

УДК 616.24 - 006

ЧАСТОТНЫЙ, КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ ЗАБОЛЕВАЕМОСТИ РАКА ЛЕГКОГО В КАРАГАНДИНСКОЙ ОБЛАСТИ**Шериева Т.М., Кабилдина Н.А., Раззаков К.К., Исмаилов С.Т.***КГП «Карагандинский государственный медицинский университет»,
Казахстан, Караганда, @-mail: lioncub-93@mail.ru*

Проведен частотный, корреляционный и регрессионный анализ заболеваемости рака легкого в Карагандинской области.

Ключевые слова: рак легкого, частотный, корреляционный, регрессионный анализ

FREQUENCY, CORRELATION AND REGRESSION ANALYSIS OF THE INCIDENCE OF LUNG CANCER IN KARAGANDA REGION**Sheriyeva T.M., Kabildina N.A., Razzakov K.K., Ismailov S.T.***MSE «Karaganda State Medical university», Kazakhstan, Karaganda, @-mail: lioncub-93@mail.ru*

It have been held frequency, correlation and regression analysis of the incidence of lung cancer in the Karaganda region.

Keywords: lung cancer, frequency, correlation, regression analysis

Введение

В настоящее время во всех странах мира самым распространенным из злокачественных новообразований является рак легкого (РЛ) и частота его растет с каждым годом. По данным экспертов МАИР рак легкого в 2010 году занял первое место среди всех онкологических заболеваний. Эта опухоль выявлена у более миллиона населения планеты, что составило 13,2% от всех злокачественных новообразований. На индустриально развитые страны мира приходится 54% новых случаев РЛ, который является основной причиной смертности от злокачественных опухолей [1,2,3,4]. В развивающихся странах с развитием урбанизации и индустрии прогнозируется рост заболеваемости и смертности от РЛ. Свыше 50% населения земного шара проживает в странах, где противораковая борьба является одной из приоритетных задач национального здравоохранения. В Казахстане указом Президента Н.А. Назарбаева 2002 год был объявлен «годом здоровья», а Постановлением Правительства от 30.12.2001 года за № 571 онкология впервые была внесена в разряд приоритетных направлений [5,6,7].

Цель - изучение заболеваемости рака легкого в Карагандинской области с проведением частотного, корреляционного и регрессионного анализа.

Материалы и методы исследования

Изучены статистические данные первично выявленных за 20 лет (1990-2010гг.) 1981 больных РЛ и соответственно вычисленные на основе этих материалов статистические показатели.

Для выполнения поставленных задач использованы материалы из уточненных сведений: годовых отчетов Карагандинского областного онкологического диспансера (КООД); материалов из 1981 медицинских карт стационарных больных и 1242 медицинских карт амбулаторных больных РЛ, пролеченных и наблюдавшихся в торакальном, 2 радиологическом, химиотерапевтическом и диспансерном отделениях КООД за этот период, статистических талонов, извещений, выписок из медицинских карт КООД и других ЛПУ, направлении на патолого-гистологическое исследование, протоколов на случаи выявления у больных запущенной формы РЛ, контрольных талонов к карте диспансерного наблюдения, статистических карт и данных разработанного нами скринингового тест - «Кодификатора по изучению заболеваемости рака легкого».

В указанный кодификатор проведена выкопировка сведений из проанализированной медицинской документации. Она составлена таким образом, что при ее заполнении и считывании сразу производилась кодировка всех признаков в виде цифровых значений (1- 20), при этом учитывались все значения для каждого конкретного случая. Кодификатор включает общую (паспортные данные) и специальную части, которые включают ряд признаков: анамнез жизни, перенесенные болезни, профессиональная патология, стаж основной и прежней работы, наличие производственных неблагоприятных факторов и вредных привычек, данные клинико-диагностического обследования больных. Рассматривались основной диагноз и его осложнения, степень тяжести, динамика и длительность заболевания, мероприятия по лечению и диспансеризации и т.д. В дальнейшем все сведения в зашифрованном виде заносились в специально созданную базу данных.

Также использованы дескриптивные, аналитические, социальные, статистические и математические методы исследования с применением программ «Ex-

cell», «Statistica5.2.», «Lung/cancer». Первичный анализ данных проведен в модуле «Основные статистики/таблицы с использованием подпрограммы «Таблицы частот». Применялись частотный, корреляционный и регрессионный анализ с применением методов вариационной статистики. Достоверность различия сравниваемых показателей определялась с учетом критерия Стьюдента $t = 1,97$ при $p < 0,05$.

Коэффициент корреляции для выборки значений определен по уравнению:

$r = \frac{\sum x1' * y1'}{\sqrt{\sum(x1')^2 * \sum(y1')^2}}$, где $x1'$ и $y1'$ - отклонения фактических значений переменных от средних (x и y), то есть $x1' = x1 - \bar{x}$ и $y1' = y1 - \bar{y}$.

Для проверки существенности при малом числе наблюдений применяли формулу:

$t = r \sqrt{n-2} / \sqrt{1-r^2}$, где t - критерий распределения Стьюдента (сопоставляется с табличным значением tp при $p-2$ степенях свободы для вероятности $a=0,95$). Также использован пошаговый регрессионный анализ по Pearson для оценки линейной зависимости между несколькими независимыми переменными ($X1...n$) и зависимой переменной (Y):

$\hat{y} = (a + b * \hat{x}) \pm tasy$; где \hat{y} - зависимая переменная; a - свободный член модели; b - угловой коэффициент для независимого признака; \hat{x} - независимая переменная; $\pm tasy$ - величина доверительных границ \hat{y} ; ta - табличное значение t - распределения Стьюдента, при $p-2$ степенях свободы; sy - выборочная оценка дисперсии случайных величин.

При проведении множественной регрессии вычислены коэффициенты множественной регрессии (r), детерминации (d^2), Фишера (F) и показатель статистической значимости (p). Применяли значимые параметры, вероятность которых равна или более 0,95 ($p < 0,05$). Множественная регрессия выражена в уравнении линейной регрессии: $\hat{y} = (a + b1 * \hat{x}1 + b2 * \hat{x}2 + \dots + bp * \hat{x}p)$.

Результаты исследования и их обсуждение

Первым этапом нашего исследования было проведение частотного анализа основных показателей из разработанного кодификатора. При этом использовались все качественные параметры, каждому из которых было присвоено одно из значений (по кодам). Основная цель частотного анализа - выявление частоты определенного признака в анализируемой выборке (в процентах). Для дальнейшей статистической обработки использованы все показатели, которые подверглись частотному анализу - численность заболевших РЛ в поло-возрастном аспекте, по месту жительства, количество больных по роду деятельности (по разным специальностям - основной и прежней), профессионального стажа, наличие различных вредных производственных факторов. Нами учитывались следующие виды вредностей: работа в условиях влажности, сырости, переохлаждения, перегревания, пыли, шума, выхлопных отработанных газов, сварочной аэрозоли, вредных испарений химических веществ, вибрации, излучения и др. Также нами выделены такие характеристики тру-

да, как работа в условиях: вынужденного положения (конвейерные цеха), стресса (диспетчеры, машинисты), эмоционального напряжения (служащие) и физического напряжения (рабочие, с умеренной и значительной физической нагрузкой). Отдельно проанализирована частота случаев профессиональной патологии, различных перенесенных заболеваний, фоновых заболеваний легких (ФЗЛ), вредные привычки, результаты обследования в КООД, случаи цитоморфологического подтверждения диагноза РЛ и другие.

Нами изучены материалы 1981 больных РЛ за период 1990-2010гг.

При обследовании в поликлинике КООЦ у больных выявлены следующие стадии заболевания: II-6%, III-67,7% и IV-26,3%. Как видно, большинство больных выявлено с III-IV стадией процесса - почти 94%. В 28,2% случаях при морфологических исследованиях выявлены метастазы и в 7,1% - опухолевый распад.

При этом анализ заболеваемости больных по половой принадлежности выявил, что из 1981 больных 87,1% составляли мужчины и 12,9% женщины. Соотношение численности заболевших мужчин и женщин 7:1. По месту жительства - городские жители составили 76,4%, сельские -23,6%, то есть 2/3 больных РЛ составили горожане.

Распределение больных по возрастным группам следующее: 50-59 лет - 44% пациентов; 60-69 лет - 28,8%; 40-49 лет-15,3%; 70-79 лет - 8,3% и другие - 3,6%. Как видно, среди обследованных отмечено преобладание лиц трудоспособного, социально активного возраста - 50-59 лет - 44%. По этническим группам больные распределились следующим образом: русские - 52,3%, казахи - 13,2%, украинцы - 11,6%, немцы - 9,4% и другие национальности -13,5%. Среди пациентов РЛ преобладали лица европейской национальности над коренными в 5,5 раз, их доля составила 73,3%.

При изучении социального статуса и профессионального анамнеза заболевших на момент обследования, выяснилось что, работающие составили 66,1%, неработающие - 33,9%. Из числа первых, рабочие составили - 34,8%, пенсионеры - 28,5%, водители - 12,8%, служащие - 9,7%, инвалиды - 5,4%, шахтеры - 3,8%, металлурги - 0,3% и другие - 2,6%, не работали - 2% и 0,1% - не указали профессию. Мы также учитывали изменения профессиональной деятельности пациентов, при этом профессии отмеча-

ли как основные (на момент обследования) и прежние (в анамнезе, так как, при смене профессии, раньше могли иметься вредные факторы на работе). Изучение профессионального анамнеза пациентов, имеющих прежние профессии, позволило выяснить следующее: рабочие составили - 11,9%, шахтеры - 11%, водители - 4,9%, служащие - 2,5%, металлурги - 1,3% и другие - 1,2%.

Среди работающих профессиональный стаж по основной специальности колебался от минимального (менее 5 лет указывали - 0,5% пациентов) до максимального (40 лет стажа и выше - 0,8%). Стаж по предыдущей специальности: менее 10 лет указали - 3,2%; менее 20 лет - 7,2%; 2,2% человек отрицали смену профессии.

В профессиональном анамнезе вредные производственные факторы (как одиночные, так и сочетанные) выявлены у 86,6% больных. В этой группе, в числе других вредностей чаще назывались: работа в условиях запыленности - 70,2%; переохлаждения, сквозняков - 68,3%, меньше всех - излучения при сварочных работах - 2,4% и сырости, влажности - 2,1%. В большинстве случаев обнаружены сочетанные вредности - 88,5%. При исследовании материалов, в признаки, показывающих напряженность труда, были включены некоторые характеристики. Из них, чаще встречались вынужденное положение при работе - 17,5%, эмоциональное напряжение - 27,5% и физический труд от умеренной - 66,4% до значительной степени тяжести - 4,9%. У шахтеров, металлургов, токарей-фрезеровщиков, работников транспорта наблюдались сочетанное действие этих факторов.

Данные анамнеза по поводу различных перенесенных заболеваний выглядят следующим образом: 88,3% человек подтвердили наличие сопутствующей патологии, 11,7% отрицали их (или не помнили). Профессиональные заболевания выявлены у 3,3% больных. При этом указывались профессиональный пылевой бронхит и силикоз - по 1%. Антракосиликоз назван в 0,6% случаях, вибрационная болезнь - в 0,2%. Антракоз и сочетанная профессиональная патология по 0,2% случая каждый.

В структуре сопутствующей патологии у пациентов, на которую приходится почти 90%, первое место занимает сочетанная патология - 63,5%, второе - заболевания сердечно-сосудистой системы

-8,5% пациентов, третьи - инфекционные болезни - 2,8%, далее - различные травмы, раны, ожоги - 2,2%. Патология со стороны эндокринной и урогенитальной систем была лишь у 0,4% и 0,3% больных соответственно. Из всей группы (1756) 3 пациента состояли на учете в КООЦ (с диагнозами рак нижней губы, кожи и молочной железы). 10 пациентов были прооперированы по поводу миом, кист яичника и язвы желудка. Из всей группы на диспансерном учете у терапевта с соматическими заболеваниями находились лишь 18,6% больных.

Нами также была проанализирована частота имеющихся фоновых заболеваний легких в анамнезе: их подтвердили 65,3% больных; отрицали (или не помнили) - 34,7%. Обращает на себя внимание высокий процент пневмонии - 38,1% и хронического бронхита - 33,6% соответственно. Реже встречались травмы, ранения и оперативные вмешательства на легких - 0,1%. Сочетанные заболевания легких (бронхит и пневмония, пневмония и туберкулез, и другие) отмечали - 33,7% пациентов. Из всего количества больных с фоновыми заболеваниями легких на Д-учете состояли только 8,6% больных.

Отягощение наследственности было выявлено в 0,4% случаях наблюдений. Из них в 3 случаях отмечен РЛ у родственников первой и второй линии, в 6 случаях - генетическая предрасположенность к различным онкологическим заболеваниям.

При изучении вредных привычек, в анамнезе чаще всего отмечался факт курения, подтвердили его 67,6% больных. При этом, стаж курения называли разный: до 10 лет - 1,3% пациентов, до 20 л - 6,1%, до 30 л - 13,8%, до 40 л - 26,5%, до 50 л - 13,7%, до 60 л - 3,1%, свыше 60 л - 0,3% больных. 9,1% курильщиков стаж курения не указали. Большинство из курящих имели стаж курения до 40 лет - 26,5%. Также варьировал режим курения: 10 сигарет в день выкуривали 0,9% больных, менее 20 - 9,9%, 25 - 0,6%, менее 40 сигарет - 4,3%, до 60 - 0,3% и до 80 сигарет в день всего лишь 0,2%. Из всей группы курящих, большинство число выкуриваемых сигарет не уточнили - 51,3%.

Из других вредных привычек, закладывание насвая признали 2,7% пациентов. Алкоголь употребляли 24,3% больных. Количество выпивки варьировало

в пределах умеренного - 9,6%, среднего - 12,5% и значительного - 2,2%.

Важным моментом онкоэпидемиологического исследования является проведение корреляционного и регрессионного анализа, который является наиболее доступным методом выбора при оценке приоритетных показателей. Основная цель корреляционного анализа - поиск внутри каждой подсистемы изучаемых явлений и признаков, которые между собой тесно взаимосвязаны и взаимозависимы. Коэффициент корреляции для выборки значений вычисляли по уравнению: $r = \frac{\sum x_1' \cdot y_1'}{\sqrt{\sum (x_1')^2 \cdot \sum (y_1')^2}}$. Сила связи зависит от величины коэффициента корреляции Пирсона, где 0 - связь отсутствует, 0 - 0,3 - слабая, 0,3 - 0,7 - умеренная связь, 0,7-1,0 - сильная, 1,0 - полная связь (функциональная). Чем выше значение r , тем теснее связь между переменными и с тем большим основанием данная взаимосвязь может быть использована для последующего регрессионного анализа. Пошаговый регрессионный анализ проводили по Пирсону: $\hat{y} = (a + b \cdot x) \pm t_{asy}$. Множественная регрессия определена по уравнению линейной регрессии: $\hat{y} = (a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p)$.

В приведенной модели коэффициенты регрессии (b) представляли собой вклады каждой независимой переменной в зависимую переменную. Их значения не сопоставимы и зависят от единиц измерения, диапазонов измерения соответствующих переменных.

Нами выделены факторы риска возникновения РЛ на основе указанных уравнений. Также в работе мы определяли оценку частоты связи между отдельными признаками при РЛ для проведения прогнозирования тяжести течения и исхода заболевания у отдельно взятого больного. Отбор значимых признаков для занесения в математическую модель рассчитан по коэффициентам регрессии с включением F при $p < 0,05$.

Статистически выявлено, что значимыми факторами риска возникновения РЛ являются: производственные вредности, (где коэффициент корреляции r составил 0,99), экологически неблагоприятный район места жительства ($r=0,65$) и фоновые заболевания легких ($r=0,49$). Также проведенный регрессионный анализ показал, что исход РЛ достоверно зависит от соматического состояния больного, стадии опухолевого процесса, наличия осложнений, метастазов и перенесенных заболеваний в анамнезе ($p < 0,001$), а тяжесть течения РЛ - от стадии

процесса, наличия осложнений и сопутствующей патологии ($p < 0,001$).

Выводы

Таким образом, проведенное исследование выявило, что большая часть больных РЛ представлена лицами трудоспособного возраста - 20-59 лет (62,6%). Изучение этнического состава пациентов показало выраженное преобладание лиц европейской национальности - 73,3% над лицами коренной национальности (13,2%) в 5,5 раз. На момент обследования, из числа работающих, 56,5% составили лица физического труда. У 86,6% пациентов были выявлены профессиональные вредности, 67,6% злоупотребляли курением. Также в анамнезе больных выявлены различные заболевания: профессиональная патология в 3,3%, различные сопутствующие заболевания в 88,3% и фоновые заболевания легких в 65,3% случаях наблюдений. Расчетные данные показали, что частота РЛ у жителей городской и сельской местностей в области не совпадают. У сельчан заболеваемость РЛ ниже в 3,2 раза, чем у городских жителей. Разница в заболеваемости РЛ жителей городов и сельских районов может быть обусловлена комплексным воздействием природных и антропогенных факторов, выраженных в разной степени на отдельных территориях. При проведении корреляционного анализа взаимозависимости между частотой РЛ и местом жительства выявлена связь умеренной силы с положительным знаком корреляции $r = 0,65$).

Также выявлено, что значимыми факторами риска возникновения РЛ могут быть производственные вредности ($r=0,99$) и фоновые заболевания легких ($r=0,49$). Регрессионный анализ показал, что исход РЛ достоверно зависит от соматического состояния больного, стадии опухолевого процесса, наличия осложнений, метастазов и перенесенных заболеваний в анамнезе ($p < 0,001$), а тяжесть течения РЛ - от стадии процесса, наличия осложнений и сопутствующей патологии ($p < 0,001$).

Список литературы

1. Заридзе Д. Г. Эпидемиология и профилактика рака // Вопросы онкологии. - 2001. - №9. - С. 6-14.
2. Нидюлин В.А., Эрдниева Б.В. Об эпидемиологии рака легких // Медицинский вестник Башкортостана: обзорная статья. - Башкирский Государственный Медицинский Университет, 2009. - т. 4. - № 1. - с. 66-71.

3. Шаназаров Н.А. Рак легкого: учеб.-метод. пособие / под ред. А.В.Важенина. - Челябинск: УГМАДО, 2002.
4. Duan J.C., An T.T., Wu M.N., Yang L. et al. Correlation between the efficacy of epidermal growth factor receptor tyrosine kinase inhibitors and EGFR mutations in advanced squamous cell lung cancer. *Zhonghua Jie He He Hu Xi Za Zhi*. 2012 May.
5. Hanada S., Nishiyama N., Mizuguchi S., Yamano S. et al. Clinicopathological significance of combined analysis of cytokeratin 19 expression and preoperative serum CYFRA21-1 levels in human lung squamous cell carcinoma. *Osaka City Med. J.* 2013 Jun.
6. Li S.J., Wang W.Y., Li B., Chen B. et al. Expression of NDRG2 in human lung cancer and its correlation with prognosis. *Med. Oncol.* 2013 Mar.
7. Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta – analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. *Am. J. Epidemiol.* 2012 Jan 1.