

УДК 616

Бенчак Г.Ю., Зайцева Е.А., Леготкина Л.А., Бразовская Н.Г., Денисова О.А.,
Калачева Т.П.

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДЛЯ ОПРЕДЕЛЕНИЯ ТИПА ТЕЧЕНИЯ САРКОИДОЗА ОРГАНОВ ДЫХАНИЯ

Сибирский государственный медицинский университет, г. Томск

Саркоидозу подвержены люди обоих полов, всех возрастов и рас. Врачей разных специальностей привлекает определение факторов, влияющих на тип течения саркоидоза. Целью исследования является построение оптимальной модели классификации типа течения саркоидоза, основанной на анамнестических и клинико-лабораторных факторах, а также данных инструментального обследования пациентов. В данной работе описано построение классификационных моделей типа течения саркоидоза. Модели были построены методами логистической регрессии и линейного дискриминантного анализа. Выявлены факторы, ассоциированные с типом течения саркоидоза.

Ключевые слова: саркоидоз, модель, логистическая регрессия, дискриминантный анализ

Саркоидоз является системным воспалительным заболеванием неизвестной этиологии, характеризующимся образованием неказеифицирующихся гранулем, мультисистемным поражением и активацией Т-клеток в месте гранулем. [3]

Саркоидозу подвержены люди обоих полов, всех возрастов и рас. Распространенность саркоидозом в мире варьируется от 1 до 40 случаев на 100 тыс. населения, а в России находится в пределах от 22 до 47 случаев на 100 тыс. [5]

В зависимости от типа течения саркоидоза выбирается стратегия лечения. Ежегодно растет количество случаев саркоидоза с неблагоприятным течением и формированием осложнений.

При благоприятном типе медикаментозное лечение не требуется, а при неблагоприятном течении необходимо назначение терапии системными глюкокортикостероидами (СГКС) и цитостатиками. Несвоевременное назначение данной терапии приводит к серьезным осложнениям, вплоть до развития острой дыхательной недостаточности [1]. В связи с этим большое внимание фтизиатров и пульмонологов привлекает определение факторов, влияющих на тип течения саркоидоза. Степень активности данной патологии может быть различной и зависит как от выраженности общевоспалительных симптомов, так и от числа органов, вовлеченных в патологический процесс, и степени их структурно-функциональных нарушений.

Возникает необходимость определения методики выявления пациентов с неблагоприятным течением саркоидоза на основании клинико-лабораторных и инструментальных показателей.

Цель работы

Построить оптимальную модель классификации типа течения саркоидоза, основанную на анамнестических, клинико-лабораторных факторах и данных инструментального обследования пациентов.

Соответствие нормам этики

Исследование проводилось в соответствии с Хельсинской декларацией и правилами GCP, было одобрено решением локального комитета по этике ФГБОУ ВО СибГМУ Минздрава России от 28.11.2006, регистрационный номер 5045.

Всем госпитализированным в стационар пациентам были оформлены истории болезни. Пациентам, наблюдавшимся в амбулаторных условиях, оформлены амбулаторные карты. Участниками исследования подписано информированное согласие о том, что данные их истории (амбулаторной карты) могут использоваться в научных целях.

Материал и методы

Проведено ретроспективное исследование пациентов с морфологически верифицированным диагнозом «саркоидоз легких», у которых изучались основные клинические проявления при саркоидозе органов дыхания (СОД) в зависимости от особенностей течения заболевания. Обследовано 298 пациентов (женщин – 191, мужчин – 107) в возрасте от 19 до 74 лет (46(35;56)), которые были разделены на две клинические группы: 1-ая группа – 163 пациента с благоприятным типом течения саркоидоза; группа 2 – 135 больных с неблагоприятным типом течения саркоидоза. Благоприятный тип течения саркоидоза определялся при наличии самопроизвольного регресса заболевания, в т.ч. спонтанного, без применения системных глюкокортикостероидов (сГКС) или на фоне случайного непродолжительного назначения небольших доз сГКС, отсутствие рецидивов заболевания, похудания и генерализованных форм саркоидоза. Неблагоприятный тип течения саркоидоза – при наличии прогрессирования и рецидивов заболевания, генерализованных форм саркоидоза. [4] При описании характера течения заболевания использовалась терминология, в частности, понятия активной фазы (прогрессирования), фазы регрессии (спонтанной или под влиянием лечения) и фазы стабилизации (стационарной фазы), представленная в Федеральных клинических рекомендациях по диагностике и лечению саркоидоза.

Исследование включало более 200 параметров обследования пациентов. У всех больных оценивались жалобы и анамнестические данные, а также лабораторные показатели (общий и биохимический анализ крови, общий анализ мочи), параметры инструментальных методов исследования (спирометрии, рентгенографии органов грудной клетки, компьютерной томографии легких, ультразвукового исследования органов брюшной полости, почек, лимфатических узлов, ЭКГ, Эхо-КГ), наличие поражений внутренних органов.

Статистический анализ данных проводился с помощью пакета программы RStudio версии 4.3.1.

Статистические расчеты включали межгрупповые сравнения с помощью критерия Манна-Уитни, корреляционный анализ Спирмена, ROC-анализ. Статистически значимыми считались различия с p -уровнем меньше 0,05. Проводилось заполнение пропущенных данных с помощью метода k -ближайших соседей. Классификационные модели были построены с использованием методов линейного дискриминантного анализа и бинарной логистической регрессии. Оценка значимости моделей классификации типа течения саркоидоза проводилась с помощью метода ROC-анализа, расчёта чувствительности и специфичности моделей.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Результаты и обсуждение

На первом этапе было проведено сравнение групп пациентов с благоприятным и неблагоприятным течением саркоидоза. Выделены факторы, которые рассматривались для построения моделей. Переменной отклика, принимающей одно из двух значений – 0 (благоприятное течение) и 1 (неблагоприятное течение), в моделях является тип течения саркоидоза.

Была построена модель на основе показателей трансторакальной эхокардиографии (ЭхоКГ), рентгенологического обследования и данных компьютерной томографии высокого разрешения. Данные ЭхоКГ могут косвенно отражать поражение сердца при саркоидозе, которое сложно верифицировать в клинической практике. В данном исследовании изучается изолировано влияние эхокардиографических показателей на течение заболевания.

Для построения модели был выбран метод бинарной логистической регрессии, так как переменная отклика является категориальной дихотомической, а предикторы – как количественными, так и категориальными.

Перед моделированием в отобранных признаках были удалены значения с пропущенными данными. Полученная выборка составила 146 человек и была разделена на обучающую (116 человек) и тестовую (30 человек).

Решающее правило строилось по обучающей выборке.

В модель вошли: зоны гипокинезов по ЭхоКГ, толщина задней стенки левого желудочка, рентгенологическая стадия, наличие немногочисленных очагов поражения в легких.

Таблица 1

Предикторы типа течения саркоидоза

Признак	коэффициент	ОШ	ДИ		p-уровень
			2,5%	97,5%	
Свободный член	-3,9				<0,025
Толщина задней стенки левого желудочка	0,1	1,16	0,84	1,62	0,383
Немногочисленные очаги поражения в легких	-0,8	0,43	0,16	0,91	0,049
Зоны гипокинезов по ЭхоКГ	1,3	3,51	0,52	69,28	0,265
Рентгенологическая стадия	1,4	4,15	1,79	10,94	0,002

Уравнение логистической регрессии имеет вид:

$$p = \frac{1}{1 + e^{-(-3,886 + 1,255 \cdot \text{ЗГ} + 0,145 \cdot \text{ТЗСЛЖ} + 1,423 \cdot \text{РС} - 0,839 \cdot \text{НОПЛ})}}, \quad (1)$$

где ЗГ – зоны гипокинезов по ЭхоКГ (есть, нет); ТЗСЛЖ – толщина задней стенки левого желудочка (мм); РС – рентгенологическая стадия (нет изменений, 1, 2, 3 стадии); НОПЛ – наличие немногочисленных очагов поражения в легких (есть, нет).

В качестве точки отсечения взято стандартное значение 0,5. Решающее правило (1) работает так: если вероятность p , рассчитанная по формуле (1) больше 0,5, то тип течения классифицируется как «неблагоприятный», если меньше 0,5, то – «благоприятный».

Пример применения модели. Пациент с неблагоприятным типом течения саркоидоза, 3 рентгенологической стадией, без зон гипокинезов по ЭхоКГ, без немногочисленных очагов поражения в легких, с толщиной задней стенки левого желудочка 11 мм. Рассчитанное значение $p=0,99$, что больше 0,5, следовательно, тип течения классифицируется как «неблагоприятный». Предсказанное значение совпадает с действительным.

Качество модели оценивалось с помощью тестовой выборки. Чувствительность составила 80%, специфичность – 73%, процент правильных решений – 77%. Площадь под ROC-кривой равна 0,81.

Большинство параметров, оцениваемых у пациентов в клинической практике, относятся к категориальным признакам, которые определяются врачами субъективно. Поэтому более информативными можно считать модели, построенные с использованием количественных признаков. Методом статистического моделирования, в котором в качестве предикторов выступают количественные признаки, является линейный дискриминантный анализ.

С его помощью была построена модель, в которую вошли следующие предикторы: возраст на момент начала заболевания, уровень кальция в крови, индекс Тиффно, толщина

задней стенки левого желудочка, толщина межжелудочковой перегородки, объем правого желудочка, фракция выброса левого желудочка, уровень АСТ.

При работе с медицинскими данными есть проблема пропусков в базах данных [6]. Это может быть связано с изменениями в протоколах или с ошибками заполнения документации. Пропущенные данные могут повлиять на качество статистического моделирования. Есть несколько подходов к заполнению отсутствующих данных: заполнение средними или медианными значениями, линейной регрессией, методами с использованием машинного обучения.

Признаки, отобранные для построения модели, были проверены на пропуски данных. Часть значений, с большим количеством пропусков, была удалена. После удаления доля пропусков составила 23%. Для заполнения оставшихся пропусков использовался метод k-ближайших соседей. Значения гиперпараметра k равно 5 (значение k – количество ближайших точек, используемых для расчета целевого значения) [2]. После было проведено масштабирование данных, таким образом, что среднее значение каждой переменной равно 0, а стандартное отклонение – 1.

Полученная выборка составила 188 человек и была разделена на обучающую (152 человек) и тестовую (36 человек). Решающее правило строилось по обучающей выборке.

Полученная модель является статистически значимой (значение лямбды Уилкса равно 0,83; $F(8,143)=3,7$; $p<0,001$). Уравнение имеет вид:

$$y = 0,380 * Ca^{2+} + 0,096 * ПЖ - 0,941 * ИТ + 0,349 * T_{ЗСЛЖ} - 0,173 * T_{МЖП} - 0,098 * ФВ + 0,407 * АСТ + 0,063 * В, (2)$$

где Ca^{2+} – уровень кальция в крови (ммоль/л); ПЖ – объем правого желудочка (мл); ИТ – индекс Тиффно (%); $T_{ЗСЛЖ}$ – толщина задней стенки левого желудочка (мм); $T_{МЖП}$ – толщина межжелудочковой перегородки (мм); ФВ – фракция выброса левого желудочка (%), АСТ (Ед/л); В – возраст на момент начала заболевания (год).

Подставляя в дискриминантное уравнение средние значения предикторов в группе с благоприятным и неблагоприятным течением саркоидоза, рассчитали значения функции, которые являются центроидами (для группы с неблагоприятным течением равно значению равно -0,46, для группы с благоприятным течением равно 0,44). Точка разделения классов равна -0,01. Решающее правило (2) работает так: если значение y, рассчитанное по формуле (2) меньше -0,01, то тип течения классифицируется как «неблагоприятный», если больше -0,01, то – «благоприятный».

Пример применения модели. Пациент с благоприятным типом течения саркоидоза, в возрасте 50 лет (0,56), уровнем кальция в крови 2,54 ммоль/л (0,36), индексом Тиффно 100% (-3,21), толщиной задней стенки левого желудочка 10 мм (-0,40), толщиной межжелудочковой

перегородки 9 мм (-0,21), объемом правого желудочка 22 мл (1,06), фракцией выброса левого желудочка 69% (-0,20), уровнем АСТ 36 Ед/л (-0,71). Рассчитанное значение u равно 2,93, что больше -0,01, следовательно, тип течения классифицируется как «благоприятный». Предсказанное значение совпадает с действительным.

Качество модели оценивалось с помощью тестовой выборки. Чувствительность составила 86%, специфичность – 87%, процент правильных решений – 85%.

Заключение и выводы

В результате проведенного исследования были получены модели классификации типа течения саркоидоза. Они построены с применением разных методов статистического моделирования. Модели показывают, что параметры ЭхоКГ, данные рентгенологического обследования имеют связь с типом течения саркоидоза. Обе модели имеют достаточное качество классификации.

Исследование позволило выделить совокупность характеристик, ассоциированных с тяжестью течения саркоидоза, что в будущем позволит запланировать исследования по построению индивидуальных прогнозов для больных саркоидозом органов дыхания.

ЛИТЕРАТУРА

1. Визель И.Ю. Саркоидоз: возможность спонтанной ремиссии // Вестник современной клинической медицины. – 2012. – №2.
2. Литтл, Р. Дж. А. Статистический анализ данных с пропусками : пер.с англ. / Р. Дж. А. Литтл, Д. Б. Рубин. – М. : Финансы и статистика, 1990. – 336 с.
3. Межрегиональная общественная организация «Российское Респираторное Общество». Общероссийская общественная организация «Российское научное медицинское общество терапевтов». Общероссийская общественная организация «Педиатрическое респираторное общество». Клинические рекомендации. *Саркоидоз, 2022*. Ссылка на рекомендации: https://cr.minzdrav.gov.ru/recomend/736_1
4. Постникова ЛБ, Гудим АЛ, Болдина МВ, Кубышева НИ. Клинические фенотипы больных саркоидозом в реальной клинической практике // Сибирское медицинское обозрение. – 2022. – №3. – С. 91-99. DOI: 10.20333/25000136-2022-3-91-99
5. Чучалин А. Г., Визель А. А., Илькович М. М. и др. Диагностика и лечение саркоидоза: резюме Федеральных согласительных клинических рекомендаций. Часть II. Классификация, этиопатогенез, клиника // Вестник современной клинической медицины. – 2014. – Т. 7, № 5. – С. 73-81. doi. org/10.20969/vskm
6. Burton A., Altman D. G. Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines. *British Journal of Cancer*, 2004, 91(1):4–8.