

Некоторые статистические показатели, используемые при обработке данных о вторых опухолях

Кулева С.А., Карицкий А.П.

*ФГБУ «НИИ онкологии им. Н.Н. Петрова» Минздрава России,
19775, С-Петербург, п. Песочный, Ленинградская ул., д. 68*

Phone: (812).596 65 43

Fax: (812).596 89 47

E-mail: Kulevadoc@yandex.ru

Резюме. Основная цель статьи - ознакомление с некоторыми понятиями статистики, на которые опираются исследователи при обработке данных. Арсенал статистических методов очень широк, и не всегда понятно, какой алгоритм организации работы следует использовать. В статье представлены определения статистических показателей, используемых при обработке данных о вторых опухолях, которые даны на понятийном уровне, теоретические основы сведены к минимуму, и главное внимание уделено практической стороне анализа данных.

Ключевые слова: вторые опухоли, статистические показатели

Some statistics being used at data processing about the second tumors

Kulyova S.A., Karytsky A.P.

*Department of Children's Oncology, N.N. Petrov Research Institute of Oncology,
68 Leningradskaya Str., Pos. Pesochny, St. Petersburg, 197758, Russia*

Phone: (812).596 65 43

Fax: (812).596 89 47

E-mail: Kulevadoc@yandex.ru

Abstract. The objective of the article is acquaintance with some statistics on which researchers lean at data processing. The arsenal of statistical methods is very wide, and it is not always clear, what algorithm of the organization of work should be used. In article definitions of statistics processing about the second tumors which are given at conceptual level are presented, theoretical bases are minimized, and the main attention is given to a practical side of the analysis of data.

Key words: second tumors, statistics

«Статистика, возможно, знает все, но ее знают не все»

Александр Самойленко

Основная цель статьи – ознакомление и приобщение к некоторыми понятиями статистики, на которые опираются исследователи при обработке данных. Арсенал статистических методов очень широк, и не всегда понятно, какой алгоритм организации работы следует использовать. Иногда аббревиатура переменной ставит в тупик читателя специальной литературы. Особенно это часто встречается в публикациях с онкологической тематикой. Статистические показатели при обработке данных о вторых опухолях еще более не известны, хотя, в какой-то мере, и пересекаются с показателями, характеризующими первичное новообразование. В статье представлены определения, которые даны на понятийном уровне, теоретические основы сведены к минимуму, и главное внимание уделено практической стороне анализа данных.

Основным статистическим параметром в онкологической практике можно считать заболеваемость, показатели которой могут вычисляться в абсолютных значениях, однако, для аналитической статистики целесообразно использовать относительные величины, т.е. принимать во внимание численность населения, в которой эти случаи заболевания возникли [1].

Кумулятивный коэффициент. Частота возникновения второй опухоли может определяться с помощью кумулятивного коэффициента или коэффициента риска (CIR - *cumulative incidence rate*), представленного следующей формулой:

$$CIR = \frac{\text{число заболевших лиц в определенный период времени}}{\text{численность популяции риска} * \text{в тот же период времени}}$$

* - число лиц, у которых существует вероятность возникновения второй опухоли

И числитель, и знаменатель представляют собой число лиц, которые в начале искомого периода являются относительно здоровыми (излеченными от первичного злокачественного заболевания) и подвергаются риску заболеть второй опухолью. Другими словами, коэффициент риска – это средний риск заболеть второй опухолью для пациента за рассматриваемый период времени.

Когда получившийся показатель имеет несколько нулей после запятой, то с целью облегчения информации используется множитель 10^n .

Пример. Если в популяции риска численностью 100.000 возникло 10 случаев вторых опухолей, то кумулятивный риск составляет 0,0001. Воспользовавшись множителем 10^n , CIR будет равен 1 случай на 10.000.

Кумулятивный коэффициент представляет собой долю людей, которые могут заболеть второй опухолью за исчисляемый период времени. Таким образом, числитель представляет собой сокращенную версию знаменателя. Он может рассматриваться, как средний риск заболеть для члена группы в течение этого временного периода.

Коэффициент заболеваемости (IR - incidence rate) является основным показателем частоты заболеваемости второй опухолью.

$$IR = \frac{\text{число случаев заболевания в группе населения на протяжении определенного промежутка времени}}{\text{сумма длительности риска заболевания каждого лица в данной группе}}$$

Сумма длительностей риска измеряется в годах и называется человеко-годы. Для каждого лица в данной когорте риск заболевания является временем, в течение которого он принадлежит к исследуемой группе и не имеет данного заболевания, и, следовательно, рискует заболеть. Периоды риска для каждого члена когорты суммируются. Коэффициент заболеваемости равен делению числа случаев заболевания на производное численности группы населения и продолжительность периода времени, равного сумме периодов времени для каждого лица в данной группе. В формуле учитываются продолжительность периода наблюдения, а также лица, которые присоединяются к данной когорте или покидают ее в период наблюдения вследствие смерти по какой-либо иной, не связанной с второй опухолью причине. Данная формула учитывает ошибки, встречающиеся при вычислении показателя кумулятивной заболеваемости.

Пример. По данным R. Sankila и соавт. (1996) среди излечившихся от первичной лимфомы Ходжкина (n=1641) отмечено 7 случаев острого миелобластного лейкоза. Число человеко-лет в этой группе составило 17.000. Коэффициент заболеваемости равен:

$$IR = \frac{7}{17.000} = 0,0004$$

Кумулятивный коэффициент при этом составляет:

$$CIR = \frac{7}{1641} = 0,004$$

В эпидемиологической практике нередко сравниваются частота заболевания среди лиц, имеющих определенную патологическую нозологию, с соответствующей частотой среди тех, у кого этого заболевания нет. Сравнимые группы являются экспонированными и неэкспонированными.

Частота заболевания может характеризоваться как кумулятивным коэффициентом, так и коэффициентом заболеваемости. Сравнение этих показателей может быть абсолютным и относительным. Абсолютное сравнение – это разница в частоте заболеваемости между экспонированной и неэкспонированной группами. Относительное сравнение – отношение частоты заболевания в этих группах. Относительное сравнение имеет более широкое распространение и в большей степени соответствует целям научных исследований.

Показатель относительного сравнения обычно называется *относительным риском* (*RR – relative risk*).

Пример. Кумулятивная заболеваемость второй опухолью за 40-летний период составляет 0,0030 для наследственно-обусловленной ретинобластомы и 0,0001 для ненаследственной формы ретинобластомы [3]. Относительный риск в данном примере составляет:

$$RR = \frac{0,0030}{0,0001} = 30$$

Вероятность появления второй опухоли в группе с наследственно-обусловленной ретинобластомой в 30 раз превышает таковую в группе спонтанных ретинобластом.

Показатели смертности и заболеваемости часто вычисляются для неоднородных по своему составу совокупностей (возрастному, половому составу). Эта неоднородность отражается на показателях. В данных случаях следует уравнивать группы и вычислить показатели. Для этой цели использует метод стандартизации, который позволяет исключить влияние неоднородности состава на величины заболеваемости и смертности населения [1].

Пример. В табл. 1 представлены результаты обследования двух совокупностей, которые должны сравниваться на частоту заболевания. Сравнение можно проводить, руководствуясь грубыми коэффициентами – 0,015 и 0,023. Повозрастное распределение в них различно. Совокупность с более высоким грубым коэффициентом заболеваемости имеет большую долю взрослых людей.

Табл. 1

Число человеко-лет и число случаев заболевания в экспонированной и контрольной совокупностях

ПОВОЗРАСТНОЕ РАСПРЕДЕЛЕНИЕ	ЭКСПОНИРОВАННАЯ СОВОКУПНОСТЬ			КОНТРОЛЬНАЯ СОВОКУПНОСТЬ		
	Число человеко- лет	Число случаев	Коэффициент заболеваемост и	Число человеко- лет	Число случаев	Коэффициент заболеваемости
Дети	3.000	30	0,010	1.000	5	0,005
Взрослые	1.000	30	0,030	9.000	225	0,025
ВСЕГО	4.000	60	0,015	10.000	230	0,023

Для представленной в табл. 1 экспонированной группы грубый коэффициент заболеваемости равен:

$$(3.000/4.000) \times 0,010 + (1.000/4.000) \times 0,030 = 0,015$$

Одним из способов повышения обоснования сравнения в группах является стандартизация. В данном примере необходимо произвести перерасчет для стандартизации по возрасту. Вычисляются грубые коэффициенты, если бы возрастное распределение групп было аналогично возрастному распределению в стандартной группе. В таком случае экспонированная и контрольная совокупности сливаются с образованием некоей стандартной группы. Тогда в стандартной группе будет 4.000/14.000 детей и 10.000/14.000 взрослых. Стандартные коэффициенты заболеваемости составят:

$$(4.000/14.000) \times 0,010 + (10.000/14.000) \times 0,030 = 0,024$$

И

$$(4.000/14.000) \times 0,005 + (10.000/14.000) \times 0,025 = 0,019$$

Разница или отношение этих стандартизованных показателей могут использоваться для проведения абсолютного или относительного сравнений. В примере отношение стандартизованных показателей составляет:

$$\frac{0,024}{0,019} = 1,26$$

Данный пример является *методом прямой стандартизации*.

Пример. В случае отсутствия информации о распределении по возрасту, сравнение часто производится между наблюдаемым числом случаев (O – observed) и ожидаемым числом случаев (E – expected) в экспонированной совокупности (табл. 2).

Табл. 2

Число человеко-лет и число случаев заболевания в экспонированной и контрольной совокупностях

ПОВОЗРАСТНОЕ РАСПРЕДЕЛЕНИЕ	ЭКСПОНИРОВАННАЯ СОВОКУПНОСТЬ			КОНТРОЛЬНАЯ СОВОКУПНОСТЬ		
	Число человеко- лет	Число случаев	Коэффициент заболеваемости и	Число человеко- лет	Число случаев	Коэффициент заболеваемости
Дети	3.000	-	-	1.000	5	0,005
Взрослые	1.000	-	-	9.000	225	0,025
ВСЕГО	4.000	60	0,015	10.000	230	0,023

В данном примере O = 60. Ожидаемое число случаев – это число случаев, которые произошли бы в экспонированной когорте населения, если бы все повозрастные коэффициенты заболеваемости для экспонированного населения были бы такими же, как для контрольной.

$$E = 3.000 \times 0,005 + 1.000 \times 0,025 = 40$$

Наблюдаемое и ожидаемое числа случаев связаны с возрастным распределением в экспонированной группе. Наблюдаемое число случаев зависит от коэффициентов заболеваемости экспонированной совокупности, а ожидаемое число случаев – от этих коэффициентов контрольной совокупности. Отношение наблюдаемого к ожидаемому числу случаев – есть стандартизованное относительное сравнение коэффициентов

заболеваемости в экспонированной и неэкспонированной группах населения. Данный метод, учитывающий различия в возрастном распределении, называется косвенной стандартизацией.

Отношение наблюдаемого числа заболеваемости к ожидаемому выражается в процентах и называется *стандартизованным отношением заболеваемости* (*SIR – standardized incidence ratio*), а при анализе смертности – *стандартизованным отношением смертности* (*SMR – standardized mortality ratio*).

В нашем примере мы имеем:

$$SIR = \frac{60}{40} \times 100 = 150$$

Кроме того, наблюдаемое и ожидаемое число случаев используется для расчета *избыточного абсолютного риска* (*AER – absolute excess risk*), характеризующего число «лишних» событий:

$$AER = \frac{O - E}{PY}$$

где PY (person-years) — объем наблюдения, человеко-лет.

В нашем примере AER составляет

$$AER = \frac{60 - 40}{4.000} = 0,005$$

или 50 на 10.000 человеко-лет

Показатель частоты заболевания должен характеризоваться *доверительным интервалом* (*CI – confidence interval*), определяющим точность полученного значения. Например, 95% доверительный интервал представляет собой диапазон значений, выстроенный таким образом, что включение истинного значения показателя составляет 95%. Другими словами, 95% всех величин будут истинными, а 5% - ложными. Т.е., доверительный интервал дает информацию о точности исследования. Для расчета доверительного интервала необходима вероятностная модель, которая обычно опирается

на биномиальное распределение или на распределение Пуассона. При достаточно большом количестве наблюдений модели часто расчет осуществляется путем нормального распределения или распределения Гаусса с целью упрощения формул.

Средняя ошибка вычисляется по формуле:

$$m = \pm \sqrt{\frac{P1 \times P2}{n}} ,$$

где P1 – процент наличия признака, P2 – процент отсутствия признака, n – число случаев.

Пример. При использовании алкилирующих препаратов у 300 пациентов при лечении лимфомы Ходжкина у 20% возникли вторые опухоли (МДС/ОЛЛ). У 200 больных в схемы лечения эти препараты не входили, вторые неоплазии диагностированы в 12% случаев [2]. Необходимо оценить статистическое различие показателей.

Средняя ошибка при использовании алкилирующих агентов составляет:

$$m1 = \pm \sqrt{\frac{20 \times 80}{300}} \% = \pm \sqrt{5,33\%} = \pm 2,31\%$$

Средняя ошибка при отсутствии в схемах алкилирующих препаратов составляет:

$$m2 = \pm \sqrt{\frac{12 \times 88}{200}} \% = \pm \sqrt{5,28\%} = \pm 2,30\%$$

Для оценки значимости двух показателей используем критерий Стьюдента:

$$t = \frac{P1 - P2}{\sqrt{m1^2 + m2^2}} = \frac{20 - 12}{\sqrt{5,33 + 5,28}} = \frac{8}{\sqrt{10,61}} = 2,4$$

По таблице, используя n' и t, находим, что P < 0,05. Это доказывает, что различия статистически достоверны и применение в схемах терапии лимфомы Ходжкина алкилирующих агентов связано с большей частотой вторичных гемобластозов.

Используемые статистические формулы давно экстраполированы в компьютерные программы. Благодаря современным компьютерным технологиям даже неподготовленный пользователь может успешно решать вопросы анализа данных. Однако

расширение знаний о вычислении показателей статистической обработки позволяет повысить самообразование и избежать некорректного применения в решении поставленных задач медицинского исследования.

Список литературы

1. Мерабишвили В.М. Выживаемость онкологических больных. – СПб.: фирма Коста, 2006. – 440 с.
2. Bhatia S., Robinson L., Meadows A. High risk of second malignant neoplasms (SMN) continues with extended follow-up of childhood Hodgkin's disease (HD) cohort: Report from the Late Effects Study Group // Blood. - 2001. – V. 98. – P. 275a.
3. Marres T., Moll A.C., Imhof S.M. et al. Risk of second malignancies in survivors of retinoblastoma: more than 40 years of follow-up // JNCI. – 2008. – V. 100. – P. 1771-1779.
4. Sankila R., Garwicz S., Olsen J.H. et al. Risk of subsequent malignant neoplasms among 1641 Hodgkin's disease patients diagnosed in childhood and adolescence: a population-based cohort study in the five Nordic countries. Association of the Nordic Cancer Registries and the Nordic Society of Pediatric Hematology and Oncology // J. Clin. Oncol. – 1996. – V. 14. – P. 1442-1446.