

Выводы

Разработан индивидуальный интегральный показатель здоровья, основанный на динамической модели исследуемого объекта, заключающийся в вычислении евклидова расстояния исследуемого объекта от эталонного и отличающийся шкалированием исходных данных с помощью функции принадлежности Харрингтона. Показатель служит для объективной оценки состояния здоровья пациентов с заболеваниями щитовидной железы и доказательного определения эффективности лечения.

СПИСОК ЛИТЕРАТУРЫ

1. Кормление и питание грудных детей и детей раннего возраста / Методические рекомендации для Европейского региона ВОЗ с особым акцентом на республики бывшего Советского Союза. — Всемирная организация здравоохранения [Электронный ресурс]. — 2003. — Режим доступа: <http://www.who.int/res-ea/rch/ru/>. — 22.04.2009.
2. Мурашко Е.В., Осокина Г.Г. Современные аспекты синдрома вегето-сосудистой дистонии у детей и подростков: клиника, диагностика. Подходы к терапии // Вестник аритмологии. — 2000. — № 18. — С. 49–50.
3. Щеплягина Л.А. Медико-социальные последствия роста напряженности зубной эндемии для детей и подростков / Тирод Россия. Сб. лекций. — М.: Медицина, 1997. — 142 с.
4. Пузырев В.П. Генетика мультифакториальных заболеваний: между прошлым и будущим // Медицинская генетика. — 2003. — Т. 2. — № 12. — С. 498–508.
5. Бубнов Е.А., Скороходов Д.И. Шкалирование входной информации в корабельных системах информационной поддержки [Электронный ресурс]. — 2000. — Режим доступа: <http://grinda.info/control/skalir/skalir.htm>. — 22.04.2009.
6. Дегтева О.А. Инвестирование в высшее образование как фактор развития социально-экономических систем: Автореф. дис. ... канд. экон. наук. — М., 2007. — 25 с.
7. Зазнобина Н.И. Оценка экологической обстановки в крупном промышленном центре по степени антропогенной нагрузки с помощью обобщенной функции желательности // Вестник Нижегородского университета им. Н.И. Лобачевского. — 2007. — № 2. — С. 115–118.
8. Казаков В.Ф., Серебряков В.Г. Бальнеотерапия ишемической болезни сердца. — М.: Медицина, 2004. — 256 с.
9. Савченко В.М. Унифицированная стандартизация значений показателей исследования в клинической пульмонологии // Украинский пульмонологический журнал. — 2002. — № 3. — С. 22–25.

Поступила 22.04.2009 г.

УДК 004.9:519.252

ИНФОРМАЦИОННАЯ СИСТЕМА ОЦЕНКИ ФАКТОРОВ РИСКА ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ

О.А. Ананина, Л.Ф. Писарева, В.А. Фокин*

НИИ онкологии ТНЦ СО РАМН, г. Томск

*ГОУ ВПО «Сибирский государственный медицинский университет», г. Томск

E-mail: epidem@oncology.tomsk.ru

Разработана информационная система для оценки факторов риска возникновения онкологических заболеваний на основе анализа результатов анкетирования населения. Методом мета-анализа выявлены обобщенные факторы риска и построена прогностическая модель оценки индивидуального риска возникновения рака молочной железы.

Ключевые слова:

Информационные системы, факторы риска, злокачественные новообразования.

Введение

Компьютерные технологии и разрабатываемые на их основе программно-технические средства и информационные системы являются универсальным инструментом, использующимся в биомедицинских исследованиях и практическом здравоохранении для формирования структурированных массивов данных, их анализа и принятия решений на основе алгоритмов обработки в соответствии с

В результате выделения информативных показателей число анализируемых лабораторных данных сокращено в два раза, что позволило существенно снизить расходы на обследование пациентов.

Рассчитан обобщенный индекс здоровья для двух групп детей с различными формами тиреопатологии, подтвердивший правильность выделения информативных показателей и позволивший доказательно оценить эффективность проведенного курса реабилитации.

целями пользователя [1–3]. В практику здравоохранения прочно вошло понятие «скрининг» — массовое периодическое обследование практически здорового населения с целью выявления ранних стадий опухолей, лечение которых дает хороший прогноз. При этом важное место отводится как диагностическому (опухолевый процесс распознается на стадии прогрессии опухоли), так и профилактическому скринингу, при котором формируются

группы онкологического риска, а диагностические методы используют для дальнейшего обследования отобранного контингента на базе специализированных медицинских учреждений.

В основе профилактического скрининга лежит анализ результатов анкетирования отдельных групп населения по выявлению этиопатогенетических факторов, влияющих на возникновение и частоту распространения онкологических заболеваний. Анкетирование как метод выявления факторов риска обладает рядом очевидных преимуществ, таких как широта охвата, простота в реализации, малозатратность, отсутствие воздействия на здоровье, гибкость учета специфики изучаемых популяционных групп населения.

В данной работе описана информационная система обобщения и анализа результатов анкетирования по выявлению факторов риска возникновения онкозаболеваний и формированию на их основе групп повышенного риска.

Оценка значимости факторов риска

Анализ первичных материалов, накопленных в лаборатории эпидемиологии НИИ онкологии ТНЦ СО РАМН, по выявлению факторов риска онкологических заболеваний показал, что структура анкет, при проведении исследований в рамках изучения одной и той же проблемы, может быть различной. В частности, анкетирование проводимое в отдельных регионах, в разные временные периоды, на различных группах населения различается количеством и типом вопросов в анкете, градацией ответов на них и т. п. Такая разнородность анкет, позволяющая гибко учитывать те или иные особенности влияния факторов риска, в то же время может приводить к трудностям их совместного исследования и обобщения.

Необходимость совместного анализа результатов анкетирования, полученных в различных исследованиях, обусловлена двумя причинами. Во-первых, это способствует повышению статистической значимости получаемых оценок, а во-вторых, позволяет выявить факторы риска, общие для всей совокупности анализируемых данных. Для обобщенного анализа и выявления связи различных факторов с риском возникновением онкозаболеваний нами был использован метод мета-анализа [4, 5]. Международная эпидемиологическая ассоциация определяет мета-анализ как методику «объединения результатов различных исследований ... складывающуюся из качественного компонента (например, использование таких заранее определенных критериев включения в анализ, как полнота данных, отсутствие явных недостатков в организации исследования и т. д.) и количественного компонента (статистическая обработка имеющихся данных)» [6].

Количественно величина риска возникновения заболевания при наличии или отсутствии изучаемого фактора оценивается отношением шансов [7]:

$$OR_i = \frac{a_i \cdot d_i}{b_i \cdot c_i},$$

где OR_i — отношение шансов, рассчитанное по результатам i -го исследования; a_i — число случаев наличия фактора риска группе больных; b_i — число случаев отсутствия фактора риска в группе больных; c_i — число случаев наличия фактора риска в группе здоровых; d_i — число случаев отсутствия фактора риска в группе здоровых. Отношение шансов может принимать значения в диапазоне от 0 до ∞ , причем $OR_i=1$ соответствует случаю отсутствия связи между исследуемым фактором и риском возникновения заболевания. Такая несимметричность распределения отношения шансов приводит к тому, что непосредственно оценить границы доверительного интервала для него достаточно сложно. Поэтому для получения оценок используют логарифмическое преобразование отношения шансов, которое приводит к распределению $\ln(OR_i)$, близкому к нормальному. Границы $(1-\alpha)\%$ -го доверительного интервала CI_i для полученной оценки OR_i определяются следующим образом:

$$CI_i = \exp(\ln(OR_i) \pm t_\alpha \sqrt{V_i}),$$

где t_α — коэффициент Стьюдента для уровня статистической значимости α , V_i — дисперсия логарифма отношения шансов.

В предположении, что основным источником вариативности оценок, получаемых в различных исследованиях, является случайная составляющая, суммарная оценка величины риска возникновения заболевания может быть рассчитана как средневзвешенная величина оценок для всех анализируемых исследований:

$$OR = \sum_{i=1}^m (OR_i \cdot w_i) / \sum_{i=1}^m w_i,$$

а соответствующий доверительный интервал будет равен

$$CI = \exp \left(\ln(OR) \pm t_\alpha \sqrt{\sum_{i=1}^m (1/w_i)} \right),$$

где $w_i=1/V_i$ вес i -го исследования, m — количество исследований.

Вывод о связи изучаемого фактора риска с возникновением заболевания делается, если доверительный интервал для OR не включает 1. Все выявленные по совокупности исследований статистически значимые факторы риска используются в дальнейшем для построения прогностической модели оценки индивидуального риска и формирования групп повышенного риска возникновения онкозаболеваний.

Прогностическая модель оценки факторов риска

Прогностическая модель строится на основе байесовского классификатора для переменных, у

которых выявлена статистически значимая связь с риском возникновения онкозаболевания. Значениями переменных, описывающих выявленные факторы риска, являются варианты ответа на вопрос анкеты (градации фактора). Каждая градация фактора x_i оценивается величиной диагностического коэффициента:

$$DK_{i,j} = 10^q \lg \left(\frac{p(x_{i,j} / A_2)}{p(x_{i,j} / A_1)} \right),$$

где $p(x_{i,j}/A_1)$ и $p(x_{i,j}/A_2)$ – вероятности значений j -й градации фактора x_i при состояниях здоров (A_1) или болен (A_2), соответственно; 10^q – числовой множитель.

Накопление диагностической информации для оценки риска возникновения заболевания происходит путем суммирования диагностических коэффициентов $DK_{i,j}$ по всем значимым факторам риска:

$$DK = \sum_{i=1}^n DK_{i,j},$$

где n – количество факторов риска.

При превышении диагностическим коэффициентом DK заданного порогового значения A , обследуемого следует отнести к группе риска возникновения онкозаболеваний и, наоборот, при DK меньшем пороговой величины B к группе отсутствия риска. Случай $B < DK < A$ соответствует группе неопределенности. Диагностические пороги принятия решения задаются исходя из уровней ошибок I (α) и II (β) рода [8]:

$$A = 10^q \lg \left(\frac{1-\alpha}{\beta} \right), \quad B = 10^q \lg \left(\frac{\alpha}{1-\beta} \right).$$

В частности, при уровне ошибок первого рода (гипердиагностика: обследуемый здоров, но по результатам анкетирования отнесен к группе риска) 10 %, уровне ошибок второго рода (гиподиагностика: обследуемый относится к группе больных, но по результатам анкетирования отнесен к группе отсутствия риска) 5 % и $q=1$, диагностические пороги равны $A=+12,5$ и $B=-9,8$.

Структура информационной системы

Для анализа результатов анкетирования и оценки факторов риска возникновения онкозаболеваний на основе обобщения данных различных исследований, была разработана информационная система, структурная схема которой представлена на рис. 1.

В информационной системе предусмотрены 3 базы данных: описаний вопросов и ответов, электронных бланков анкет, результатов анкетирования. Поскольку каждому исследованию может соответствовать своя анкета с определенной структурой вопросов и ответов, то в единую базу описаний вопросов и ответов заносятся все различные варианты вопросов и соответствующих им ответов, со-

держащихся в анкетах. В дальнейшем из нее генерируются электронные бланки анкет, которые используются для непосредственного заполнения базы результатов анкетирования данными, например, с бумажных носителей. Ведение общей базы вопросов значительно оптимизирует работу по разработке бланков анкет для новых, по своим задачам, исследований.

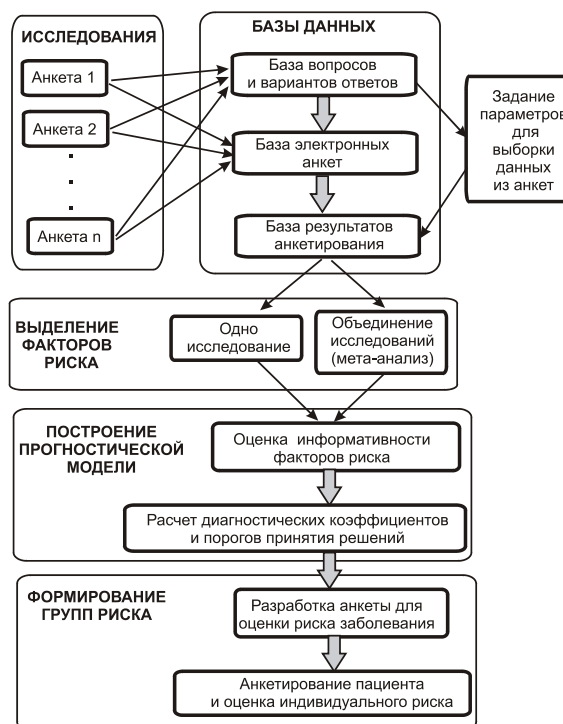


Рис. 1. Структура информационной системы анализа факторов риска

Для выявления и оценки факторов риска формируются выборки первичных данных из анкет, вопросы в которых удовлетворяют заданным условиям. В зависимости от поставленной цели анализа, выборки могут быть представлены вопросами и вариантами ответов полученных в отдельном исследовании, либо совокупностью результатов нескольких исследований. Выявленные статистически значимые факторы риска используются для построения прогностической модели оценки индивидуального риска и формирования групп повышенного риска возникновения онкозаболеваний.

Информационная система реализована в виде программного комплекса состоящего из двух независимых модулей «ПИФАРО» [9] и «РИСК», первый из которых предназначен для выявления факторов риска, а второй для проведения анкетирования населения на основе построенной прогностической модели. Такое разделение на модули позволяет разделить научно-исследовательскую деятельность по анализу и обобщению результатов анкетирования и скрининга на территориях по выявлению и формированию групп риска.

Прогностическая модель для формирования групп риска РМЖ

Разработанная система была использована для анализа анкет накопленных в лаборатории эпидемиологии НИИ онкологии ТНЦ СО РАМН (руководитель проф. Л.Ф. Писарева) по исследованию факторов риска развития рака молочной железы (РМЖ) среди женского населения, проживающего на различных территориях региона Сибири и Дальнего Востока. Всего было проанализировано 684 анкеты женщин с морфологически верифицированным диагнозом РМЖ и 1209 анкет контрольной группы без онкологических заболеваний, распределение которых по отдельным территориям представлено в табл. 1.

Таблица 1. Количество анкет по территориям региона

Территория	Больные РМЖ		Здоровые		Всего
	Абс.	%	Абс.	%	
Приморский край	123	18,0	366	30,3	489
Иркутская область	125	18,3	132	10,9	257
Красноярский край	122	17,8	247	20,4	369
Республика Хакасия	71	10,4	82	6,8	153
Республика Тыва	127	18,5	252	20,8	379
Республика Алтай	116	17,0	130	10,8	246
Всего	684	100,0	1209	100,0	1893

В зависимости от территориальной специфики (степень урбанизации, интенсивность миграционных потоков и т. п.) анкеты, используемые в различных исследованиях, различались между собой структурой, количеством (от 95 до 120), типом вопросов и градациями ответов на них. Формулировка вопросов анкет была направлена на выявление социально-гигиенических (экзогенных) и медико-биологических (эндогенных) факторов риска. К первой группе относятся факторы, характеризующие демографические данные, миграционную подвижность, уровень образования и дохода, профессиональную принадлежность, вредные привычки, воздействие радиации, особенности питания, стрессовые ситуации, а ко второй – возраст, национальность, конституциональные особенности, функционирование репродуктивной системы, наличие сопутствующих заболеваний, наследственная предрасположенность и т. п.

В результате проведенного анализа была построена прогностическая модель, включающая 26 вопросов анкет, ответы на которые были статистически значимо связаны с наличием или отсутствием факторов риска возникновения РМЖ. Оценка качества прогностической модели проводилась на экзаменационной выборке состоящей из анкет 118 больных и 178 здоровых женщин, не включенных в исходную выборку. Диагностическая точность модели составила 86,3 %. На рис. 2 приведено распределение диагностических коэффициентов, рассчитанных по построенной прогностической модели в группах больных РМЖ и здоровых женщин. Пунктирными линиями отмечена область неприятия решения.

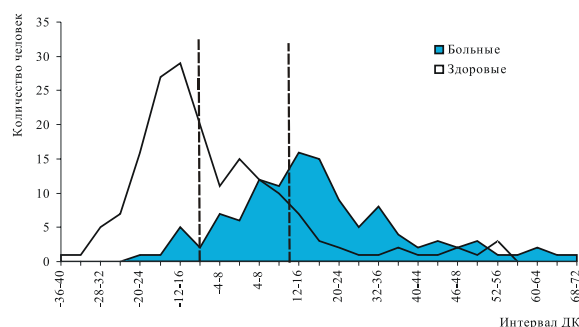


Рис. 2. Распределение значений в группах больных РМЖ и здоровых женщин

На последнем этапе верификации разработанной информационной системы протестирована группа из 101 женщины, пришедшей на специализированный маммологический прием в диагностический центр ГУ НИИ онкологии. После тестирования проводилось УЗИ-исследование или рентгенологическое маммографирование, на основании которого врачом-маммологом устанавливался клинический диагноз (табл. 2).

Таблица 2. Сопоставление диагноза врача и тестирования по прогностической модели

Группы	Диагноз врача	Тестирование
Здоровые	23 (22,7 %)	17 (16,8 %)
Больные	78 (77,3 %)	62 (61,4 %)
Группа неопределенности. В том числе:		22 (21,8 %)
гипердиагностика		6 (5,9 %)
гиподиагностика		16 (15,9 %)

Видно, что формирование групп риска, основанное на использовании построенной прогностической модели, является достаточно эффективным. В частности, все женщины из обследуемой группы с подтвержденным клиническим диагнозом РМЖ были отнесены по результатам тестирования либо к группе больных, либо к группе неопределенности постановки диагноза, включение в которую предполагает рекомендацию на проведение дополнительного специализированного обследования.

Выводы

1. Разработана информационная система для оценки факторов риска онкологических заболеваний, включающая модули выявления факторов риска и формирования групп повышенного риска на основе анкетирования, что позволяет эффективно организовывать хранение и накопление данных анкетирования населения, проводить анализ по выявлению факторов риска развития онкозаболеваний.
2. Проведение анкетного тестирования с использованием прогностической модели оценки индивидуального риска может быть рекомендовано как первый этап популяционного скрининга населения для формирования групп повышенного риска развития рака молочной железы.

СПИСОК ЛИТЕРАТУРЫ

1. Генкин А.А. Новая информационная технология анализа медицинских данных (программный комплекс ОМИС). — СПб.: Политехника, 1999. — 191 с.
2. Дюк В., Эммануэль В. Информационные технологии в медико-биологических исследованиях. — СПб.: Питер, 2003. — 528 с.
3. Богомолов А.В., Гридин Л.А., Кукушкин Ю.А., Ушаков И.Б. Диагностика состояния человека: математические подходы. — М.: Медицина, 2003. — 464 с.
4. L'Abbe K.A., Detsky A.S., O'Rourke K. Meta-analysis in clinical research // Ann. Intern. Med. — 1987. — V. 107. — P. 224–233.
5. Кукушкин Ю.А., Бухтияров И.В., Богомолов А.В. Обобщение результатов независимых экспериментальных исследований методом мета-анализа // Информационные технологии. — 2001. — № 6. — С. 48–54.
6. Last J.M. A dictionary of epidemiology / Ed. J.M. Last. — N.Y.: Oxford University Press, 1988. — 141 p.
7. Плавинский С.Л. Биостатистика: Планирование, обработка и представление результатов биомедицинских исследований при помощи системы SAS. — СПб.: Издательский дом СПбМАПО, 2005. — 560 с.
8. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. — Л.: Медицина, 1978. — 296 с.
9. Свид. № 2007613323 РФ. Программа для ЭВМ «Исследование факторов риска в онкологии ПИФАРО» / О.А. Ананина, Л.Ф. Писарева, В.А. Фокин. Заявка № 20076612390; Заявлено 13.06.2007; Опубл. 08.08.2007.

Поступила 20.04.2009 г.

УДК 519.256

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ ПОДХОД К ОПИСАНИЮ БИМЕДИЦИНСКИХ ДАННЫХ

А.В. Дубровин, В.А. Фокин

ГОУ ВПО «Сибирский государственный медицинский университет», г. Томск
E-mail: azygos.sky@gmail.com

Предлагается подход к описанию медико-биологических данных на основе объектно-ориентированных технологий. Проектирование информационных систем для хранения и анализа биомедицинских данных с его использованием позволяет сохранять концептуальную взаимосвязь между результатами отдельных исследований.

Ключевые слова:

Базы данных, объектно-ориентированные технологии, медико-биологические данные.

Массивы медико-биологических данных, в совокупности характеризующие состояние биосистемы, представляют собой разнообразные и разнообразные по источникам, методам получения и т. п. результаты клинических испытаний или биологического эксперимента, мониторинга или рутинных действий клинициста. Это создает определенные трудности при проектировании баз данных для интеграции результатов исследований в единую систему, позволяющую проводить их многогранный, основанный на компьютерных методах, анализ [1–3]. В то же время биомедицинские данные всегда имеют естественную концептуальную основу для систематизации и интеграции: они относятся к одному и тому же объекту — человеческому организму.

Объектно-ориентированные технологии, как средство решения проблем структуризации биомедицинских данных, является эффективным способом организации научных исследований в целом [4, 5]. В медицине и биологии объектно-ориентированный подход (ООП) находит все более широкое применение, в частности при проектировании экспертных систем [6, 7], описании и классификации биологических объектов [8] и т. п. В данной работе предлагается концепция структурированного описания результатов медико-биологических исследований с применением ООП.

В основе концепции ООП лежит представление о том, что предметную область необходимо проектировать как совокупность взаимосвязанных друг с другом объектов, рассматривая каждый объект как экземпляр определенного класса. Класс является описываемой на языке терминологии исходного кода моделью ещё не существующей сущности, так называемого объекта, а объект — экземпляр, реализация определенного класса. Причем объект — это пара: идентификатор и значение. Классы обладают свойством наследования, т. е. возможностью иерархического порождения одного класса другим с сохранением всех свойств класса-предка, и добавления необходимых новых свойств и методов.

Планирование и проведение медико-биологического исследования представляет собой алгоритм действий, направленный на изучение свойств объекта исследования. Его формальное описание заключается в установлении соответствий между свойством объекта исследования (измеряемый параметр) и его значением. Измерение любого параметра характеризуется условиями, которые включают в себя всю необходимую дополнительную информацию, позволяющую однозначно идентифицировать свойства показателей и условия проведения измерений и т. п. Это описание назовем **Бланком** исследования.