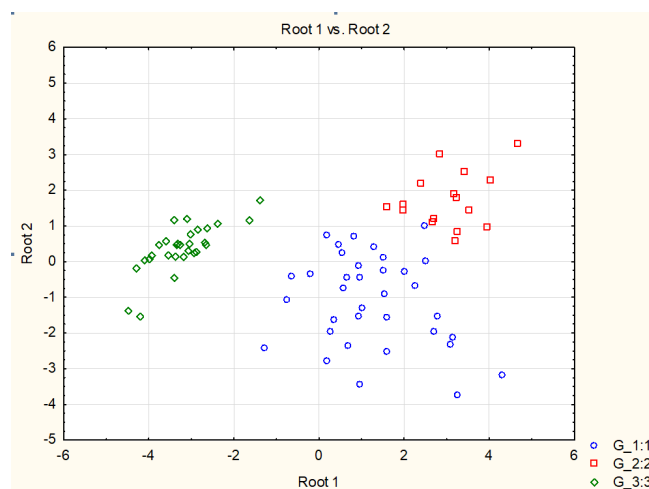


Рис. 1. График расположения трех анализируемых групп в осях ординат на основании дискриминантной функции: вертикальная ось (Root2) – отражает совокупность 8 признаков, характеризующих МЦР и анамнез курения (средние значения артериоло-венулярного коэффициента, не курит (лет), число сигарет в день, число капилляров на конъюнктиве левого глаза, распространенность агрегации в центральной зоне бульбарной конъюнктивы, стаж курения, индекс пачка/лет, средние диаметры венул (по обоим глазам)), жирным выделены признаки, вносящие максимальный вклад в смещение меток по этой оси; горизонтальная ось (Root1) отражает совокупность тех же 8 признаков, характеризующих МЦР и анамнез курения (средние значения артериоло-венулярного коэффициента, не курит (лет), число сигарет в день, число капилляров на конъюнктиве левого глаза, распространенность агрегации в центральной зоне бульбарной конъюнктивы, стаж курения, индекс пачка/лет, средние диаметры венул (по обоим глазам)), жирным выделены признаки, вносящие максимальный вклад в смещение меток по этой оси. G_1:1 – группа прекративших курить; G_2:2 – группа курящих; G_3:3 – группа некурящих

Fig. 1. Graph of the location of the three analyzed groups in the ordinate axes, based on the discriminant function: the vertical axis (Root2) – reflects a set of 8 features characterizing MCB and smoking history (mean arteriolo/venular ratio, no smoke (years), number of cigarettes per day, number of capillaries of the left eye of the conjunctiva surface, prevalence of red blood cell aggregation in the central zone of the bulbar conjunctiva, smoking experience, pack/year index, mean venule diameters (in both eyes)), the features that make the maximum contribution to the shift of labels along this axis are highlighted in bold; the horizontal axis (Root1) reflects the totality of the same 8 features characterizing the MCB and smoking history (mean arteriolo/venular coefficient values, no smoke (years), the number of cigarettes per day, the number of capillaries of the left eye of the conjunctiva surface, the prevalence of red blood cell aggregation in the central zone of the bulbar conjunctiva, smoking experience, pack/year index, average venule diameters (in both eyes)), the features that make the maximum contribution to the shift of labels along this axis are highlighted in bold. G_1:1 – a group of people who stopped smoking; G_2:2 – group of smokers; G_3:3 – a group of non-smokers

зующие курение. В результате получена новая дискриминантная функция, содержащая 5 признаков: «среднее значение артериоло-венулярного коэффициента», «число капилляров на конъюнктиве левого глаза», «распространенность агрегации в переходной зоне бульбарной конъюнктивы», «средние диаметры венул (по обоим глазам)» и «степень внутрисосудистой агрегации эритроцитов (ВСАЭ) в артериолах бульбарной конъюнктивы». Приведенная в табл. 3 классификационная матрица демонстрирует высокий процент правильной переклассификации, который составил 81,25 %.

Таким образом, используя 5 количественных признаков, включенных в данную дискриминантную функцию, можно со 100 %-й уверенностью предсказать отнесение обследованного индивидуума

Таблица 3

Классификационная матрица

Table 3

Classification matrix

Группа наблюдения	Строки: фактически наблюдаемые классификации. Столбцы: предсказанные классификации			
	процент правильного предсказания	1-я группа (прекратили курить)	2-я группа (курят)	3-я группа (не курят)
1-я (прекратили курить)	77,1429	27	0	8
2-я (курят)	100,0000	0	16	0
3-я (не курят)	75,8621	6	1	22
Общий процент	81,25	33	17	30

Таблица 4

Показатели стандартизованных коэффициентов признаков, включенных в дискриминантную функцию

Table 4

Standardized feature coefficients included in the discriminant function

Включенные признаки	Стандартизованные коэффициенты дискриминантной функции	
	ось x	ось y
Среднее значение артериоло-венулярного коэффициента	0,971249	-0,526266
Число капилляров на конъюнктиве левого глаза	0,447450	0,237240
Распространенность агрегации в переходной зоне бульбарной конъюнктивы	-0,265086	-0,562107
Средние диаметры венул (по обоим глазам)	0,495657	0,406833
Степень внутрисосудистой агрегации эритроцитов в артериолах бульбарной конъюнктивы	-0,338713	-0,088024

к группе курящих, с 77 %-й вероятностью – к группе прекративших курить и с почти 76 %-й вероятностью – к группе некурящих. Стандартизованные коэффициенты дискриминантных функций по двум осям ординат, приведены в табл. 4.

Приведенные выше стандартизованные коэффициенты позволяют понять, что максимальный вклад в смещение показателей по горизонтальной оси X дает признак «среднее значение артериоло-венулярного коэффициента», затем примерно равнозначный вклад вносят признаки «средние диаметры венул (по обоим глазам)» и «число капилляров на конъюнктиве левого глаза». По вертикальной оси Y максимальный вклад вносят показатели «распространенность агрегации в переходной зоне бульбарной конъюнктивы» и «среднее значение артериоло-венулярного коэффициента». График, демонстрирующий взаимное расположение трех анализируемых групп, приведен на рис. 2.

На рис. 2 отчетливо видна локализация группы 2 – курит, она смещена по горизонтальной оси X левее группы 1 – прекративших курить и группы 3 – некурящих. Тогда как разделение (со взаимным перекрытием) групп 1 – прекративших курить и 3 – некурящих, наблюдается по вертикальной оси Y. На полученное пространственное распределение анализируемых групп по оси X, в первую очередь, влияет значение артериоло-венулярного коэффициента. Следующие показатели, максимально смеща-

ющие группу курящих по оси X влево, – это «средние диаметры венул (по обоим глазам)» и «число капилляров на конъюнктиве левого глаза», которые принимают наименьшие значения. По этим признакам группа прекративших курить отличается и отделяется по оси X от курящих и приближается к группе некурящих. По оси Y наблюдается сложное взаимное расположение всех групп, так как в смещение по этой оси весомые вклады вносят не только признак «распространенность агрегации в переходной зоне бульбарной конъюнктивы», но и показатели артериоло-венулярного коэффициента и средних диаметров венул на конъюнктиве обоих глаз.

Из 23 количественных признаков, отражающих состояние микрососудов конъюнктивы, микрогемодинамику и микрогемореологию, статистически значимыми оказались только 5 признаков.

Таким образом, у молодых людей, прекративших курение табака, в сравнении с курящими, в состоянии МЦР и МЦ появляются позитивные сдвиги. После отказа от курения происходит увеличение средних значений артериоло-венулярного коэффициента, диаметров венул на конъюнктиве обоих глаз и числа капилляров на конъюнктиве левого глаза, уменьшается распространенность агрегации в переходной зоне бульбарной конъюнктивы и степень ВСАЭ в артериолах. Подобные результаты, полученные ранее в наших работах с помощью более простых статистических методик изучения состояния

МЦР и МЦ у прекративших курить, с применением дискриминантного анализа получили полное подтверждение.

В последующем дискриминантный анализ применялся нами при изучении динамики МЦР-параметров в зависимости от сроков отказа от курения в группах молодых лиц, пациентов с АГ и ИБС.

Второй метод, на котором мы хотели подробно остановиться в данной статье, – логистическая регрессия (логит-регрессия). Этот метод позволяет проводить статистический анализ, используя все, как количественные, так и качественные признаки, и устанавливать взаимосвязь этих признаков с одним группирующим (качественным) признаком [14].

Результаты использования этого метода в анализе данных позволяют выйти на ограниченный спектр количественных и качественных характеристик изучаемого процесса, что делает возможным упрощать процедуры, используемые для первоначальной фиксации МЦ-параметров в эксперименте, закрепляя это патентом на изобретение [15, 16].

Приводим пример использования нами метода логистической регрессии в исследовании МЦ при системной склеродермии (ССД). Реабилитируя метод БМБК в исследовании МЦР при ССД, в анализ изображений МЦР были введены новые, ранее не используемые признаки (число капилляров в четырех участках бульбарной конъюнктивы, наличие аневризм артериол и ряд других), поэтому важно было оценить их диагностическую значимость.

В качестве группирующего признака взяли «группу наблюдения», имеющего две градации – «больные с ССД» и «контроль» – и анализировали взаимосвязи с подмножеством количественных и качественных признаков МЦ [17, 18] с пошаговым алгоритмом включения и исключения предикторов. Результаты оценки уравнений логистической регрессии представлены набором коэффициентов регрессии, достигнутыми уровнями значимости для каждого коэффициента, а также оценкой показателя согласия (Concordant) фактической принадлежности пациента к той или иной группе и теоретической принадлежностью, полученной по уравнению логит-регрессии. Для наглядности полученных уравнений логит-регрессии использовали графики с ROC-кривыми [14, 19–21] и таблицами

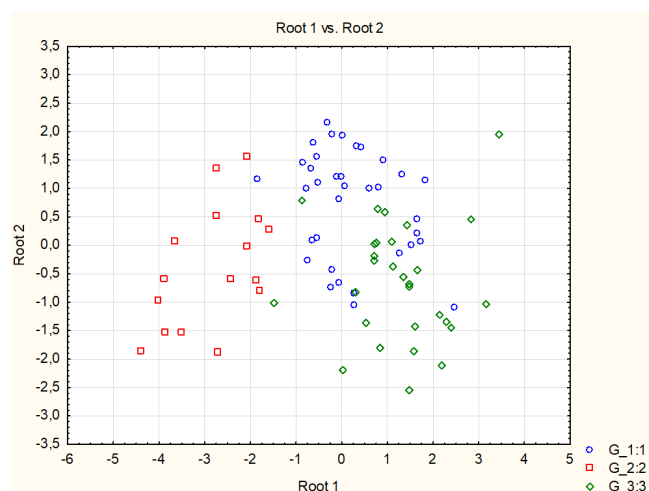


Рис. 2. График расположения трех анализируемых групп в осях ординат на основании дискриминантной функции: вертикальная ось (Root2) – отражает совокупность 5 признаков, характеризующих МЦР и МЦ (среднее значение артериоло-венулярного коэффициента, число капилляров на конъюнктиве левого глаза, распространенность агрегации в переходной зоне бульбарной конъюнктивы, средние диаметры венул (по обоим глазам), степень внутрисосудистой агрегации эритроцитов в артериолах бульбарной конъюнктивы), жирным выделены признаки, вносящие максимальный вклад в смещение меток по этой оси; горизонтальная ось (Root1) – отражает совокупность тех же 5 признаков, характеризующих МЦР и МЦ (среднее значение артериоло-венулярного коэффициента, число капилляров на конъюнктиве левого глаза, распространенность агрегации в переходной зоне бульбарной конъюнктивы, средние диаметры венул (по обоим глазам), степень внутрисосудистой агрегации эритроцитов в артериолах бульбарной конъюнктивы), жирным выделены признаки, вносящие максимальный вклад в смещение меток по этой оси. G_1:1 – группа прекративших курить; G_2:2 – группа курящих; G_3:3 – группа некурящих

Fig. 2. Chart of the three analyzed groups in the ordinate axes, based on the discriminant function: the vertical axis (Root2) – reflects a set of 5 features that characterize the MCB and MC (mean value of the arteriolo/venular ratio, the number of capillaries of the left eye of the conjunctiva surface, the prevalence of red blood cell aggregation in the transitional zone of the bulbar conjunctiva, the average diameters of venules (in both eyes), the intravascular red blood cell aggregation degree in the arterioles of the bulbar conjunctiva), bold marks indicate the features that make the maximum contribution to the shift of labels along this axis; the horizontal axis (Root1) reflects the totality of the same 5 features that characterize the MCB and MC (the average value of the arteriolo/venular ratio, the number of capillaries of the left eye of the conjunctiva surface, the prevalence of red blood cell aggregation in the transitional zone of the bulbar conjunctiva, the average diameters of venules (in both eyes), the degree intravascular red blood cell aggregation in the arterioles of the bulbar conjunctiva), the features that make the maximum contribution to the shift of marks along this axis are highlighted in bold. G_1:1 – a group of people who stopped smoking; G_2:2 – group of smokers; G_3:3 – a group of non-smokers

Таблица 5

Уравнение логит-регрессии № 1

Table 5

Logit-regression equation № 1

Параметр	Коэффициенты для включения в уравнение	Стандартная ошибка (среднего)	Статистика Вальда χ^2	Достигнутые уровни значимости p	Стандартизованные коэффициенты
Свободный член	-7,4667	2,1235	12,364	0,0004	
Средняя плотность капилляров в зоне перилимба	2,6673	0,7341	13,2024	0,0003	1,2646
ВСАЭ в венулах	-1,5313	0,692	4,8964	0,0269	-0,6695

Примечание: association of Predicted Probabilities and Observed Responses Concordant = 95 %, Somers'D 0,901.

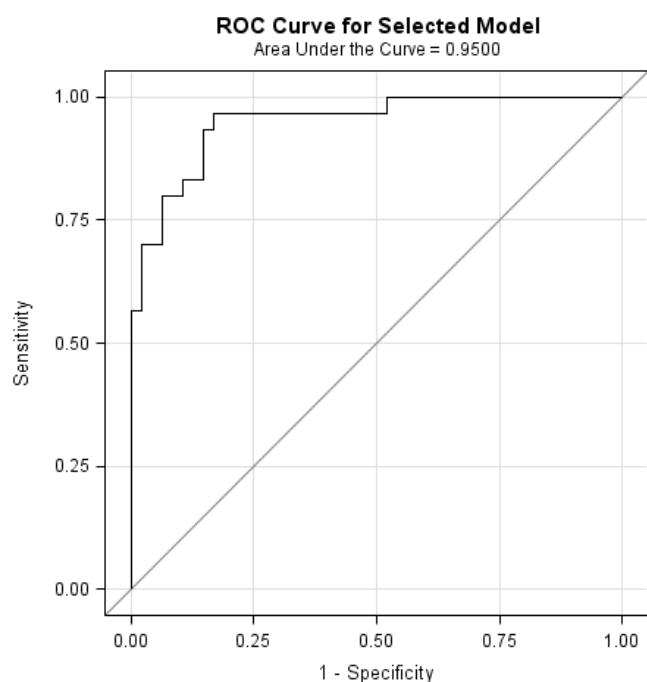


Рис. 3. ROC-кривая, демонстрирующая уравнение логистической регрессии с двумя количественными переменными, приведенными в табл. 5

Fig. 3. ROC-curve showing a logistic regression equation with two scale variables shown in Table 5

Frequency Expected Cell Chi-Square Percent Row Pct Col Pct	контроль	больные ССД	Total
контроль	25 10.472 20.154 34.72 86.21 96.15	4 18.528 11.391 5.56 13.79 8.70	29 40.28
больные ССД	1 15.528 13.592 1.39 2.33 3.85	42 27.472 7.6825 58.33 97.67 91.30	43 59.72
Total	26 36.11	46 63.89	72 100.00

Frequency Missing = 6

Statistics for Table of FACT_F by PRED_P			
Statistic	DF	Value	Prob
Chi-Square	1	52.8200	<.0001
Likelihood Ratio Chi-Square	1	61.4159	<.0001
Continuity Adj. Chi-Square	1	49.2468	<.0001
Mantel-Haenszel Chi-Square	1	52.0864	<.0001
Phi Coefficient		0.8565	
Contingency Coefficient		0.6505	
Cramer's V		0.8565	

Рис. 4. Таблица сопряженности, демонстрирующая перегруппировку случаев наблюдения на основании МЦ-признаков, отобранных первым уравнением логит-регрессии

Fig. 4. Contingency table showing the regrouping of cases based on the MC features selected by the first logit regression equation

Таблица 6

Количественные значения признаков, входящих в уравнение логистической регрессии № 1

Table 6

Quantitative values of features included in the logistic regression equation № 1

Признак	Контроль, n=30 (M±σ)	Пациенты с ССД, n=48 (M±σ)	Уровень статистической значимости p (значения V-коэффициента Крамера)
Плотность капилляров в зоне перилимба, ед./мм ²	2,9±0,2	1,8±0,1	p<0,0001
Средняя степень ВСАЭ в венулах	1,6±0,2	2±0,1	p=0,03

Таблица 7

Уравнение логит-регрессии № 2

Table 7

Logit-regression equation № 2

Параметр	Коэффициенты для включения в уравнение	Стандартная ошибка (среднего)	Статистика Вальда χ ²	Достигнутые уровни значимости p	Стандартизованные коэффициенты
Свободный член	-8,384	2,623	10,2166	0,0014	
Средняя плотность капилляров на 1 мм ² поверхности конъюнктивы	2,6698	0,8604	9,6284	0,0019	1,1053
ВСАЭ в капиллярах	-1,3181	0,6295	4,3839	0,0363	-0,6995

Примечание: Association of Predicted Probabilities and Observed Responses Concordant = 93,7 %, Somers'D 0,873.

распределения [22–24]. В процессе анализа были получены 5 эффективных моделей логит-регрессии. Нами было отобрано два уравнения и ROC-кривые к ним, которые продемонстрировали максимальные степени правильной перегруппировки пациентов с ССД, используя только отобранные уравнением

логит-регрессии МЦ-признаки. В табл. 5 приводим первое уравнение логит-регрессии.

Данные табл. 5 показывают, что из двух предикторов, участвующих в уравнении логит-регрессии, у признака, характеризующего среднюю плотность капилляров в зоне перилимба, наибольший стандар-

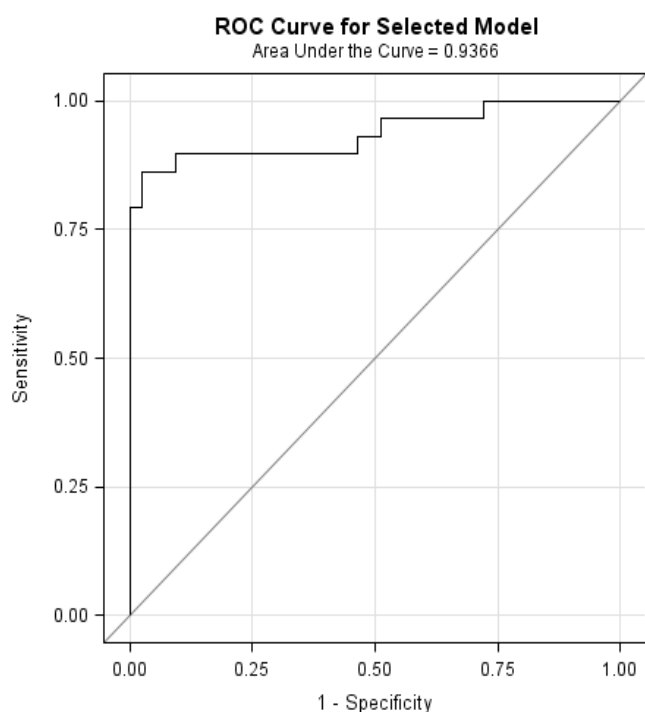


Рис. 5. ROC-кривая, демонстрирующая уравнение логистической регрессии с двумя количественными переменными, приведенными в табл. 7

Fig. 5. ROC-curve showing a logistic regression equation with two scale variables shown in Table 7

Frequency Expected Cell Chi-Square Percent Row Pct Col Pct	контроль	больные ССД	Total
контроль	25	4	29
	11.278	17.722	
	16.696	10.625	40.28
	34.72	5.56	
	86.21	13.79	
	89.29	9.09	
больные ССД	3	40	43
	16.722	26.278	
	11.26	7.1657	59.72
	4.17	55.56	
	6.98	93.02	
	10.71	90.91	
Total	28	44	72
	38.89	61.11	100.00

Frequency Missing = 6

Statistics for Table of FACT_P by PRED_P

Statistic	DF	Value	Prob
Chi-Square	1	45.7477	<.0001
Likelihood Ratio Chi-Square	1	51.1976	<.0001
Continuity Adj. Chi-Square	1	42.4746	<.0001
Mantel-Haenszel Chi-Square	1	45.1123	<.0001
Phi Coefficient		0.7971	
Contingency Coefficient		0.6233	
Cramer's V		0.7971	

Рис. 6. Таблица сопряженности, демонстрирующая перегруппировку случаев наблюдения на основании МЦ-признаков, отобранных вторым уравнением логит-регрессии

Fig. 6. Contingency table showing the regrouping of cases based on MC features selected by the second logit regression equation

Таблица 8

Количественные значения признаков, входящих в уравнение логистической регрессии № 2

Table 8

Quantitative values of features included in the logistic regression equation № 2

Признак	Контроль, n=30 (M±σ)	Пациенты с ССД, n=48 (M±σ)	Уровень статистической значимости p (значения V-коэффициента Крамера)
Средняя плотность капилляров по всем зонам, ед./мм ²	3,1±0,1	2,3±0,1	p<0,0001
Средняя степень ВСАЭ в капиллярах	1±0,1	1,9±0,1	p<0,0001

тизированный коэффициент (1,2646). На рис. 3 показана ROC-кривая, большая площадь под которой указывает на достоверность полученного уравнения.

Средние значения признаков, входящих в первое уравнение логистической регрессии, отдельно для группы пациентов с ССД и контрольной группы приведены в табл. 6. На значения этих параметров следует опираться в интерпретации полученных при БМБК данных.

На рис. 4 приведена таблица сопряженности, которая показывает проценты правильной переклассификации анализируемых групп (больные ССД и контроль), используя только два МЦ-предиктора.

Так, на основании данных о средней плотности капилляров в зоне перилимба и степени ВСАЭ в венулах с 97,67 %-й вероятностью можно отнести пациента к группе больных ССД и с 86,21 %-й – к группе контроля (p<0,0001).

Следующее уравнение логистической регрессии, содержащее два МЦ-признака, показано в табл. 7.

В приведенном в табл. 7 уравнении логит-регрессии присутствуют также два МЦ-предиктора, причем признак, характеризующий среднюю плотность капилляров на 1 мм² поверхности конъюнктивы, вносит в данное уравнение значительный вклад, о чем свидетельствует высокое значение стандартизованного коэффициента (1,1053) и высокий процент конкордантности – 93,7 %. Значимость предложенной модели логистической регрессии с включением в уравнение только двух количественных признаков подтверждает большая площадь под ROC-кривой, показанной на рис. 5.

Средние значения признаков, входящих во второе уравнение логистической регрессии, на которые следует ориентироваться в интерпретации полученных при БМБК данных, отдельно для группы пациентов с ССД и контрольной группы приведены в табл. 8.

На рис. 6 приведена таблица сопряженности, которая демонстрирует проценты правильной переклассификации анализируемых групп только на

основании отобранных уравнением логит-регрессии двух признаков.

Так, на основании данных о средней плотности капилляров на 1 мм² поверхности конъюнктивы и степени ВСАЭ в капиллярах с вероятностью 93,02 % можно отнести пациента к группе больных ССД и с вероятностью 86,21 % – к группе контроля, что подтверждается высоким уровнем статистической значимости ($p < 0,0001$).

Основываясь на двух представленных уравнениях логистической регрессии из большого перечня анализируемых БМБК-признаков микроциркуляции, для упрощения алгоритма анализа МЦ-картины бульбарной конъюнктивы у больных с ССД было предложено использовать лишь признаки, характеризующие плотность капилляров в различных областях бульбарной конъюнктивы и среднее значение числа капилляров на 1 мм² поверхности конъюнктивы, а также количественные характеристики, отражающие степень ВСАЭ в венах и капиллярах как наиболее чувствительные и показательные.

Заключение

Использование методов биомедицинской статистики является важным и неотъемлемым этапом работы ученого-медика (биолога). В то же время это сложная быстро развивающаяся наука, требующая специальных глубоких знаний, позволяющих корректно применять те или иные статистические методы в анализе данных. Сотрудничество ученого-медика (биолога) и специалиста по биомедицинской статистике крайне актуально в современной науке. Оно позволит избежать методологических ошибок анализа собственных данных, получать правдивые результаты, извлекать максимальное количество полезной, новой информации из первичного материала и делать открытия там, где, не используя в полной мере возможности статистического анализа, можно было бы «пройти мимо».

Конфликт интересов / Conflict of interest

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

Литература / References

1. Корнеева Н. В., Леонов В. П., Жмеренецкий К. В. Конъюнктивальная биомикроскопия: методология анализа: монография. – Хабаровск: Изд-во ДВГМУ, 2020. [Korneeva NV, Leonov VP, Zhmerenetsky KV. Conjunctival biomicroscopy: methodology of analysis. Khabarovsk, FESMU publishing, 2020. (In Russ.)].
2. Кобзарь А. И. Прикладная математическая статистика: пособие для инженеров и научных работников. – М.: Физматлит, 2006. [Kobzar' AI. Prikladnaja matematicheskaja statistika: posobie dlja inzhenerov i nauchnyh rabotnikov. Moscow, Fizmatlit, 2006. (In Russ.)].
3. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. Second Edition. 2021. Doi: 10.1007/978-1-4614-7138-7.
4. Ланг Т., Сесик М. Как описывать статистику в медицине: руководство для авторов, редакторов и рецензентов / пер. с англ. под ред. В. П. Леонова. – М.: Практическая медицина, 2016. [Lang TA, Sesik M. How to

describe statistics in medicine. The management for authors, editors and reviewers the translation from English under Leonov V. P. version. Moscow, Prakticheskaja Medicina, 2011. (In Russ.)]. Doi: 10.1097/00006205-199705000-00022.

5. Wasserman L. All of Statistics: A Concise Course in Statistical Inference. – Springer Texts in Statistics, 2010.

6. Прохоров Ю. В. Вероятность и математическая статистика // Энциклопедия. М.: Большая российская энциклопедия, 1999. [Prohorov JuV. Veroyatnost' i matematicheskaja statistika. Jenciklopedija. Moscow, Bol'shaja Rossijskaja Jenciklopedija, 1999. (In Russ.)].

7. Бард Й. Нелинейное оценивание параметров. – М.: Статистика, 1979. [Bard J. Nelinejnoe ocenivanie parametrov. Moscow, Statistika, 1979. (In Russ.)].

8. Лапко А. В. Непараметрические методы классификации и их применение. – Новосибирск: Наука, 1993. [Lapko AV. Neparametricheskie metody klassifikacii i ih primenenie. Novosibirsk, Nauka, 1993. (In Russ.)].

9. Леонов В. П. Логистическая регрессия в медицине и биологии. URL: http://www.biometrika.tomsk.ru/logit_9.htm (дата обращения: 28.06.2022). [Leonov VP. Logisticheskaja regressija v medicine i biologii. Available at: http://www.biometrika.tomsk.ru/logit_9.htm (accessed: 28.06.2022)].

10. Корнеева Н. В. Оценка состояния микроциркуляторного русла и микроциркуляции у практически здоровых людей молодого возраста, прекративших курение табака // Дальневосточ. мед. журн. – 2016. – № 3. – С. 88–91. [Korneeva NV. A study of the microvasculature and microcirculation in otherwise young subjects former smokers // Dal'nevostochnyj medicinskij zhurnal. 2016;(3):88–91. (In Russ.)].

11. Прикладной многомерный статистический анализ / под ред. С. А. Айвазян, А. И. Орлов. – М.: Наука, 1978. [Ajvazjan SA, Orlov AI. Prikladnoj mnogomernyj statisticheskij analiz. Moscow, Nauka, 1978. (In Russ.)].

12. Барковский С. С., Захаров В. М., Лукашов А. М. и др. Многомерный анализ данных методами прикладной статистики: учебное пособие. Казань: КГТУ, 2010. [Barkovskij SS, Zaharov VM, Lukashov AM, Nurutdinova AR, Shalagin SV. Mnogomernyj analiz dannyh metodami prikladnoj statistiki: uchebnoe posobie. Kazan', KGTU, 2010. (In Russ.)].

13. Факторный, дискриминантный и кластерный анализ / под ред. И. С. Енюкова. – М.: Финансы и статистика, 1989. [Faktornyj, diskriminantnyj i klasternyj analiz / eds by IS Enjukova. Moscow, Finansy i statistika, 1989. (In Russ.)].

14. Roderick JA, Little, Donald B. Rubin. Statistical Analysis with Missing Data. 3rd ed. 2019.

15. Патент РФ на изобретение № 2725281 C1. Способ неинвазивной оценки наличия атеросклеротического поражения коронарного русла у человека in vivo / Корнеева Н. В., Сиротин Б. З. Заяв. № 2019112008 от 19.04.2019; опубл. 30.06.2020. [Patent RF na izobretenie № 2725281 C1. Method for non-invasive assessment of presence of atherosclerotic lesion of human coronary bed in vivo / Korneeva NV, Sirotnin BZ. Zayavka № 2019112008 ot 19.04.2019. 30.06.2020].

16. Корнеева Н. В. Упрощенный алгоритм видео-биомикроскопии бульбарной конъюнктивы у молодых людей, прекративших курение, полученный с помощью метода логистической регрессии // Регионарное кровообращение и микроциркуляция. – 2016. – Т. 15, № 4 (60). – С. 75–81. [Korneeva NV. The simplified algorithm of the video biomicroscopy of a bulbar conjunctiva in the young quitted smoking subjects by means of the logistic regression method. Regional blood circulation and microcirculation. 2016; 15(4(60)):75–81. (In Russ.)].

17. Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. М.: Финансы и статистика, 1987. [Vuchkov I, Bojadzhieva L, Solakov E. Prikladnoj

lineynyy regressionnyy analiz. Moscow, Finansy i statistika, 1987. (In Russ.).]

18. Гайдышев И. Анализ и обработка данных. Специальный справочник. СПб.: Питер, 2001. [Gajdyshv I. Analiz i obrabotka dannyh. Special'nyy spravochnik. SPb., Piter, 2001. (In Russ.).]

19. Леонов В. П. Логистическая регрессия и ROC-анализ. URL: http://www.biometrika.tomsk.ru/logit_4.htm (дата обращения: 28.06.2022). [Leonov VP. Logisticheskaja regressija i ROC-analiz. Available at: http://www.biometrika.tomsk.ru/logit_4.htm. (accessed: 28.06.2022)].

20. Kottmann JE. *Statistics & Statistical Analysis Illustrated: Variance Everywhere Paperback*. 2022.

21. Корнеева Н. В., Оттева Э. Н., Жмеренецкий К. В. и др. Конъюнктивная биомикроскопия в диагностике микроциркуляторных нарушений при системной склеродермии: новый взгляд на «старый» метод // Науч.-практ. ревматология. – 2021. – Т. 59, № 4. – С. 426–433. [Korneeva NV, Otteva EN, Zhmerenetsky KV, Kovaleva MA, Panevin TS. Conjunctival biomicroscopy in the diagnosis of microcirculatory disorders in systemic sclerosis: A new look at the «old» method // Nauchno-Prakticheskaya Revmatologia = Rheumatology Science and Practice. 2021;59(4):426–433. (In Russ.).]. Doi: 10.47360/1995-4484-2021-426-433.

22. Точный критерий Фишера для таблиц сопряженности признаков 2x2. URL: <http://www.biometrika.tomsk.ru/exact.htm> (дата обращения: 28.06.2022). [Tochnyy kriterij Fishera dlja tablic sopryazhonnosti priznakov 2x2. Available at: <http://www.biometrika.tomsk.ru/exact.htm> (accessed: 28.06.2022)].

23. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. 2-е изд. М.: Физматлит, 2012. [Kobzar' AI. Prikladnaja matematika.

statistika. Dlja inzhenerov i nauchnyh rabotnikov. 2-e izdanie. Moscow, Fizmatlit, 2012. (In Russ.).]

24. Statistics By Jim Publishing. 2020. Available et: <https://statisticsbyjim.com/> (accessed: 28.06.2022).

Информация об авторах

Корнеева Наталья Вячеславовна – д-р мед. наук, доцент, профессор кафедры факультетской и поликлинической терапии с курсом эндокринологии, Дальневосточный государственный медицинский университет, г. Хабаровск, Россия, e-mail: Gladkova1982@mail.ru, ORCID: 0000-0001-9878-180X.

Леонов Василий Петрович – канд. техн. наук, доцент, НЦ «БИОСТАТИСТИКА», редактор сайта БИОМЕТРИКА (<http://www.biometrika.tomsk.ru/>), e-mail: leo.biostat@mail.com, ORCID: 0000-0002-7968-7620.

Жмеренецкий Константин Вячеславович – д-р мед. наук, доцент, член-корреспондент РАН, ректор, Дальневосточный государственный медицинский университет, г. Хабаровск, Россия, e-mail: rec@mail.fesmu.ru, ORCID: 0000-0002-6790-3146.

Authors information

Korneeva Natalia V. – PhD, doctor of medical science, professor of faculty and ambulance medicine department with a course of endocrinology FESMU, Khabarovsk, Russia, e-mail: Gladkova1982@mail.ru, ORCID: 0000-0001-9878-180X.

Leonov Vasily P. – associate professor, SC Biostatistics, editor of the site «Biometric» (<http://www.biometrika.tomsk.ru/>), e-mail: leo.biostat@mail.com, ORCID: 0000-0002-7968-7620.

Zhmerenetsky Konstantin V. – PhD, corresponding member of the Russian Academy of Sciences, rector, Far Eastern State Medical University, Khabarovsk, Russia, e-mail: rec@mail.fesmu.ru, ORCID: 0000-0002-6790-3146.