

11. Bole-Feysot Ch., Goffin V., Edery M. et al. // *Endocrine Rev.* — 1998. — Vol. 19, N 3. — P. 225–268.
12. Cavaco B., Leite Y., Santos M. A. et al. // *J. Endocrinol. Invest.* — 1993. — Vol. 16, Suppl. — P. 1–83.
13. Ciccarelli E., Camanni F. // *Drugs.* — 1996. — Vol. 51, N 6. — P. 954–965.
14. Colao A., Di Sarno A., Sarnacchiaro F. et al. // *J. Clin. Endocrinol.* — 1997. — Vol. 82. — P. 876–882.
15. Elshalbr H., Lew A., Paul W., Sundmarp V. // *J. Biol. Chem.* — 1991. — Vol. 266, N 304. — P. 22919–22925.
16. Haro L. S., Lee D. W., Singh R. N. // *J. Clin. Endocrinol.* — 1990. — Vol. 71, N 2. — P. 379–383.
17. Lamberts W. J., McLeod R. M. // *Physiol. Rev.* — 1990. — Vol. 70, N 2. — P. 279–308.
18. Pepperell R., Crosignani P. G., Franks S. *Progress in Prolactin Lowering Therapy.* — 1991.

Поступила 20.04.09

◆ ЛЕКЦИЯ

© КОЛЛЕКТИВ АВТОРОВ, 2009

УДК 616-036.8:31

П. О. Румянцев, В. А. Саенко, У. В. Румянцева, С. Ю. Чекин

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА В КЛИНИЧЕСКОЙ ПРАКТИКЕ

Часть II. Анализ выживаемости и многомерная статистика

ГУ Медицинский радиологический научный центр (дир. — акад. РАМН А. Ф. Цыб) РАМН, Обнинск

Статистический анализ является интегральной частью клинического исследования. Цель настоящей работы — помочь клиницистам разобраться в сути и принципах применения различных методов обработки медицинских данных, не углубляясь в детали математических расчетов. В обзоре рассмотрены наиболее востребованные и популярные статистические методы анализа данных, применяемые в клинической и экспериментальной медицине. В I части обзора были рассмотрены основы одномерной статистики, II часть посвящена принципам анализа выживаемости и многомерной статистики.

Ключевые слова: *методы статистического анализа, медицина, логистическая регрессия, множественная логистическая регрессия, отношение рисков, анализ выживаемости, таблицы дожития, метод Каплана—Мейера, лог-ранк, модель Кокса, моделирование, многомерная статистика.*

P.O. Rumyantsev, V.A. Saenko, U.V. Rumyantseva, S.Yu. Chekin

STATISTICAL METHODS FOR THE ANALYSES IN CLINICAL PRACTICE. PART 2. SURVIVAL ANALYSIS AND MULTIVARIATE STATISTICS

Medical Radiological Research Centre, Russian Academy of Medical Sciences, Obninsk

Statistical analysis is an integral component of clinical studies. The objective of the present paper is to assist clinicians in getting deeper insight into basic principles underlying different methods available for the statistical treatment of medical data without a detailed description of relevant mathematical calculations. The most popular and widely used methods of statistical analysis are considered with special reference to their practical application in clinical and experimental medicine. Part I of the review was devoted to foundations of descriptive statistics and univariate analysis. Part II is focused on the principles of survival analysis and multivariate methods.

Key words: *methods of statistical analysis, medicine, descriptive statistics, logistic regression, multiple logistic regression, hazard ratio, survival analysis, mortality tables, Kaplan--Meir method, log-rank, Cox model, simulation, multivariate statistics*

1. Методы анализа выживаемости

Под методами оценки выживаемости (survival) понимается изучение закономерности появления ожидаемого события у представителей наблюдаемой выборки во времени. Такое событие — не обязательно летальный исход, как можно предположить из названия анализа. Им может быть рецидив заболевания или, наоборот, выздоровление, в общем слу-

чае — происхождение определенного события. Точкой отсчета могут быть дата (час) выполнения процедуры, назначения лекарственного препарата, возраст на момент диагноза и т. п. Период времени от начального события (например, постановки диагноза) до итогового (летальный исход, рецидив, выздоровление) называется временем до события (time to event), или временем ожидания.

Исходно термин "выживаемость" заимствован из лексикона страховых компаний, использующих его в статистических расчетах при страховании жизни своих клиентов. С помощью определенной методики компания оценивает потенциальный риск летального исхода (страховой случай) или среднее время выживания (выживаемость, время до события) клиента с учетом сопутствующих рисков, что и определяет размер индивидуальных страховых взносов.

Сведения об авторах

Для контактов:

Румянцев Павел Олегович, канд. мед. наук, вед. науч. сотр.

Адрес: 249036, Обнинск, ул. Королева, 4

Телефон: (48439) 93241

roum@mrrc.obninsk.ru

Саенко Владимир Александрович, канд. биол. наук, ст. науч. сотр.

Румянцева Ульяна Викторовна, канд. мед. наук, ст. науч. сотр.

Чекин Сергей Юрьевич, ст. науч. сотр.

Исходя из общей постановки задачи, т. е. анализа среднего времени выживания и проведения на его основе, например, оценки эффективности нового метода лечения, казалось бы, можно воспользоваться параметрическими и непараметрическими статистическими методами, описанными в разделе 6 части I обзора. В принципе это возможно, но анализ выживаемости имеет важное отличие в способе построения выборки. В то время как для рассмотренных ранее статистических методов объем и структура выборки являются постоянными, в анализе времени до события они могут меняться. Проблема заключается в том, что время до события не обязательно может быть определено для всех пациентов выборки в ходе запланированного срока наблюдения. Значение этого показателя становится определенным только среди тех лиц, у которых произошло интересующее событие. Для всех остальных объектов наблюдения показатель остается неизвестным до наступления события, которое может вообще не произойти за период наблюдения. Кроме того, пациенты могут выбывать из исследования в силу разных обстоятельств (смена места жительства и т. п.), включаться в исследование в его середине или в конце, а также ожидаемое событие может быть вызвано иной причиной (например, летальный исход не от заболевания, а в результате несчастного случая). Все это приводит к (нерегулярным) качественным и количественным изменениям в анализируемых данных и определяет необходимость применения специальных методов, в которых можно было бы учесть и использовать неполную информацию.

Данные, которые содержат неполную информацию, называют цензурированными (censored). С такими выборками приходится иметь дело, когда наблюдаемый параметр является временем до наступления события, а период наблюдения ограничен (например, у пациента рецидив заболевания не обнаружен за 6 мес до того, как он переехал в другой город и дальнейшая информация о нем недоступна). При анализе выживаемости, как и при других методах статистического анализа, вся информация о выборке содержится в соответствующей ей функции распределения вероятности (в данном случае — времен ожидания), но используется она не в виде плотности распределения вероятности значений, а в виде функции выживания (survival function). Кумулятивная функция распределения $F(t)$ времен ожидания отражает вероятность того, что время ожидания события меньше t . Соответственно функция выживания $S(t) = 1 - F(t)$ равна вероятности того, что событие не состоится ранее, чем по истечении времени t .

Наиболее распространенными описательными методами исследования цензурированных данных являются построение таблиц дожития (mortality table) и метод Каплана—Мейера (Kaplan—Meier method). Для анализа используют несколько подходов, из которых мы остановимся на лог-ранк-тесте (логарифмический ранговый тест; англ. log-rank test) и модели пропорциональных интенсивностей Кокса (или модель пропорциональных рисков Кокса; англ. Cox Proportional Hazards Model).

1.1. Таблицы дожития

Таблицы дожития — один из наиболее традиционных методов исследования данных о выживаемости (происхождение интересующего нас события). В таблицах дожития время наступления события разбивается на интервалы, для каждого из которых определяются число и доля объектов: а) у которых событие не произошло на момент начала данного интервала времени; б) у которых событие произошло в течение данного интервала; в) которые были изъяты или цензурированы на данном интервале. По существу таблица дожития является расширенной таблицей частот. Считается, что для получения надежных оценок основных показателей (функции выживания, плотности вероятности и интенсивности, см. ниже) размер группы должен быть не менее 30.

На основании таблицы рассчитывается ряд индикаторов. Число изучаемых объектов — количество объектов, у которых событие не произошло на момент начала данного интервала времени, минус половина числа объектов, которые были изъяты или цензурированы. Доля "умерших" — отношение числа объектов, у которых событие произошло в течение данного интервала, к числу изучаемых объектов на данном временном интервале. Доля выживших — единица минус доля "умерших". Функция выживания (выживаемость) — кумулятивная доля объектов, у которых событие не произошло на момент начала определенного интервала времени; ее рассчитывают как произведение долей выживших на всех предыдущих интервалах. Плотность вероятности — оценка вероятности наступления события в каком-либо интервале; рассчитывается как отношение разности между значениями функции выживания на любом данном и последующем интервале к продолжительности данного интервала времени. Функция интенсивности представляет собой вероятность того, что на данном интервале произойдет событие у того объекта, у которого оно еще не произошло на момент начала этого интервала; вычисляется как отношение числа событий, происшедших в течение данного интервала, к числу объектов, у которых событие не произошло до момента времени, находящегося в середине этого интервала. Медиана ожидаемого времени жизни — точка на оси времени, в которой значение функции выживания равно 0,5; медиана ожидаемого времени жизни совпадает с точкой выживания 50% наблюдений только в том случае, если до этого момента времени цензурированных наблюдений не было. Аналогично через значения функции выживания можно определить и квантили (25-й и 75-й процентиля) ожидаемого времени жизни.

1.2. Метод Каплана—Мейера

Метод Каплана—Мейера используется для оценки доли объектов наблюдения (пациентов), у которых событие не произошло (функция выживания, выживаемость), для любого момента времени в течение всего периода наблюдения. Поскольку разбиение данных по временным интервалам

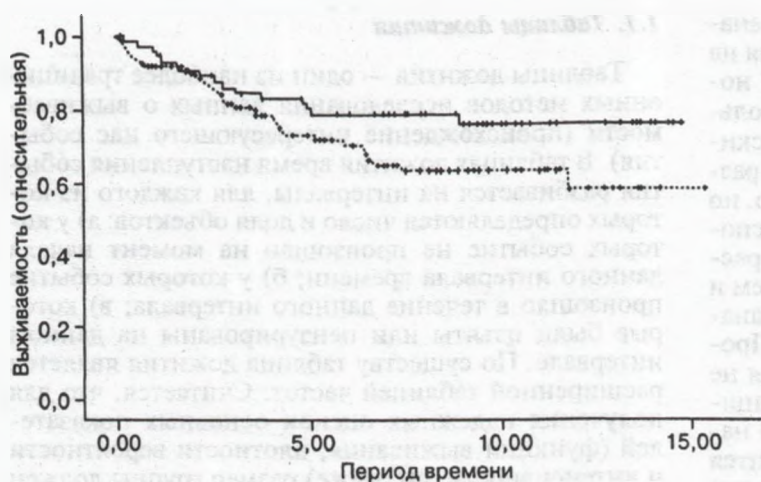


Рис. 1. Метод Каплана—Мейера (пример). Вертикальные штрихи отображают цензурированные наблюдения.

(группировка) не производится, суть метода Каплана—Мейера несколько отличается от таблиц дожития. В то же время результаты, получаемые с помощью этих двух методов, принципиально близки по смыслу. Оценка функции выживания в методе Каплана—Мейера представляет собой произведение выживаемости в данный момент времени на выживаемость в следующий момент времени, когда событие произошло.

Как и таблицы дожития, метод Каплана—Мейера полностью применим к цензурированным данным. Для расчетов используется истинное количество объектов, у которых событие еще не произошло в любой момент времени, для которого производится оценка. Отметим, что цензурированность данных может оказывать влияние на оценку функции выживаемости, в связи с чем метод Каплана—Мейера использует следующие предположения: а) цензурированные объекты ("выбывшие") имеют те же показатели выживаемости, что и те, которые продолжают наблюдаться (цензурирование не влияет на прогноз выживаемости); б) оценки выживаемости одинаковы для объектов, включенных в исследование на более ранних или более поздних сроках; в) событие происходит именно в анализируемый момент времени. Последнее предположение может искусственно завязать оценку выживаемости, если измерения производятся редко, так как определение момента времени наступления события откладывается до следующего обследования.

Метод Каплана—Мейера широко используется в клинических испытаниях, например, с целью оценки эффективности нового лекарственного препарата в изучаемой группе по сравнению с контрольной (получающей плацебо). Предположим, мы хотим выявить долю пациентов, у которых через 2 нед после начала применения нового препарата уровень холестерина в крови не понизился (как ожидалось) до определенного значения; определение холестерина производится 1 раз в неделю. Допустим, что в течение 1-й недели эффект наблюдался у 5 из 50 пациентов (оценка функции выживания 0,90; рассчитывается как $(50-5)/50$). Из 45 пациентов, у которых уровень холестерина не сни-

зился к началу 2-й недели, в течение нее эффект был отмечен у 9. "Выживаемость" в течение 2-й недели составила, таким образом, $36/45 = 0,80$. Общая 2-недельная выживаемость (доля пациентов, у которых эффект не наблюдался) в этом случае была $0,90 \cdot 0,80 = 0,72$. Поскольку для вычислений используется операция умножения, метод Каплана—Мейера называют также множительной оценкой.

Графическое представление метода Каплана—Мейера заключается в построении кривой выживаемости, отражающей пропорцию пациентов, у которых ожидаемое событие не произошло к определенному моменту времени. Временные интервалы определяются либо периодичностью контрольных обследований, либо временем до события в реальном масштабе (если известен момент происхождения события). Когда у объекта наблюдения происходит ожидаемое событие, производят перерасчет пропорции оставшихся в исследовании объектов, у которых событие не произошло, что отображается "ступенькой" вниз на кривой, как показано на рис. 1.

Кривые, построенные с помощью метода Каплана—Мейера, часто используются для оценки собственно выживаемости или безрецидивной выживаемости онкологических больных.

Бесспорное преимущество метода состоит в том, что он не требует знания о предполагаемой форме кривой выживаемости или характере распределения показателей выживаемости во времени. С другой стороны, будучи описательным средством, метод Каплана—Мейера имеет недостаток: он не позволяет сравнить выживаемость между группами, т. е. оценить достоверность различий кривых выживаемости.

1.3. Лог-ранк-тест

С помощью лог-ранк-теста (логарифмического рангового теста) можно оценить общую выживаемость в двух группах и более за весь период наблюдения, что является важным отличием от умозрительного сравнения показателей выживаемости в любой момент времени. Лог-ранк-тест принимает за нулевую гипотезу то, что выживаемость в сравниваемых группах пациентов не различается. Для всего периода наблюдения определяются ожидаемые и фактические показатели выживаемости для всех моментов времени происхождения события. Дальнейшие вычисления (сравнение ожидаемых и фактических значений) производятся с использованием теста χ^2 с целью выявления достоверности (без оценки ее степени и без доверительных интервалов) различий. Ряд статистических программ имеет специальные модули для выполнения лог-ранк-теста.

Лог-ранк-тест, как и тест Каплана—Мейера, применим к цензурированным выборкам и основывается на тех же предположениях о влиянии цензурированности данных на результат. С помощью лог-ранк-теста различия выживаемости в

группах обнаружить легче, если риск возникновения события в одной группе существенно и последовательно выше, чем в другой. Если кривые выживаемости вдруг пересекаются (например, при сравнении результатов хирургического лечения и тактики пассивного наблюдения прогрессирующей коронарной окклюзии), лог-ранк-тест вообще не способен выявить различие. В связи с этим при выполнении анализа с использованием теста необходимо вначале представить кривые выживания на графике.

1.4. Модель пропорциональных интенсивностей Кокса

Модель Кокса, часто называемая в литературе "пропорциональной моделью Кокса", — наиболее используемый в современных публикациях и рекомендуемый инструмент анализа данных выживаемости. В ее основе лежит метод множественной регрессии (см. раздел 2), и в качестве выходного параметра модель возвращает значение отношения рисков и его доверительный интервал. Отношение рисков (hazard ratio — HR) — это оценка отношения интенсивностей (показателей, уровней, функций) риска в экспериментальной и контрольной группах, рассчитанная для любого момента времени наблюдения. Модель предполагает, что HR у членов экспериментальной и контрольной групп остается неизменным в течение всего периода наблюдения (предположение о пропорциональности, англ. proportionality assumption). Интенсивность риска представляет собой вероятность того, что событие, не произошедшее к определенному моменту времени, случится в следующий интервал времени, отнесенную к продолжительности этого интервала. Временной интервал может быть установлен очень коротким, поэтому оценку можно делать для любого момента времени. Говоря другими словами и применительно к клиническому испытанию, в котором ожидаемым результатом является, например, выздоровление пациента, HR отражает относительную вероятность быстрого выздоровления у больных, получающих лечение, по отношению к пациентам контрольной группы для любого момента времени.

Данная модель позволяет включать в исследование всех интересующих нас пациентов, невзирая на цензурирование (частичную неполноту данных), поскольку использует базисное допущение о том, что пациенты выбывают случайным образом и с одинаковой вероятностью как в изучаемой, так и в контрольной группе. Кроме того, изначально предполагается, что пациенты, у которых произойдет или не произойдет событие, выбывают из исследования с одинаковой вероятностью (правила пропорциональности модели).

Пропорциональная модель Кокса в последнее время получает все большее признание и популярность в биомедицинских исследованиях. С точки зрения информативности выходных статистических характеристик, она предоставляет возможность провести более точный и взвешенный анализ выживаемости, чем рассмотренные выше, по-

скольку позволяет включить в расчеты целый набор переменных, влияющих или предположительно влияющих на исход.

Ввиду частого использования рассматриваемой модели попытаемся поглубже понять интерпретацию результатов обработки данных выживаемости с ее помощью. Сразу заметим, что хотя HR может быть применено к любому моменту периода наблюдения, сам по себе этот показатель не дает непосредственного представления о времени до события. HR может показывать наличие положительного эффекта препарата в клиническом испытании (когда HR достоверно превышает 1), что действительно предполагает укорочение времени до выздоровления. При этом значение HR может быть меньше, больше или иногда равным отношению медиан ожидаемого времени жизни (см. раздел 1.1), что свидетельствует о том, что это две разные статистические характеристики.

В литературе можно встретить такие суждения, основанные на значении HR, как: "ускорение периода выздоровления", "выздоровление было в столько-то раз (или на столько-то процентов) более быстрым". К примеру, $HR = 2$ может быть истолковано исследователями (это встречается в литературе) в том смысле, что пациенты, получавшие препарат, выздоравливали в 2 раза быстрее. Определение "в 2 раза быстрее" может (теоретически) быть понято так, что медиана ожидаемого времени эффекта (выздоровления) снизилась в результате лечения в 2 раза; что количество выздоровевших на какой-то день было в 2 раза больше в группе, получившей лечение; что ожидаемое количество выздоровевших на какой-то день было в 2 раза больше в группе, получившей лечение. Ни одно из этих утверждений не является примером верной интерпретации результата анализа. Значение HR не должно восприниматься, как имеющее отношение к реальной "скорости" процесса выздоровления. $HR = 2$ в общем предполагает более быстрое выздоровление, но его понимание должно восприниматься под специфическим "вероятностным" углом зрения. Наиболее корректной "расшифровкой" $HR = 2$ было бы, что пациент, получающий препарат и не выздоровевший до какого-то момента времени, имеет в 2 раза больший шанс выздороветь к следующему моменту времени, чем тот, кто получал плацебо. Эта интерпретация кардинально отличается от тех интуитивных формулировок, которые были приведены выше. В более широком смысле HR эквивалентно шансу того, что у члена группы высокого риска событие наступит раньше, чем у члена группы меньшего риска. Вероятность того, что событие наступит раньше, может быть рассчитана из показателя HR по формуле: $p = HR / (1 + HR)$. Таким образом, $HR = 2$ соответствует 67% шансу более раннего наступления события (например, выздоровления) у пациента, получавшего препарат, чем у того, который получал плацебо.

Алгоритм пропорциональной модели Кокса и расчета HR, а также, что очень важно, оценка доверительных интервалов реализованы в некоторых статистических программных пакетах.

2. Многомерный анализ

Методы многомерного анализа (англ. — multivariate, или multivariable analysis) разработаны для оценки одновременного влияния более чем одного фактора на результат (исход). В отличие от одномерной статистики, которая дает оценку того, как каждая (одна) переменная связана с интересующим нас результатом, многомерная статистика дает информацию о степени влияния на исход каждой из (многих) переменных, а также об эффекте взаимодействия этих переменных между собой. В отечественной литературе многомерный анализ часто называют многофакторным анализом.

Например, выдвигается предположение о том, что пациенты травматологического отделения, оперированные в 9 ч утра, имеют более высокий показатель смертности, чем пациенты, оперированные в 9 ч вечера. Анализируя смертность после операции одномерным (параметрическим или непараметрическим) статистическим методом, исследователь, допустим, действительно может обнаружить достоверную разницу. Однако было бы ошибочным полагать, что время суток является единственным определяющим фактором. В анализ должны быть включены и другие переменные, такие как тяжесть травмы, возраст больного, плановые или срочные показания к операции и т. п. После этого время операции, скорее всего, будет исключено из набора факторов, влияющих на смертность пациентов, или его вклад окажется очень малым по сравнению с истинными причинами. Лишь используя многомерный анализ, исследователь может сделать обоснованный вывод о причинах, влияющих на вариабельность результата, и оценить степень одновременного влияния на него этих нередко (в той или иной степени) взаимосвязанных причин.

Факторы (причины), влияющие на исход, принято называть факторами риска (risk factors), независимыми (independent) или объясняющими переменными (explanatory variable), а сам исход (outcome) — зависимой (dependent) или переменной отклика (response variable), или эффектом.

Важным моментом, обуславливающим необходимость многомерного анализа, является именно многообразие потенциальных факторов риска, возможно связанных с исходом. Экспериментальная проверка совместного влияния многих факторов в клинической практике чаще всего просто невозможна или недопустима по этическим соображениям. Положим, нужно выяснить, повышает ли курение вероятность ишемической болезни сердца (ИБС) в двух случайных выборках людей, одни из которых курят, другие — нет (заставить человека курить или не курить на протяжении долгого времени и невозможно, и неэтично). Хотя предварительный одномерный статистический анализ показывает, что ИБС у курящих развивается с большей вероятностью, чем у некурящих, этот результат сам по себе еще не является доказательством причинной связи курения с ИБС, хотя может и указывать на наличие таковой. В принципе нельзя исключить, что более вескими причинами развития ИБС у курящих является то, что в большинстве своем

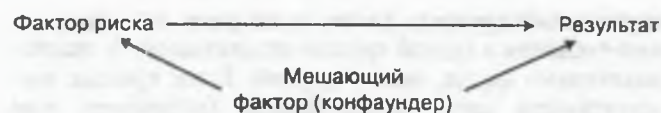


Рис. 2. Схема взаимодействия факторов, влияющих на результат (в данном случае — ИБС).

это малообеспеченные и ведущие малоподвижный образ жизни мужчины. Привычка к курению характерна именно для такой категории лиц, а все перечисленные качества повышают риск ИБС, что известно из других исследований.

В рассмотренном примере на связь между курением и ИБС могут влиять так называемые мешающие факторы (конфаундеры; от англ. — confounders). О мешающих факторах говорят, когда на видимую связь между фактором риска и независимой переменной (результатом) воздействует третья переменная, влияющая как на фактор риска, так и непосредственно (причинно) на сам результат, как показано на рис. 2. Мужской пол, малообеспеченность и малоподвижный образ жизни вполне могут оказаться мешающими факторами, поскольку они ассоциированы как с курением, так и с ИБС.

В итоге многомерный анализ показывает, что даже с поправкой (учетом, нормализацией; англ. adjustment) на мужской пол, малообеспеченность или малоподвижный образ жизни курение оказывает независимое влияние на развитие ИБС (рис. 3).

Определить, какая переменная является независимым фактором риска, а какая конфаундером, иногда невозможно: один и тот же фактор может как оказывать независимый эффект на результат, так и быть мешающим фактором, влияющим на другую переменную. Возвращаясь к ситуации с ИБС, малообеспеченность является мешающим фактором взаимосвязи между курением и заболеванием: у малообеспеченных людей повышена вероятность злоупотребления курением, а у курящих — вероятность развития ИБС. С другой стороны, малообеспеченность сама по себе влияет на развитие ИБС: даже с учетом (поправки) курения, уровня холестерина крови, артериальной гипертензии и прочих факторов у малообеспеченных лиц ИБС развивается с большей вероятностью, чем у людей с высоким достатком.

Заметим, что многомерный анализ — единственный статистический метод учета влияния или исключения конфаундеров. Оценить влияние фактора риска на исход можно также с помощью условного анализа, при котором изучаемую группу последовательно разбивают на подгруппы (страты), в которых какая-либо потенциально мешающая переменная "фиксируется". При этом модель применяется отдельно на каждой группе данных. И на каждой группе будут получены разные эффект и разные оценки параметров модели. Например,

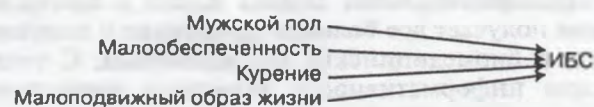


Рис. 3. Оценка степени влияния различных потенциальных факторов на результат (в данном случае — ИБС).

мужской и женский пол: далее рассматривают только курящих или некурящих мужчин и только курящих или некурящих женщин. Условный анализ эффективен, когда изучают относительно небольшое число факторов (2–3). Если же потенциально мешающих факторов много, дробление приводит к образованию большого числа малочисленных выборок, в которых оценки рисков могут оказаться нестабильными (см. раздел 4 в части I).

Стратификация (stratification) является способом включения в одну модель всех групп данных с разными значениями мешающих факторов таким образом, чтобы зависимость эффекта от интересующих исследователя независимых переменных подгонялась по всему множеству данных, а зависимость от мешающих факторов — только по соответствующим подгруппам. Например, пропорциональная модель радиационного риска может состоять из произведения двух членов: фоновой риска, который зависит от пола, и относительного радиационного риска, который считается одинаковым для обоего пола. В этом случае говорят, что в модели риска произведена стратификация фоновой риска по полу. В общем случае, чтобы определить коэффициенты модели при интересующих независимых переменных, нужно оценить всю модель, включая коэффициенты при мешающих факторах.

В случае применения стратификации иногда удается построить модель, в которой оценка интересующих исследователя переменных может быть проведена без оценки мешающих факторов. Построение таких моделей (а соответственно и схем исследования или выборок) с математической точки зрения весьма нетривиально. Примером является модель случай — контроль с подбором контролей к случаям по значениям мешающих параметров, разбитых на страты (matched case-control study). В этой модели отношение шансов по интересующему фактору определяется без оценки мешающих факторов. В силу математических обстоятельств такие модели чаще всего применяются на условных выборках, которые абсолютно искажают величину эффекта в популяции, но позволяют все-таки оценить степень статистической связи между изучаемой переменной и эффектом. Например, в условной подобранной модели случай — контроль (conditional matched case-control model) к каждому из N случаев подбирается M контролей по полу, возрасту, месту проживания и ряду других признаков. Ясно, что в полученной выборке частота случаев не имеет никакого отношения к частоте случаев в исходной популяции. Но отношение шансов, как мера статистической связи между изучаемым фактором и эффектом, определяется незави-

симо от мешающих факторов, по которым производился подбор контролей к случаям. Такой способ позволяет анализировать влияние факторов риска с высокой мощностью даже в сравнительно небольших выборках. Однако надо отдавать себе отчет в том, что условные модели, основанные на условных выборках — это не то же самое, что модели, основанные на натуральных выборках (не искажающих статистических свойств популяции), в которых оценивается полный набор всех переменных. Например, отношения шансов, полученных в когортной модели заболеваемости и в условной модели случай — контроль, различны, хотя и численно близкие в не слишком экзотических случаях. Напомним, что терминологически не следует путать условный анализ путем независимой подгонки модели на независимых подгруппах и условные модели, основанные на условных выборках.

2.1. Виды многомерного анализа

В клинических исследованиях в зависимости от задачи и типов данных чаще всего используют 3 метода: множественная линейная регрессия (multiple linear regression), множественная логистическая регрессия (multiple logistic regression) и модель пропорциональных интенсивностей Кокса. В таблице представлены наиболее существенные характеристики этих трех моделей.

Множественная линейная регрессия используется для изучения изменения зависимой переменной (y) в ответ на различные значения других переменных (x_1, x_2, x_3), которые представляют собой непрерывные (численные интервальные или относительные) переменные. Модель предполагает, что с увеличением (или уменьшением) значений независимых переменных зависимая переменная (исход) изменяется линейно. Для линеаризации нелинейно влияющих независимых переменных часто используется математическое приведение, такое, например, как логарифмирование. Для интервальных переменных также предполагается, что для равных промежутков на всей шкале интервала степень влияния на исход будет одинакова. Величина коэффициента при независимой переменной и его знак в конечной модели показывают степень и характер связи между этой переменной и исходом. Примером множественной линейной регрессии может быть модель оценки костной плотности у женщин в менопаузе, в которую входят возраст и индекс массы тела.

Логистическая регрессия используется, когда значение переменной результата является бинарным, таким как выживаемость (да/нет), развитие

Основные модели многомерного анализа

Модель	Тип данных	Специфические особенности
Множественная линейная регрессия	Диапазон значений исхода (например, показатели артериального давления)	Коэффициенты при переменной линейно связаны с результатом (или влияют на него)
Логистическая регрессия	Дихотомические результаты (да/нет)	Модель ограничивает вероятность исхода от 0 до 1
Модель пропорциональных интенсивностей	Период времени до события (время до выздоровления/рецидива/смерти)	Применяется для продолжительных исследований, в которых объекты могут быть потеряны во время наблюдения

заболевания (да/нет), положительный результат диагностического теста (да/нет), и может включать одну независимую переменную или более. К примеру, для прогнозирования летальности пациентов травматологического отделения (см. выше) может быть использована модель, которая учитывает возраст пациента, степень тяжести и причину травмы. Логистическая регрессия при использовании модели пропорционального риска позволяет оценить относительный риск (relative risk — RR), отношение шансов и границы их доверительных интервалов (confidential intervals — CI), а также степень достоверности (величину p) отличия этих величин от 1 (значение при нулевой гипотезе). Значения OR и 95% CI предоставляют наглядную информацию о взаимосвязи независимой переменной с исходом. OR близкое к 1 свидетельствует о слабой взаимосвязи. Широкий CI наводит на мысль о невысокой надежности оценки и о необходимости проверки данных на предмет их точности. Если границы CI включают значение 1, связь независимой переменной с исходом не может быть признана достоверной, насколько бы значение OR не отличалось от 1.

Модель пропорциональных интенсивностей Кокса оценивает шансы более раннего наступления события у членов изучаемой группы по сравнению с контрольной группой с помощью показателя HR, как рассмотрено в разделе 1.4.

2.2. Включение независимых переменных в модель

Любая модель многомерного анализа должна включать как минимум один или несколько факторов риска и потенциальные мешающие факторы (конфаундеры). Однако универсального рецепта включения как факторов риска, так и потенциальных конфаундеров не существует. В связи с этим подход к вопросу должен быть очень осторожным. В идеале в модель нужно включить все переменные, которые были определены с помощью теоретических рассуждений или установлены в предыдущих исследованиях как факторы риска или конфаундеры изучаемого исхода. Так, для выяснения влияния гиперальбуминемии на смерть от сердечно-сосудистых заболеваний в многомерную модель пропорциональных рисков должны быть включены возраст, пол, статус курения, наличие артериальной гипертензии, дислипидемия, диабет, ожирение, уровень креатинина в крови и другие факторы, известные как связанные с исходом.

С другой стороны, важно не только включение в анализ всех потенциально важных переменных, но и исключение посторонних. Например, в модель изучения связи курения и рака легкого нет смысла включать наличие ИБС в список факторов риска или конфаундеров, хотя риск ИБС повышен у курящих. Ни теоретически, ни по опыту предыдущих исследований ИБС никак не была связана с развитием изучаемой онкопатологии. Если массив данных включает очевидно связанные между собой или сильно коррелирующие независимые переменные, рекомендуется произвести сознательный выбор одной как наиболее важной. Например, исследование причин неонатальной смертности показало, что поскольку масса тела при рождении и срок

гестации очень взаимосвязаны, нет необходимости включать в модель обе переменные. Авторы исключили срок гестации, аргументируя это тем, что масса тела при рождении оказывает схожее, но более сильное влияние на исход.

На выбор включаемых в анализ переменных также влияет предназначение создаваемой модели, которая может быть объясняющей (explanatory) или прогностической (prognostic). Назначением объясняющей модели является выяснение характера и степени влияния различных факторов на результат. Для такой модели тщательный подбор переменных и их математической формы крайне важен. Прогностические модели направлены на определение вероятности происхождения события. Если созданная модель хорошо воспроизводится и эффективно работает на независимых массивах данных, то вопросы подбора и ревизии переменных в ней уже не актуальны. В прогностической модели наиболее важна точность предсказания результата. Например, в нее могут быть включены коррелирующие независимые переменные, и модель будет хорошо работать. Но оцененные коэффициенты модели для таких переменных не имеют самостоятельного значения и не несут информации о степени влияния на эффект каждой независимой переменной в отдельности. Такая модель не может быть объясняющей.

Количество переменных в модели можно оптимизировать с помощью автоматизированных алгоритмов их подбора. Эти алгоритмы помогают компьютеру отобрать переменные на основе критериев, определенных исследователем. Отбор осуществляется методами прямого пошагового отбора, обратного пошагового удаления и наилучшего подмножества. Для прямого пошагового отбора переменная, оказывающая наиболее сильное влияние на исход по результатам одномерного анализа, вводится первой, следом за ней добавляется переменная со следующим наиболее сильным влиянием и т. д. до тех пор, пока все переменные, влияющие на результат (с уровнем значимости, определенным исследователем; обычно в многомерном регрессионном анализе он устанавливается $> 90\%$, т. е. $p < 0,1$), не будут включены в модель. Любая ранее введенная в модель переменная, которая перестает быть значимой при введении следующей переменной, последовательно исключается. В методе обратного пошагового удаления в модель сначала включаются все переменные. Затем они последовательно удаляются, начиная с переменной, имеющей наиболее слабую ассоциацию с результатом. Удаление продолжается до тех пор, пока в модели не останутся только те переменные, которые достоверно влияют на исход. Метод наилучшего подмножества подразумевает выбор путем подстановки такого набора переменных, которые наилучшим образом удовлетворяют условиям, определенным исследователем. При автоматизированном подходе к выбору переменных может получиться так, что не все основные и мешающие переменные окажутся в модели или в модели могут отсутствовать наиболее значимые клинические показатели. Поэтому после создания модели в автоматизированном режиме от

исследователя требуется ее критическая оценка на адекватность.

2.3. Взаимодействие между переменными

О взаимодействии между переменными (interactions) в эффекте говорят, когда влияние фактора риска на исход (эффект) зависит от значения третьей синтетической переменной, составленной из двух исходных независимых переменных. При этом сама третья переменная не является независимым фактором риска или мешающей переменной. Взаимодействие между переменными также называют эффектом модификации (effect modification) в том случае, если одна из них рассматривается как основная с содержательной точки зрения.

Например, клинические испытания продемонстрировали, что некоторый препарат снижает вероятность перелома кости у пациентов с остеопорозом, но не у людей с изначально более высокой минеральной плотностью костной ткани (условный анализ на разных группах пациентов). Чтобы свести весь доступный материал в одну модель (для увеличения достоверности и мощности исследования), был введен член взаимодействия. Исследование проводили с использованием пропорциональной модели Кокса с учетом переменных приема препарата, результата анализа минеральной плотности костной ткани, а также синтетической переменной, которая представляла собой композицию (для двух непрерывных переменных было бы произведение) первых двух переменных. Достоверность влияния этой синтетической переменной на исход означала, что эффект препарата зависит от начальной минеральной плотности костной ткани.

Учет взаимодействия между переменными может оказаться клинически важным, однако для внесения эффекта взаимодействия в модель необходимо изначально иметь предположение о том, что переменные могут взаимодействовать. В противном случае начинается почти системный поиск взаимодействия путем разбиения групп на подгруппы, и чем больше ожидается взаимодействующих переменных, тем больше образуется подгрупп данных. Это может привести к тому, что в одной или нескольких из них взаимодействие будет обнаружено в силу случая (ошибка 1-го типа).

2.4. Анализ качества модели

Для оценки эффективности множественной линейной регрессии используется уже известный из корреляционного анализа коэффициент детерминации r^2 (см. раздел 7.1 части I обзора), который отражает степень рассеяния результата, возникающего благодаря вкладу многих переменных. Значение r^2 варьирует в пределах от 0 до 1, и чем ближе оно к 1, тем лучше модель описывает результат. Ввиду того что в модель может входить несколько факторов риска, коэффициент рассчитывается с поправкой на их количество.

Для логистических регрессий предложено несколько статистических критериев согласия (goodness-of-fit test), каждый из которых имеет свои дос-

тоинства и недостатки. Чаще применяется тест Хосмера—Лемешова (Hosmer—Lemeshow test). Критерии согласия обычно используются для оценки эффективности объясняющих моделей. О надежности объясняющих моделей можно судить по их воспроизводимости на других массивах данных. Если модель надежна, то и в независимом массиве данных в модель войдут те же факторы риска с коэффициентами, близкими к тем, что наблюдались в оригинальной модели.

Для прогностических моделей рекомендуется более точная количественная оценка. Для ее получения рассчитываются такие показатели, как чувствительность, специфичность и точность (см. раздел 8 части I обзора) для определенных пороговых условий (например, исходя из предположения, что у всех лиц, у которых предсказанная вероятность заболевания была 40% и выше действительно развилось заболевание). Исходя из значений чувствительности и специфичности, строится характеристическая кривая (ROC-кривая, см. раздел 8 части I), по форме которой и величине площади под которой (AUC) можно судить об удачности модели. Надежность прогностической модели также может быть проверена на независимых массивах данных, в которых она должна предсказывать вероятность исхода с высокой эффективностью.

Завершая рассмотрение многомерного анализа, заметим, что, хотя его алгоритмы реализованы в ряде статистических программ, он требует специальных знаний и подготовки в математической статистике. Поскольку наиболее важные для клиницистов выводы обычно делаются на основании именно многомерного анализа, его корректное выполнение и интерпретация имеют особое значение.

Заключение

В данном обзоре мы остановились на ряде наиболее часто используемых в медицинских исследованиях методов статистического анализа. Статистический анализ является неотъемлемой частью практически любого исследования, и только с его помощью можно пополнить доказательную базу. Исходя из собственного опыта, осмелимся высказать мнение, что наиболее значимые и глубокие выводы делаются на основании всестороннего и тщательно проведенного статистического анализа, в котором могут использоваться довольно сложные алгоритмы. В связи с этим его самостоятельное выполнение клиницисту не всегда по силам. Во многих случаях необходимо участие специалиста с профессиональной подготовкой в области математической статистики. Именно в ходе сотрудничества клинициста с математиком можно рассчитывать на проведение глубокого и корректного статистического анализа данных.

Чтобы сделать диалог более продуктивным, клиницист должен произвести ревизию информации со статистических позиций. Убедившись, что исследование обладает достаточной мощностью, необходимо составить себе четкое представление об имеющихся данных: их типах, распределении и пр. Далее необходимо охарактеризовать данные с помощью описательной статистики и произвести

их одномерный статистический анализ, после чего можно сформировать вопросы для многомерного анализа. Это будет несомненным подспорьем при обсуждении дальнейшего анализа с математиком, что значительно повышает шансы на взаимопонимание и в итоге на успешный результат. Мы последовательно рекомендуем подобное сотрудничество и являемся его сторонниками. На наш взгляд, оно является наиболее рациональным подходом к правильной интерпретации результатов и формированию корректных выводов.

Выражаем надежду, что представленное описание методов статистического анализа окажется полезным клиницистам, особенно тем, кто находится в начале своего профессионального пути.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Гланц С. Медико-биологическая статистика: Пер. с англ. — М., 1999.

- Bland J. M., Altman D. G. // Br. Med. J. — 1998. — Vol. 317. — P. 1572.
 Bland J. M., Altman D. G. // Br. Med. J. — 2004. — Vol. 328. — P. 1073.
 Cassidy L. D. // J. Surg. Res. — 2005. — Vol. 128, N 2. — P. 199–206.
 Davis C. S. Statistical Methods of the Analysis of Repeated Measurements. — New York, 2002.
 Katz M. H. // Ann. Intern. Med. — 2003. — Vol. 138, N 8. — P. 644–650.
 Kirkwood B. R., Sterne J. A. Essential Medical Statistics. — 2-nd Ed. — Oxford, 2003.
 Machin D., Cheung Y., Parmar M. Survival Analysis: A Practical Approach. — 2-nd Ed. — New York, 2006.
 Petrie A., Sabin C. Medical Statistics at a Glance. — Oxford, 2005.
 Rao S. R., Schenfeld D. A. // Circulation. — 2007. — Vol. 115. — P. 119–123.
 Royston P., Parmar M. K., Altman D. G. // J. Natl. Cancer Inst. — 2008. — Vol. 100, N 2. — P. 92–97.
 Spruance S. L., Reid J. E., Grace M., Samore M. // Antimicrob. Agents Chemother. — 2004. — Vol. 48, N 8. — P. 2787–2792.

Поступила 15.01.09

Информация для авторов ТРЕБОВАНИЯ К РИСУНКАМ, представленным на магнитных носителях

Черно-белые штриховые рисунки:

- формат файла — **TIFF** (расширение *.tif), любая программа, поддерживающая этот формат (Adobe PhotoShop, CorelDRAW, Adobe Illustrator и т. п.);
- режим — **bitmap** (битовая карта);
- разрешение — **600 dpi** (пиксели на дюйм);
- серые и черные заливки должны быть заменены на косую, перекрестную или иную штриховку;
- рисунок должен быть **обрезан** по краям изображения и **очищен** от "пыли" и "царапин";
- ширина рисунка — **не более 180 мм**, желательно не использовать ширину от 87 до 150 мм;
- высота рисунка — не более 230 мм (с учетом запаса на подрисовочную подпись);
- размер шрифта подписей на рисунке — **не менее 7 pt** (7 пунктов);
- возможно использование сжатия LZW или другого;
- носители — floppy 3.5" (1,44 MB), Zip 100 MB, CD-ROM, CD-R, CD-RW;
- обязательно наличие распечатки.

Цветные изображения, фотографии и рисунки с серыми элементами:

- платформа (компьютер) — IBM PC или совместимый;
- формат файла рисунка — **TIFF** (расширение *.tif);
- программа, в которой выполнена публикация, — PageMaker 6.5; CorelDRAW 7 и 8;
- цветовая модель — **CMYK**;
- разрешение — не более 300 dpi (пиксели на дюйм) или 119,975 пикселя на 1 см;
- рисунок должен быть связан с публикацией;
- возможно использование сжатия LZW;
- не использовать цвета **PANTONE**;
- носители — Zip 100 MB; компакт-диск CD-ROM.