

УДК 002.53:004.65; 613.6.02

МЕТОДИКА ПОСТРОЕНИЯ СОПРЯЖЕННОЙ БАЗЫ ДАННЫХ ПОКАЗАТЕЛЕЙ ЗДОРОВЬЯ И СОЦИАЛЬНЫХ ФАКТОРОВ РИСКА РАЗВИТИЯ СОМАТИЧЕСКИХ НЕИНФЕКЦИОННЫХ ЗАБОЛЕВАНИЙ У РАБОТНИКОВ ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ

Кириянов Д.А., Лебедева-Несевря Н.А.

*ФГУН «Федеральный научный центр медико-профилактических технологий управления рисками здоровью населения» Роспотребнадзора, Пермь
e-mail: kda@fcrisk.ru, e-mail: natnes@fcrisk.ru*

Разработаны методические подходы к формированию сопряженных баз данных показателей здоровья и социальных факторов риска развития соматических неинфекционных заболеваний у работников промышленных предприятий на основе результатов эмпирических эпидемиологических, медико-биологических и социологических исследований. Методика может быть использована для выявления и описания закономерностей влияния факторов социального порядка на ухудшение состояния здоровья представителей различных социально-профессиональных групп.

Ключевые слова: сопряженные базы данных, риск здоровью

Основной вклад в потери трудового потенциала вносят ухудшение состояния здоровья населения трудоспособного возраста и преждевременная смертность в данной социальной группе. Среди всех факторов, детерминирующих здоровье работающего населения, при условии их взаимосвязанности и взаимозависимости, доминирующими, определяющими действие остальных являются факторы социального и социально-экономического порядка, что актуализирует, в частности, задачу установления закономерностей их влияния на здоровье работников промышленных предприятий на популяционном и индивидуальном уровнях. Решение данной задачи позволит подойти к разработке рациональной системы охраны здоровья трудящихся, обоснованию эффективных методов первичной и вторичной профилактики соматических неинфекционных заболеваний.

Исследование воздействия социально-экономических факторов риска на развитие соматических неинфекционных заболеваний у работников промышленных предприятий связано с анализом разнородной информации, которая накапливается и хра-

нится в различных организациях и ведомствах. Причем, каждый набор данных отличается своей оригинальной структурой, системой кодировок и индексов. В этой связи, при проведении углубленных исследований, использующих различные информационные массивы, важную роль приобретает процедура сопряжения данных.

При установлении закономерностей влияния социальных факторов риска на развитие соматических неинфекционных заболеваний у работников промышленных предприятий используется информация, характеризующая уровень заболеваемости, производственные вредности, результаты медицинских осмотров (периодических и дополнительных), социально-экономический статус и образ жизни работающих. В качестве источников данных служат:

- реестры пролеченных случаев заболеваний, хранящиеся в базе данных системы обязательного медицинского страхования (ОМС);
- результаты углубленных медико-биологических и социологических исследований, обследований, проводимых специализированной организацией;

- результаты периодических медицинских осмотров, хранящиеся в амбулаторных картах работников;

- документация по аттестации рабочих мест, проводимых службами Роспотребнадзора.

В указанных источниках информации данные организованы по своим установленным внутренним правилам, в своем формате и на разных носителях. Например, в большинстве ТФОМС разработаны и используются реляционные базы данных, содержащие широкий спектр персонифицированных данных на сервере, а документация по аттестации рабочих мест зачастую хранится в виде журналов. Это диктует необходимость разработки определенных форматов данных, позволяющих формировать информационные таблицы из различных источников.

Особо следует отметить, что большая часть информации содержит персональные данные, поэтому при их передаче и последующей обработке необходимо учитывать требования Федерального закона Российской Федерации от 27 июля 2006 г. № 152-ФЗ «О персональных данных» [3]. Согласно закону обработка персональных данных осуществляется для статистических или иных научных целей при условии обязательного обезличивания персональных данных. Поэтому, при формировании сопряженных баз данных не допускается наличие сведений, позволяющих идентифицировать работника.

Вместе с тем, при согласовании данных из разных источников желательно использовать универсальные идентификаторы, определяющие однозначные ссылки на разные наборы информации. Поскольку в данной работе основные сведения носят медицинский характер, в качестве такого универсального идентификатора предпочтительно использовать серию и номер страхового медицинского полиса. Использование номеров полиса позволит однозначно привязывать данные из ФОМС и амбулаторных карт. Для привязки остальных информационных массивов используются дополнительные идентификаторы.

В силу того, что при решении задач оценки связи состояния здоровья работаю-

щих с факторами риска объем информации довольно большой как в плане количества объектов исследования, так и анализируемых параметров, то формирование сопряженной базы данных необходимо проводить на основе принципов построения реляционных баз данных.

Следует отметить, что требование проектирования реляционных баз данных не является обязательным. В целом по форме представления информации базы данных разделяют на плоские, иерархические, сетевые, реляционные, многомерные, объектно-ориентированные. В плоских базах данных вся информация располагается в единственной таблице, каждая запись в которой содержит идентификатор конкретного объекта. Реляционные базы данных состоят из нескольких таблиц, связь между которыми устанавливается с помощью совпадающих значений одноименных полей. Использование того или иного типа баз данных определяется как объемом самих данных, так и уровнем потребности в конструировании серьезных структур при решении частных (зачастую разовых) задач. В настоящее время существует несколько различных моделей представления данных, которые пока не получили такого широкого распространения среди разработчиков и пользователей, как реляционная модель. Однако при разработке систем управления базами данных, реляционная модель фактически является стандартом.

В принципе, всю информацию, содержащуюся в базе, можно разместить в одной таблице, но такая структура данных является неэффективной, поскольку в этой таблице будет достаточно много повторяющихся данных. Такая организация данных может привести к следующим проблемам:

- наличие повторяющихся данных приведет к неоправданному увеличению размера файла базы данных. Кроме нерационального использования дискового пространства, это также вызовет заметное замедление работы приложения;

- ввод пользователем большого количества повторяющейся информации неизбежно приведет к возникновению ошибок;

- изменение одного из часто используемых параметров потребует значительных усилий по корректировке каждой записи, содержащей эти данные.

Процесс уменьшения избыточности информации в базе данных посредством разделения ее на несколько связанных друг с другом таблиц называется нормализацией данных. Существует шесть уровней нормализации базы данных, которые получили название нормальных форм. Нормальные формы различаются между собой по сложности структур хранения данных от простых (первая нормальная форма) до более сложных (третья и последующие нормальные формы). Вместе с ростом сложности структур хранения данных уменьшается избыточность и противоречивость информации, но, вместе с этим, также уменьшается «прозрачность» данных и увеличивается время обработки информации. Т.е. при формировании сопряженной базы данных необходимо искать оптимум в каждой практической реализации.

Для построения сопряженных данных существует несколько способов практической реализации. Наиболее правильным, универсальным и оптимальным является использование специальных систем управления реляционными базами данных (СУРБД). Наиболее популярными СУРБД являются Oracle, MS SQL Server, MS Access, DB2, MySQL [2]. Однако, при всей своей правильности, этот путь не всегда является рациональным, поскольку требует наличия квалифицированных специалистов в области IT-технологий и баз данных. Кроме того, доступ к данным осуществляется с помощью специально разработанных программных средств, а сами данные не являются «прозрачными» для исследователей. Использование СУРБД оправдано при больших объемах хранения информации и организации систематического пополнения. В этом случае СУРБД даст возможность использования преимуществ одновременного доступа к данным, поддержки целостности данных, защиты информации, ведения протоколов и многих других функций, реализованных на уровне сервера баз данных. Отсюда, требова-

ния к структуре данных наиболее жесткие и модель данных должна соответствовать как минимум третьей нормальной форме, которая предусматривает разветвленную систему справочников и связанных таблиц.

Другим способом практической реализации является использование офисного приложения MS Excel [1]. В этом случае, наличие высоко квалифицированного специалиста в области баз данных не обязательно (хотя и желательно). Современные версии MS Excel позволяют хранить значительные объемы информации, проводить первичную обработку данных, отображать данные в виде таблиц, графиков и диаграмм. Кроме того, MS Excel привычен для пользователя, не является сложным в обучении и использовании. Вместе с тем, при проектировании данных необходимо понимать, что данные являются реляционными, состоят из нескольких таблиц и требуют наличия специальных полей для связи между ними, а сама модель данных должна соответствовать первой нормальной форме.

Первая нормальная форма:

- а) запрещает повторяющиеся столбцы (содержащие одинаковую по смыслу информацию);
- б) запрещает множественные столбцы (содержащие значения типа списка);
- в) требует определить первичный ключ для таблицы, то есть тот столбец или комбинацию столбцов, которые однозначно определяют каждую строку.

Так как MS Excel не поддерживает целостности данных и позволяет записывать в ячейки информацию любого типа, необходимо отслеживать корректность заполнения на уровне пользователей. Т.е. наличие дополнительных символов в поле числового типа (например, запись «более 0,5 мг/л») является недопустимым.

Методика построения сопряженной базы данных состоит из нескольких этапов.

На первом этапе формируются списки работников, принимающих участие в исследовании. Обычно списки содержат паспортные данные, которые характеризуют пол, возраст, стаж работы, место жительства и др. Важно в таблицу со списками внести

ряд идентификаторов, отвечающих за сопряжение с другими массивами данных:

- внутренний идентификатор работника (обычно обозначается id) представляет собой номер по порядку или табельный номер, принятый на предприятии;
- серия и номер страхового полиса отвечает за связь с таблицами по заболеваниям работника, содержащимися в базе ТФОМС и амбулаторных картах;
- идентификатор рабочего места отвечает за связь с таблицами, содержащими данные аттестации рабочих мест;
- идентификатор клинико-лабораторных исследований отвечает за связь с таблицами, содержащими данные медицинских осмотров;
- прочие идентификаторы, отвечающие за связи с таблицами дополнительных данных.

В реляционных базах данных поля, содержащие идентификаторы, называют ключевыми полями. Ключевым полем может быть практически любое поле в таблице. Ключ может быть первичным или внешним. Первичный ключ однозначно определяет запись в таблице, в то время как внешний ключ используется для связи с первичным ключом другой таблицы. Для таблицы списков идентификатор работника является первичным ключом, а остальные идентификаторы – внешними ключами к другим таблицам.

Первый этап построения сопряженной базы данных является наиболее важным, и качественное его выполнение снимает множество проблем с последующей обработкой информации.

На втором этапе формируются таблицы, содержащие непосредственно интересующие данные по заболеваемости, характеристикам рабочих мест, социально-экономическим параметрам и др. Важно, чтобы эти таблицы содержали соответствующие идентификаторы, являющиеся первичными ключами. При передаче данных в электронном виде из другой базы разумно пользоваться идентификаторами (или ключами), принятыми в этой базе данных. При переносе данных с твердых копий или при создании новых таблиц можно использовать основной идентификатор работника.

Третий этап построения сопряженной базы данных заключается в формировании перекрестных ссылок на связанные таблицы. При использовании СУРБД ссылки формируются автоматически на основе первичных и внешних ключей, при использовании MS Excel ссылки формируются с помощью специальных функций. Кроме того, в MS Excel существует возможность объединения всех таблиц в одном файле на отдельных страницах.

Для решения ряда задач по оценке риска на одном из предприятий города Перми была сформирована база данных с использованием MS Excel. Модель данных содержит следующие таблицы:

- Списочный состав работников с указанием пола, даты рождения, стажа работы, место жительства. В таблице введен первичный ключ (id), указаны серия и номер полиса (Polis), введен идентификатор рабочего места (id_wp), идентификационный номер из базы клинико-лабораторных исследований (idn).
- Заболеваемость работников с указанием даты заболевания, код заболевания, длительность заболевания, стоимость лечения. В качестве первичного ключа используется серия и номер полиса (Polis).
- Результаты клинико-лабораторных исследований с указанием даты проведения обследования, измеряемого показателя и его значения. Первичный ключ – idn.
- Результаты социологического опроса с первичным ключом id.
- Результаты аттестации рабочих мест, первичный ключ id_wp.

Всего в базе данных собрана информация по 280 работникам, размер файла около 2 Мб.

Практика использования такой структуры базы данных показала, что временные издержки на обработку таблиц с помощью функций MS Excel не существенны, созданная система первичных и внешних ключей позволяет в динамическом режиме обращаться к любым данным по перекрестным ссылкам. При этом существенным моментом с точки зрения пользователя является прозрачность данных, которая дает возможность легко исправлять грубые ошибки ввода.

СПИСОК ЛИТЕРАТУРЫ

1. Гарнаев А. Использование MS Excel и VBA в экономике и финансах. – СПб.: БХВ-Петербург, 2002. – 420 с.
2. Грофф Дж., Вайнберг П. SQL: Полное руководство: пер. с англ. – 2-е изд., перераб. и доп. – К.: Издательская группа BHV, 2001. – 816 с.
3. О персональных данных: федер. закон Рос. Федерации от 27 июля 2006 г. № 152-ФЗ: принят Гос. Думой Федер. Собр. Рос. Федерации 8 июля 2006 г.: одобр. Советом Федерации Федер. Собр. Рос. Федерации 14 июля 2006 г. // Рос. газ. – 2006. – 29 июля.

**METHODICAL APPROACHES TO BUILDING A COMBINED
DATABASE OF HEALTH PARAMETERS AND SOCIAL
RISK FACTORS FOR SOMATIC NON-INFECTIOUS
DISEASES IN INDUSTRIAL WORKERS**

Kiryanov D.A., Lebedeva-Nesevrya N.A.

*FSSI «Federal Scientific Center for Medical and Prophylactic Health Risk Management
Technologies» of Federal State Service for Surveillance in the Sphere of Consumers
Protection and Human Well-Being, Perm
e-mail: kda@fcrisk.ru, e-mail: natnes@fcrisk.ru*

The study has developed methodical approaches to building of combined databases that include health parameters and indicators of social risk factors for somatic non-infectious diseases in industrial workers using the findings of empirical epidemiological, medicobiological and sociological investigations. The method can be applied to discover the mechanisms of social factors' impact on health disorders in individuals belonging to various socio-professional groups.

Keywords: combined database, health risk