

## **Таблицы сопряжённости в эпидемиологии профессиональных заболеваний.**

Федорович Г.В.

Д.ф.-м.н., технический директор ООО «НТМ-Защита»

E-mail: fedorgv@gmail.com

Таблицы сопряжённости (*далее* ТС) представляют собой форму представления статистических данных для их последующего эффективного анализа. В ТС заносят числа заполнения ячеек, на которые разбивается пространство признаков исследуемых объектов. При анализе зависимости в ТС выходных (зависимых) переменных от входных (определяющих) в статистике используются (см. напр.[1]) параметры «преобладание» и «отношение преобладаний». В эпидемиологии аналогом этих параметров являются «шансы» и «отношение шансов», дополнительно анализируются «риски». Последние совпадают с условными (по Байесу) вероятностями наблюдения определенных выходных переменных при заданных входных переменных. Информативны также «относительные риски».

Отличительной особенностью эпидемиологии профессиональных заболеваний (*далее* ПЗ) является многофакторность вредных условий труда и множественность нозологических форм ПЗ, обусловленных каждым из этих факторов. Следовательно, становятся многомерными как пространство признаков, так и соответствующие ТС.

### **1. Признаки и пространство признаков в эпидемиологии.**

Для более-менее информативного описания какой-либо совокупности объектов нужны определяющие их параметры - переменные, которым можно было бы присваивать количественные значения. Набор параметров, характеризующий определенный объект, является признаком этого объекта. Концептуальные переменные образуют систему координат (пространство признаков), в которой размещается совокупность описываемых объектов. Подробное описание использования признаков в эпидемиологии как инструмента исследований и метода решения практических задач дано в монографии [2].

Пространство признаков, образующее для аналитика систему координат, может быть различной размерности. Можно ограничиться описанием ситуации с помощью одной переменной. В этом случае пространство признаков сведется к линейному. Вдоль единственной оси координат будут откладываться значения переменной, встречающиеся при обследовании различных объектов. Однако, сбор эпидемиологической информации по одной переменной мало информативен. Нужны не просто сведения, а объяснения связей, поэтому необходимо анализировать сразу несколько переменных, т.е. работать в многомерном пространстве признаков. В области эпидемиологии ПЗ существует широкий спектр исследований профессионально обусловленных нарушений здоровья работников: от углубленного изучения ЗВУТ, до патологии хронических ПЗ. Общий вопрос, возникающий при этом: «Существует ли связь между объектами исследования – ВПФ и заболеваемостью работников?». Представляющие практический интерес ответы могут быть получены при статистическом корреляционном анализе данных в двумерном пространстве признаков.

В качестве входной переменной  $X$  следует взять интенсивность ВПФ. Для некоторых видов ВПФ (химический или биологический факторы, АПФД) это может быть концентрация вредных веществ в воздухе рабочей зоны. Для других – уровни воздействующих полей (акустического для шума и инфразвука, электромагнитных полей, теплового облучения). Для единообразия, в качестве единиц измерения можно взять отношение фактических уровней к ПДУ.

Выходную (зависимую) переменную  $Y$  следует определить через подходящим образом определенную заболеваемость работника. Это может быть физиологический показатель (например – уровень иммуноглобулинов в крови), либо показатель внешнего проявления заболеваемости (например – частота случаев заболеваний с временной утратой трудоспособности, далее ЗВУТ).

Каждый  $i$ -тый работник ( $i = 1, 2, \dots, N$ ) обследуемого предприятия характеризуется набором из двух чисел  $\{X_i; Y_i\}$ , являющимся признаком этого работника. На плоскости  $(X, Y)$  признак представляет точку с соответствующими координатами  $X_i$  и  $Y_i$ . Множество из  $N$  точек отражает положение с заболеваемостью на обследуемом предприятии. Соответствующая четвертьплоскость ( $X > 0, Y > 0$ ) представляет собой пространство признаков для этого обследования.

Набор признаков может иметь вид, представленный на рис.1.

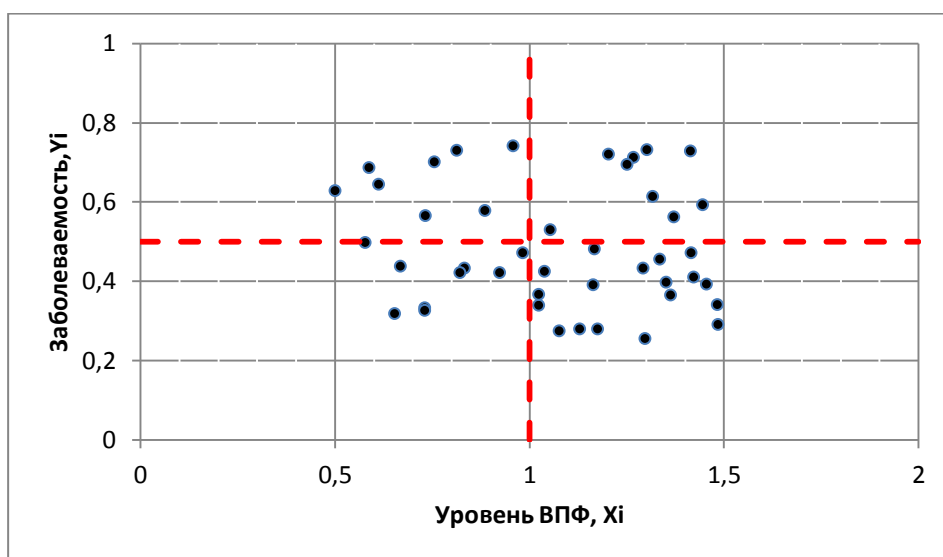


Рисунок 1. Портрет трудового коллектива в пространстве признаков.  
Зависимость ПЗ от ВПФ отсутствует

Множество изображающих точек образует более-менее симметричное во всех направлениях «облако». При любом возможном уровне ВПФ распределение заболеваемости работников одинаково. Очевидно, здесь нет корреляции между ВПФ и заболеваемостью.

Другая ситуация изображена на рис.2. Здесь явно видно, что заболеваемость работников растет с ростом уровня ВПФ.

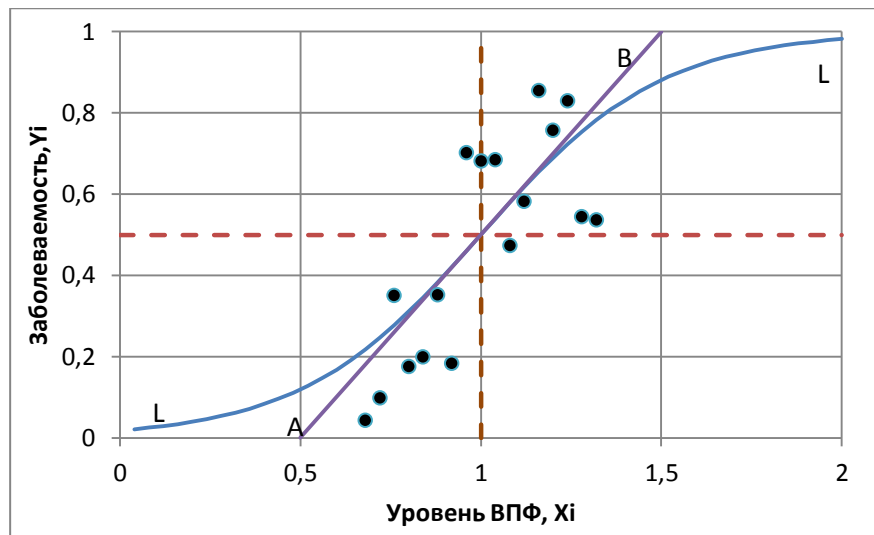


Рисунок 2. Портрет трудового коллектива, заболеваемость в котором растет с ростом уровня ВПФ.

Для количественной оценки характера и степени связи величин  $X$  и  $Y$  используется коэффициент линейной корреляции (коэффициент Пирсона):

$$r = \frac{\sum(X_i - \langle X \rangle) * (Y_i - \langle Y \rangle)}{\sqrt{\sum(X_i - \langle X \rangle)^2 * \sum(Y_i - \langle Y \rangle)^2}} \quad (1)$$

Здесь через  $\langle X \rangle$  и  $\langle Y \rangle$  обозначены средние значения соответствующих координат.

Вообще говоря, коэффициент корреляции может меняться от 1 до -1. Значению  $r = 0$  соответствует отсутствие зависимости между величинами  $X$  и  $Y$ , как на рис.1. Если существует строгая зависимость  $Y$  от  $X$ , например – прямая пропорциональность, изображаемая прямой А-В на рис.2, то  $r = 1$ . Для «облака» точек на рис.2 величина  $r$  лежит между 0 и 1.

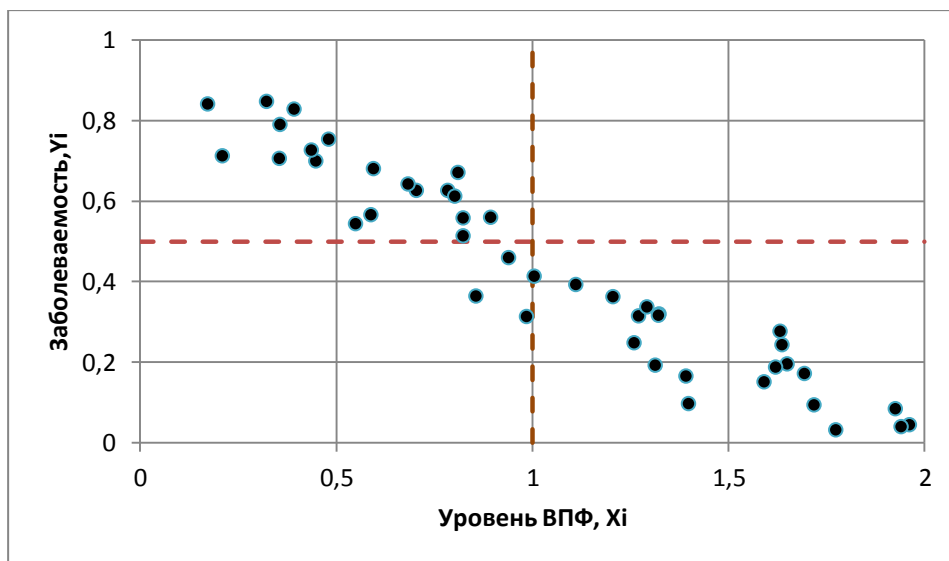


Рисунок 3. Портрет трудового коллектива, заболеваемость в котором убывает с ростом уровня воздействующего фактора.

Бывают ситуации, изображенные на рис.3, когда «облако» точек вытянуто так, что с ростом координат  $X_i$  величина координат  $Y_i$  в среднем убывает. Если в исходных данных нет систематической ошибки, следует признать, что влияние исследуемого фактора благотворно действует на здоровье работников – уменьшает их заболеваемость. Для таких данных коэффициент корреляции лежит между 0 и -1.

Коэффициент Пирсона (1) свидетельствует только о наличии или отсутствии связи между переменными  $X$  и  $Y$ . Если связь есть, т.е. портреты имеют вид рис.2 или рис.3, анализ можно продолжить, определив количественную зависимость заболеваемости  $Y$  от интенсивности воздействия  $X$ . Такие возможности предоставляет регрессионный анализ статистических данных.

Если предположить, что зависимость линейная, т.е. имеет вид

$$Y = a + b * X \quad (2)$$

то коэффициенты этой зависимости подсчитываются по формулам

$$b = \frac{\sum(X_i - \langle X \rangle) * (Y_i - \langle Y \rangle)}{\sum(X_i - \langle X \rangle)^2}, \quad a = \langle Y \rangle - b * \langle X \rangle \quad (3)$$

Для данных на рис.2 эта зависимость представлена прямой А-В. Сопоставляя выражения (1) и (3), можно заметить, что когда равен нулю коэффициент корреляции Пирсона, тогда же обращается в нуль и коэффициент  $b$  наклона регрессионной зависимости заболеваемости  $Y$  от интенсивности  $X$  воздействия ВПФ: связь  $Y$  с  $X$  отсутствует.

В некоторых случаях, при подборе функции, дающей оценку зависимости  $Y$  от  $X$ , следует учитывать эффект ограниченности возможных значений  $Y$ . Она, например, не может быть отрицательной. Кроме того, при характеристике заболеваемости количеством дней болезни на одного работника в год, следует учитывать, что эта величина не может превосходить полного количества дней в году. Количество выявляемых случаев профзаболеваний не может превосходить полного количества обследуемых работников и т.п. В биостатистике для этих случаев используется т.н. логистическая функция (кривая L на рис. 2), максимальная величина которой ограничена:

$$Y = Y_0 * \{1 + \exp[-(X - X_0) / \delta X]\}^{-1} \quad (4)$$

Здесь  $X_0$ ,  $Y_0$  и  $\delta X$  - характерные масштабы зависимости, подбираемые при анализе данных методом логистической регрессии. В эпидемиологии профзаболеваний логистическая регрессия используется редко, т.к. уровень заболеваемости редко бывает столь велик, чтобы проявлялся эффект ограниченности возможных значений  $Y$ .

## 2. Эпистемическая инверсия.

Термин в названии раздела введен в социологии. Суть эффекта эпистемической инверсии в том, что при статистических исследованиях уже не единица исследования характеризуется определенным признаком, а сам признак характеризуется частотой - количеством единиц, которым присущ данный признак. Несмотря на рутинность этого подхода, его всегда следует иметь в виду, анализируя десятки и сотни историй болезни, комбинируя самые различные сведения. Формально, ни один из этих больных сам по себе не представляет самостоятельного интереса для анализа. Интерес представляет:

- наполнение ячеек в пространстве признаков,
- распределение частот по всем значениям переменных,
- изменения распределений при увеличении размерности пространства (введении и рассмотрении дополнительных признаков),
- наличие связи между признаками и ее вариабельность в различных контекстах.

Эпистемическая инверсия делает эпидемиологию менее эмпиричной, чем это кажется на стадии сбора исходной информации, где происходит нечто вроде общения исследователя с живым человеком. Если иметь в виду только формальную структуру эпидемиологической теории, можно сказать, что в ней нет ничего, кроме пространства признаков и единиц, размещаемых в пространстве в соответствии с содержанием этих признаков.

Техника эпистемической инверсии сводится к следующему.

Предположим, что интенсивность  $x$  воздействия ВПФ имеет  $r$  градаций (или уровней)  $X_1, X_2 \dots X_r$ , а уровень заболеваемости  $y$  имеет  $s$  градаций  $Y_1, Y_2, \dots Y_s$ . Вертикальные линии, проведенные через точки  $X_i$  ( $i = 1, \dots, r$ ) и горизонтальные, проведенные через точки  $Y_k$  ( $k = 1, \dots, s$ ) разбивают фазовую плоскость на ячейки (см.рис.1-3). Если количества  $n_{ij}$  точек, попавших в соответствующие ячейки, занести в таблицу, состоящую из  $r$  строк и  $s$  столбцов, то так отформатированные результаты принимают вид ТС, обычно используемой в эпидемиологии ПЗ. Метод построения приводит к очевидному свойству таких таблиц: если между  $x$  и  $y$  имеется статистически значимая связь, максимальные величины  $n_{ik}$  группируются вдоль диагонали ТС. Статистический характер результатов приводит к тому, что некоторое число точек попадает в недиагональные ячейки таблицы. При этом перед исследователем встает задача определить – насколько уверенно можно предсказать одну величину по значению другой.

В отличие от регрессионного анализа, в данном случае представляет интерес не столько конкретные значения коэффициентов в зависимости  $Y$  от  $X$  типа (2), сколько надежная и непротиворечивая оценка степени и характера влияния ВПФ на уровень ПЗ. Переводя проблему в практическую плоскость, следует оценить условную вероятность уровня ПЗ при наблюдаемом значении ВПФ.

ТС представляют собой аппарат, с помощью которого можно оценивать как вероятность причинения ущерба здоровью, так и математическое ожидание ущерба различного вида (финансового, имущественного, потерь рабочего времени и пр.). Вообще говоря, вероятности можно подсчитывать и без ТС. В простейших случаях – поделив определенный результат испытаний на полное число испытаний, получим вероятность этого результата. Такие вероятности, однако, представляют небольшой интерес в эпидемиологии (общей и профессиональных заболеваний, в частности). Здесь более интересны условные вероятности. Разница принципиальная, в количественном отношении – очень существенная.

### **3. Эпидемиологическая трактовка результатов анализа ТС. Значимость результатов.**

Разбиение шкал интенсивности воздействия и уровня заболеваемости на несколько градаций обычно используются в фундаментальных исследованиях. В рутинных эпидемиологических исследованиях ПЗ, однако, применяется дифференцирование по бинарной шкале «да - нет». Для таких исходных данных при переходе от пространства признаков к таблице сопряженности следует использовать дихотомическую (по шкале «да-нет») градацию как интенсивности воздействия ВПФ, так и уровня заболеваемости. Суммирование идет по частям фазовой плоскости «ВПФ-ПЗ», разграниченными пунктирными линиями  $X = X_0$  и  $Y = Y_0$  на рис. 1-3. Предполагается, что ВПФ действует (ВПФ = 1) только в области  $X > X_0$  и не действует (ВПФ = 0) в остальной части фазовой плоскости ( $X < X_0$ ). Аналогично, работники с заболеваемостью  $Y < Y_0$  считаются здоровыми (ПЗ = 0), а к болеющим (ПЗ = 1) относят только тех, заболеваемость которых превышает уровень  $Y_0$ .

Результат – ТС размером 2x2, структура которой приведена ниже:

Таблица 1

Таблица сопряженности  
результатов эпидемиологических исследований ПЗ

	ПЗ нет (k = 0)	ПЗ есть (k = 1)	Всего по факторам
ВПФ нет (i = 0)	$n_{00}$	$n_{01}$	$n_{0*} = n_{00} + n_{01}$
ВПФ есть (i = 1)	$n_{10}$	$n_{11}$	$n_{1*} = n_{10} + n_{11}$
Всего по заболеваемости	$n_{*0} = n_{00} + n_{10}$	$n_{*1} = n_{01} + n_{11}$	$n^{**} = n_{00} + n_{01} + n_{10} + n_{11}$

Входные переменные (аргументы) отмечаются индексами  $i$  (воздействие ВПФ). Нулевым индексом отмечается отсутствие воздействия, а индексом единица - его наличие. Выходная переменная отклика (заболеваемость) отмечается индексом  $k$  так, что  $k = 0$  отмечает количество не болеющих (здоровых), а  $k = 1$  – количество заболевших. Сами количества людей, попавших в ту или иную группу, обозначаются переменной  $n_{ik}$  с соответствующими индексами. Например, число  $n_{00}$  обозначает количество здоровых людей, не подвергающихся воздействию ВПФ. Число  $n_{11}$  обозначает количество больных, подвергавшихся воздействию ВПФ. Аналогично интерпретируются числа  $n$  с другими индексами. Таблица 1 чисел  $\{n_{ik}\}$  дает полное описание ситуации с влиянием входной переменной (вредное воздействие) на уровень отклика (заболеваемость) в обследуемом коллективе.

В дальнейшем придется обращаться к вычислениям различных сумм чисел  $n_{ik}$  по индексам. Будем обозначать их, ставя значок \* вместо того индекса, по которому произведено суммирование. Например,

$$n_{*k} = \sum_i n_{ik} , \quad n^{**} = \sum_{ik} n_{ik} \quad (5)$$

Смысл этих чисел очевиден:  $n_{*k}$  представляют собой числа здоровых ( $k=0$ ) или больных ( $k=1$ ) людей, безотносительно к воздействию (либо к его отсутствию) ВПФ,  $n_{i*}$  представляют собой числа людей не подвергающихся воздействию ВПФ ( $i=0$ ) или подвергающихся такому воздействию ( $i=1$ ), независимо от того – больны они или нет.

Сумма  $n_{**}$  определяет объем выборки – общее число людей, вошедших в обследуемую группу.

Данные табл. 1 позволяют перейти от случаев к вероятностям. Например, вероятность заболеть  $P(\text{ПЗ} = 1)$  в обследованном коллективе определяется соотношением

$$P(\text{ПЗ} = 1) = n_{*1} / n_{**} \quad (6)$$

Вероятность попасть под действие ВПФ  $P(\text{ВПФ} = 1)$  равна

$$P(\text{ВПФ} = 1) = n_{1*} / n_{**} \quad (7)$$

Большой интерес, чем обычные вероятности, представляют условные вероятности. Можно оценить условную вероятность для работника заболеть, если на него действует ВПФ :

$$P(\text{ПЗ} = 1 | \text{ВПФ} = 1) = n_{11} / n_{1*} \quad (8)$$

Последнюю вероятность следует отличать от вероятности обнаружить воздействие на работника ВПФ, если известно, что он заболел:

$$P(\text{ВПФ} = 1 | \text{ПЗ} = 1) = n_{11} / n_{*1} \quad (9)$$

Вероятности (6-7) можно рассматривать как априорные, а (8-9) – как апостериорные. Как таковые, последние подчиняются соотношению, известному как [теорема Байеса](#) .

$$P(\text{ПЗ}=1|\text{ВПФ}=1)=P(\text{ВПФ}=1|\text{ПЗ}=1)*P(\text{ПЗ}=1)/P(\text{ВПФ}=1) \quad (10)$$

Аналогичные соотношения можно выписать и для вероятностей, соответствующих отсутствию ВПФ или заболевания.

Если зависимость «ВПФ-ПЗ» значима, можно оценить добавочную вероятность заболевания  $\delta P$ , обусловленную воздействием ВПФ. Она определяется разностью между вероятностью заболевания при действии ВПФ:  $P(\text{ПЗ}=1|\text{ВПФ}=1)$  (см. ф-лу (8)) и «фоновой» вероятностью:  $P(\text{ПЗ}=1|\text{ВПФ}=0)$  :

$$\delta P = P(\text{ПЗ}=1|\text{ВПФ}=1) - P(\text{ПЗ}=1|\text{ВПФ}=0) = (n_{11}n_{00} - n_{10}n_{01}) / (n_{1*}n_{0*}) \quad (11)$$

Связь данных в ТС с априорными и апостериорными вероятностями приводит к возможности использования аппарата вероятностной Байесовской логики. Это направление оказалось весьма плодотворным при анализе экологических и медицинских данных. Пример использования методов Байесовской классификации и Байесовского вывода в области оценки условий труда приведены, например, в [3] и [4].

#### 4. Значимость результатов.

В случае существования однозначной зависимости «ВПФ-ПЗ» в таблице сопряженности недиагональные элементы  $n_{01}$  и  $n_{10}$  должны быть равны нулю. Реально,

однако, существует «фоновая» вероятность заболевания (без влияния ВПФ). Она определяется соотношением

$$P(\text{ПЗ} = 1 \mid \text{ВПФ} = 0) = n_{01} / n_{0*} \quad (12)$$

Вообще говоря, существует также ненулевая вероятность остаться здоровым при действии ВПФ:

$$P(\text{ПЗ} = 0 \mid \text{ВПФ} = 1) = n_{10} / n_{1*} \quad (13)$$

Ненулевые вероятности (12) и (13) исключают однозначность зависимости «ВПФ-ПЗ», однако, в этом случае можно поставить вопрос о статистической значимости связи признаков ВПФ и ПЗ.

Возникающая здесь проблема связана с оценкой достоверности таких заключений. Содержимое клеток ТС может быть случайным, поэтому в статистике принято использовать специальные критерии для проверки гипотез о значимости различий. Наиболее распространен критерий  $\chi^2$ , предложенный К.Пирсоном. С помощью него оценивается значимость различий между фактическим (выявленным в результате исследования) количеством  $n_{ik}$  исходов, попадающим в каждую категорию, и случайным количеством  $m_{ik}$ , которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы.

В последнем случае для совместной вероятности  $P(\text{ВПФ}; \text{ПЗ})$  имеет место соотношение

$$P(\text{ВПФ}; \text{ПЗ}) = P(\text{ВПФ}) * P(\text{ПЗ}) \quad (14)$$

Иными словами, если бы данные (количества работников) в клетках таблицы сопряженности были бы независимы, то их можно было бы определять по формулам:  $m_{ik} = n_{i*} n_{*k} / n_{**}$ . И обратно, степень зависимости данных можно определить, вычисляя разности  $m_{ik} - n_{ik}$ . Общепринятым критерием независимости данных в таблицах сопряженности является критерий  $\chi^2$  (К.Пирсон, Р.Фишер), определяемый как сумма квадратов таких разностей, деленных на ожидаемые количества работников  $m_{ik}$ .

$$\chi^2 = \sum_{ik} (n_{ik} - m_{ik})^2 / m_{ik} \quad (15)$$

После подстановок значений  $m_{ik}$ , получим окончательно

$$\chi^2 = n_{**} \frac{(n_{11}n_{00} - n_{10}n_{01})^2}{n_{1*}n_{*1}n_{0*}n_{*0}} \quad (16)$$

Критическое значение величины  $\chi^2$  зависит от числа степеней свободы исходных данных и выбранного уровня значимости утверждения. Для четырехпольных таблиц число степеней свободы равно 1, если выбрать уровень значимости 0,05, то критическое значение будет равно 3,84.

Существуют ограничения на применение величины  $\chi^2$  в качестве критерия зависимости данных в таблице сопряженности (см. напр. [2]). Для расчета должны



использоваться только фактические данные (не нормированные доли), причем ожидаемые числа заполнения ячеек  $m_{ik}$  не должны быть менее 10. С другой стороны, величина  $\chi^2$  растет вместе с ростом объема выборки, поэтому слишком большие выборки также не дают достоверной информации.

От этого недостатка свободен критерий  $\phi$ , определяемый по формуле

$$\phi = \sqrt{\frac{\chi^2}{n_{**}}} \quad (17)$$

Величина  $\phi$  может меняться от 0 (отсутствие связи) до 1 (однозначная зависимость). Заметим, что добавочная вероятность  $\delta P$ , определяемая формулой (11), пропорциональна величине критерия  $\phi$ , хотя и не совпадает с ней.

Более детальная статистическая интерпретация величины критерия  $\phi$  приведена в табл. 2:

Таблица 2

Качественные заключения о силе связи  
в зависимости от значений критерия  $\phi$

Значение $\phi$	Сила связи
< 0,1	Несущественная
0,1 – 0,2	Слабая
0,2 – 0,4	Средняя
0,4 – 0,6	Относительно сильная
0,6 – 0,8	Сильная
0,8 – 1,0	Очень сильная

## 5. Классификация условий труда. Заболеваемость с временной утратой трудоспособности. Хронические профзаболевания.

Классы условий труда (далее КУТ) определяются через уровни заболеваемости во многих документах (см. напр. [5], [6]). Условия труда по степени вредности и опасности подразделяются на 4 класса - оптимальные, допустимые, вредные (включают 4 подкласса) и опасные.

Принятое подразделение условий труда на классы имеет дескриптивный характер и не удовлетворяет требованиям количественного описания причинно-следственных связей. Только в сопоставлении класса вредности и опасности условий труда с количественными показателями здоровья и величины утраты трудоспособности можно судить об истинной причине его повреждения. Именно ТС представляют удобный способ количественной привязки уровней заболеваемости к классам условий труда.

Заболеваемость с временной утратой трудоспособности (далее ЗВУТ) занимает (в связи с высокой экономической значимостью) особое место в общей статистике заболеваемости. Его можно считать приоритетной характеристикой состояния здоровья работников. Единицей наблюдения при изучении ЗВУТ является каждый случай невыхода

на работу в связи с заболеванием или травмой. Учетным документом служит листок нетрудоспособности, являющийся не только медицинским, но и статистическим и юридическим документом. В 90-е годы прошлого века статистические отчеты по ЗВУТ составлялись лечебно-профилактическими учреждениями по форме 16-ВН (постановление Госкомстата N42 от 24.04.94 г.). Анализ ЗВУТ по отдельным строкам формы 16-ВН, отражающим нозологические формы болезней позволял выявить производственно обусловленные заболевания, связанные с особенностями условий труда. Такая работа была проделана в середине 90-х годов в НИИ Медицины Труда РАМН. Были получены средние по стране данные по уровням ЗВУТ на предприятиях, характеризующихся различными условиями труда [7],[8].

Представляющие интерес для дальнейшего результаты исследования [7] зависимости показателей заболеваемости от КУТ приведены в колонках табл.3. Используются обозначения:  $\Delta K$  – диапазон количества случаев ЗВУТ,  $\Delta D$  – диапазон числа дней нетрудоспособности,  $\langle K \rangle$  и  $\langle D \rangle$  – соответствующие средние значения. Эти данные нормированы на  $N=100$  работников и на период наблюдения  $Y = 1$  год (365 дней) . Риск (вероятность) заболевания определялась через среднее количество одновременно болеющих  $n_{i1} = \langle D \rangle / Y$  (предпоследняя колонка) по формуле  $R_i = n_{i1} / N$ . Индексация величин  $n$  и  $R$  как выше.

Таблица 3.

Характеристики ЗВУТ  
(по всем болезням, на 100 работников в год)

КУТ	Баллы (i)	$\Delta K$ , случ.	$\Delta D$ , дни	$\langle K \rangle$	$\langle D \rangle$	$n_{i1}$	$R_i$
2	1	36,4 - 72,3	639 - 938	54,35	802,5	2,2	0,022
3.1	2	72,4 - 84,6	939-1081	78,5	1010	2,8	0,028
3.2	3	84,7 - 90,7	1082-1153	87,7	1117,5	3,1	0,031
3.3	4	90,8 - 96,8	1154-1225	93,8	1189,5	3,3	0,033
3.4	5	96,9 - 102,9	1226-1281	99,9	1253,5	3,4	0,034

Обращают на себя внимание небольшие относительные риски ПЗ при работе во вредных условиях труда. Если за контрольную группу принять работающих в условиях КУТ 2, то, например, относительный риск ПЗ при работе в условиях КУТ 3.4 равен  $\approx 1,55$ . В [9] такой относительный риск рассматривается как признак средней степени связи нарушений здоровья с работой. Одной из возможных причин этого результата может быть «смешанность» данных по нозологии ЗВУТ. Действительно, в [7] суммировались данные по всем видам и возможным причинам заболеваний. Если в качестве исходного материала брать данные, дифференцированные по нозологии ЗВУТ, можно получить более значимые результаты.

Еще сложнее технически проблема исследования связи хронических ПЗ с условиями труда. В уже цитированной работе [7] была предпринята попытка учета ПЗ в масштабах страны, аналогично тому, как это было сделано для ЗВУТ. Действующая в России государственная система регистрации ПЗ предусматривает централизованный сбор первичного материала о вновь выявленных больных. Однако для решения проблемы стандартизованных критериев сравнения эти сведения совершенно недостаточны.

Необходима информация с различных территорий России о накопленной заболеваемости, численности работающих в контакте с конкретными производственными факторами, численности подлежащих периодическим медицинским осмотрам и осмотренных, численности и составу по возрасту, стажу, нозологическим формам, степени утраты трудоспособности всего контингента больных и инвалидов с профзаболеваниями, течения и исхода заболеваний. Без решения этих вопросов регистрируемые уровни ПЗ не отражают истинной картины. Перевод предприятий из государственной в частную собственность также значительно затрудняет контроль в масштабах страны, как за условиями труда, так и за медицинским обслуживанием работников. Для практических целей необходимо разработать и внедрить в практику новые машино-ориентированные учетно-отчетные статистические формы, организовать их обработку и анализ результатов.

Принимая во внимание изложенное, можно использовать данные работ [7] и [8] только в качестве начальных (требующих корректировки) указаний на корреляцию уровней ПЗ с условиями труда.

Представляющие интерес для дальнейшего исследования [7] показатели заболеваемости (числа случаев ПЗ, регистрируемых ежегодно в расчете на 1000 человек) для различных условий труда, оцениваемых по показателям КУТ и по соответствующим баллам, приведены в третьей строке табл.4.

Таблица 4.

Профессиональная заболеваемость  
(число случаев на 1000 работников в год)  
по классам условий труда.

КУТ	1 и 2	3,1	3,2	3,3	3,4	4,0
Баллы	1	2	3	4	5	6
Диапазоны	0	0 – 1,5	1,6 - 5	5,1 - 15	15,1 - 50	> 50

Сами авторы работы подчеркивают, что регистрируемые уровни ПЗ не отражают истинной картины. Это связано с ухудшением медико-санитарного обслуживания работающих в промышленности, недостаточной квалификацией медработников по профессиональной патологии, плохим аппаратным оснащением, переводом предприятий из государственной в частную собственность. Все это значительно затрудняет контроль, как за условиями труда, так и медицинским обслуживанием работников.

Существует также неопределенность относительно того, в исследованиях какого типа получены данные НИИ МТ РАМН. Параметр «Заболеваемость» характеризует когортные исследования, т.е. длительные наблюдения за избранной группой (когортой), направленные на определение частоты новых случаев заболеваний в исследуемой группе. Предполагается, что в качестве когорт выбраны работники, трудящиеся в условиях с определенным КУТ. Именно по отношению к численности когорт определялся процент лиц, у которых в течение года обнаруживалось профессиональное заболевание. В отличие от этого, данные табл. 4 пересчитаны на «число случаев на 1000 работников данной профессии, производства». Известно, однако, (см.напр.[10]) что различные производства характеризуются различными уровнями ПЗ. Так, относительная частота ПЗ меняется от половины процента в добывающей промышленности до тысячных долей процента в

отраслях нематериального производства. Неопределенность выбора когорты в работе [7] может быть причиной недостоверности результатов.

## **6. Многомерные таблицы сопряженности в эпидемиологии профессиональных заболеваний.**

Отличительной чертой эпидемиологии профзаболеваний является комплексный характер вредных воздействий на работника. На большинстве вредных производств обычно одновременно имеют место неблагоприятные микроклиматические условия, повышенные уровни шума, вибрации и концентраций вредных химических веществ (в том числе АПФД) в воздухе рабочей зоны и пр. Более того, не только условия труда, но и такой непродуцируемый фактор, как, например, курение, способен оказать неблагоприятное влияние на уровень заболеваемости работников [11]. Этим производственные условия отличаются от «классических» эпидемиологических ситуаций, когда эпидемию вызывает, как правило, какой-либо один (например – инфекционный) агент.

Продемонстрируем эффективность и особенности использования ТС для анализа множественного воздействия ВПФ. Используются реальные данные по уровням хронических бронхолегочных заболеваний (*далее* ХБЛЗ) у работников добывающих отраслей промышленности. Значение бронхолегочной патологии во многом определяется тем, что ее доля превалирует в структуре профессиональной заболеваемости. Например, в добывающих отраслях промышленности эта доля достигает 75% . Детальные результаты разносторонних исследований в этой области профэпидемиологии опубликованы в последнее время в ряде работ [12], [13].

Влияния загрязнения воздуха рабочей зоны вредными химическими веществами, АПФД, других ВПФ, а также воздействие курения работников на уровень заболеваемости ХБЛЗ детально изучалось, например, в работе [11]. Было обследовано 1300 работников Кольской горно-металлургической компании. Обследуемые работники были разделены на группы курящих и некурящих, а также в зависимости от уровней воздействия ВПФ. Результаты (численность групп работников с учитываемыми признаками) сведены в табл.5, это пример многомерной (2x2x2) ТС.

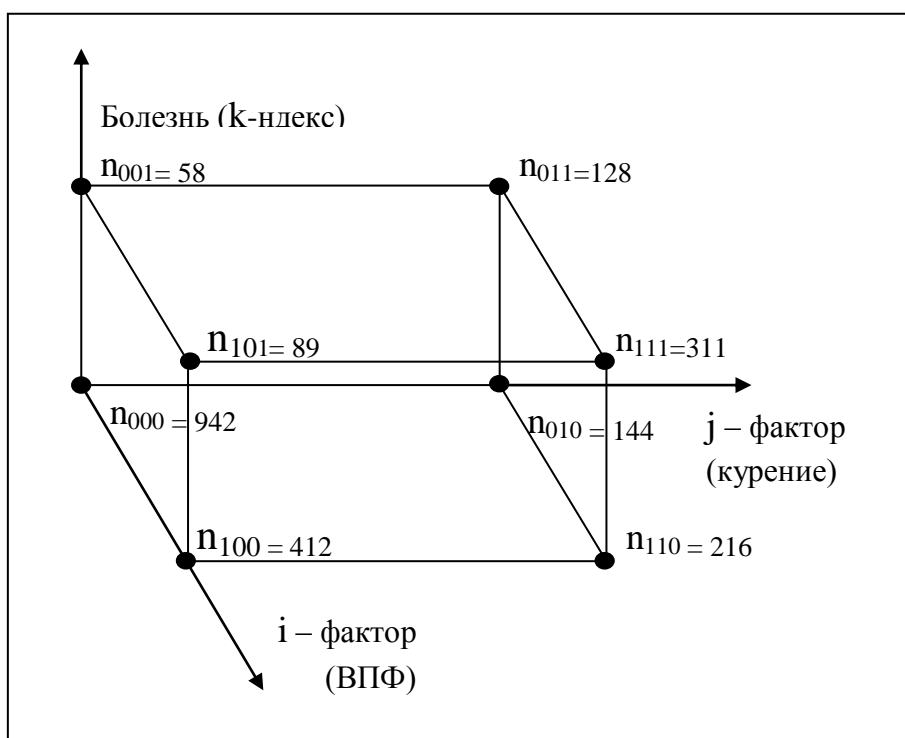
Таблица 5

Численности здоровых и больных ХБЛЗ

ВПФ	Курение	Здоров	Болен
нет	нет	942	58
	есть	144	128
есть	нет	412	89
	есть	216	311

Недостающая в работе [11] информация о контрольной группе (не подвергающихся воздействию ВПФ и не курящих) была восполнена данными статистической отчетности по стране в целом [14].

Использованные выше в разд.3 обозначения для содержимого клеток ТС естественно обобщаются на случай трехмерной  $2 \times 2 \times 2$  таблицы. Будем отмечать входные переменные (вредные воздействия) индексами  $i$  (воздействие ВПФ) и  $j$  (курение). Равенство нулю индекса обозначает отсутствие соответствующего воздействия, а индекс равный единице обозначает его наличие. Выходную переменную отклика (заболеваемость) будем отмечать индексом  $k$  так, что  $k = 0$  отмечает количество здоровых, а  $k = 1$  отмечает количество болеющих. Сами количества людей, попавших в ту или иную группу, будем обозначать переменной  $n_{ijk}$  с соответствующими индексами. Например, число  $n_{000}$  обозначает количество здоровых людей, не подвергающихся воздействию ВПФ и не курящих. Соответственно, число  $n_{111}$  обозначает количество больных, подвергающихся воздействию ВПФ и курящих. Аналогично интерпретируются числа  $n$  с другими индексами. Трехмерная таблица чисел  $\{n_{ijk}\}$  дает полное описание ситуации с влиянием входных переменных (вредных воздействий) на уровень отклика (заболеваемости) в обследуемом коллективе.



Для наглядного представления ситуации, привяжем числа  $n_{ijk}$  к вершинам куба, совмещенного с началом координатной системы  $\{i, j, k\}$  (см. рисунок). Так же, как и для  $2 \times 2$  таблицы, будем использовать различные суммы чисел  $n_{ijk}$  по индексам. Именно, будем ставить значок \* вместо того индекса, по которому произведено суммирование. Например,

$$n^{*jk} = \sum_i n_{ijk} \quad , \quad n^{**k} = \sum_{i,j} n_{ijk} \quad , \quad n^{***} = \sum_{ijk} n_{ijk} \quad (18)$$

Смысл этих чисел очевиден: например,  $n^{*jk}$  представляет собой числа здоровых ( $k=0$ ) и больных ( $k=1$ ) людей, не курящих ( $j=0$ ) или курящих ( $j=1$ ), безотносительно к воздействию (либо к его отсутствию) ВПФ. Суммы  $n^{**k}$  определяют числа здоровых ( $k=0$ ) и больных ( $k=1$ ) людей, безотносительно к воздействию (либо к его отсутствию) ВПФ или курения. Наконец, сумма  $n^{***}$  определяет общее число людей, вошедших в обследуемую группу.

Стоит отметить, что суммирование  $n_{ijk}$  по одному из индексов переводит исходную  $2 \times 2 \times 2$  ТС в таблицу меньшей ( $2 \times 2$ ) размерности, которую можно анализировать описанными выше методами. В статистике этот прием называется «сжатием» таблицы по индексу суммирования.

Перейдем к анализу рисков по данным таблицы  $2 \times 2 \times 2$ . Использование ТС в виде, представленном на рисунке, делает очевидной оценку рисков заболевания под действием различного набора  $\{i, j\}$  факторов: следует выбрать соответствующее этому набору вертикальное ребро куба и поделить число, попавшее на верхнюю грань (количество больных) на сумму чисел на этом ребре. Используя очевидную индексацию  $R_{ij}$  для обозначения риска заболевания под действием факторов  $i$  и  $j$ , получаем:

$$R_{ij} = n_{ij1} / n_{ij*} \quad (19)$$

Результирующие риски сведены в табл.б.

Таблица 6

Риски заболевания ХБЛЗ  
под действием различных факторов

ВПФ\курение	Нет ( $i=0$ )	Есть ( $i=1$ )
Нет ( $i=0$ )	0,058	0,471
Есть ( $i=1$ )	0,178	0,590

Для оценки относительного риска в качестве референтного значения следует выбрать риск  $R_{00}$  заболевания в контрольной группе, члены которой не курят и не подвергаются воздействию ВПФ. В этом случае

$$RR_{10} = R_{10}/R_{00} = 3,0 \quad ; \quad RR_{01} = R_{01}/R_{00} = 8,1 \quad ; \quad RR_{11} = R_{11}/R_{00} = 10,1 \quad (20)$$

Видно, что относительный риск заболевания ХБЛЗ для курящих ( $RR_{01}$ ) более чем вдвое превышает риск заболевания под действием ВПФ ( $RR_{10}$ ). Он еще более повышается (до  $RR_{11}$ ), если оба фактора воздействуют совместно.

Еще большие различия характерны для атрибутивных (дополнительных) рисков  $AR_{ij} = R_{ij} - R_{00}$  :

$$AR_{10} = R_{10} - R_{00} = 0,12 ; AR_{01} = R_{01} - R_{00} = 0,413 ; AR_{11} = R_{11} - R_{00} = 0,532 \quad (21)$$

Атрибутивный риск заболевания ХБЛЗ для курящих ( $AR_{01}$ ) более чем втрое превышает риск заболевания под действием ВПФ ( $AR_{10}$ ). Примечательно, что риск заболевания при совместном действии обеих факторов (ВПФ и курение) практически равен сумме рисков под действием каждого из них :  $AR_{11} \approx AR_{10} + AR_{01}$  . Это обстоятельство свидетельствует об отсутствии синергетического эффекта, когда действие одного фактора усиливало бы действие другого.

Приведенным значениям относительных рисков соответствуют этиологические доли:

$$EF_{10} = 0,672 ; EF_{01} = 0,876 ; EF_{11} = 0,901 . \quad (22)$$

Первая характеризует (см.[9]) связь заболеваемости с действием ВПФ и курением как «высокую», последние – как «почти полную».

## 7. Литературные источники.

- [1] **Аптон Г.** Анализ таблиц сопряженности (пер. с англ.). – М.: Финансы и статистика, 1982 – 143 с.
- [2] **Федорович Г.В.** Рациональная эпидемиология профессиональных заболеваний. - Saarbrucken, Deutschland: Palmarium Academic Publishing , 2014 – 343 p.
- [3] **Федорович Г.В.** Классификация условий труда по эпидемиологическим данным // Безопасность и охрана труда, № 4, 2011 г., с.49 – 52
- [4] **Федорович Г.В.** Профессиональный риск: количественная оценка и управление // Безопасность и охрана труда, № 1, 2012 г., с.60 – 64
- [5] **Федеральный закон «О специальной оценке условий труда»** № 426 – ФЗ от 28.12.2013 г.
- [6] **Р 2.2.2006 – 05.** Руководство по гигиенической оценке факторов рабочей среды и трудового процесса. М.: Минздрав России, 2005, 142 с.
- [7] **Молодкина Н.Н.** Профессиональный риск и защита здоровья работающих // РЖ "Пенсия" - М.: 1999. - 170 с.
- [8] **Молодкина Н.Н., Радионова Г.И., Денисов Э.И.** Обоснование критериев профессионального риска // Профессиональный риск / Измеров Н.Ф. (ред). – М.: Социздат, 2001. – С. 48 - 55.
- [9] **Р 2.2.1766-03** «Руководство по оценке профессионального риска для здоровья работников. Организационно-методические основы, принципы и критерии оценки» // М., - Минздрав РФ, 2004 г. – 17 с.
- [10] **Тимофеева Е.И., Федорович Г.В.** Страховые тарифы и профессиональная эпидемиология // Безопасность и охрана труда, № 4, 2011 г., с.44 – 48.
- [11] **Артамонова В.Г., Мухин Н.А.** Профессиональные болезни. – Медицина, М., 2004 – 480 с.
- [12] **Сюрин С.А.** Вдох или выдох // Безопасность и охрана труда, № 2, 2013 г., с.67 – 69.

[13] **Сюрин С.А.** Особенности формирования профессиональной патологии у работников различного передела никеля в условиях Крайнего Севера // Безопасность и охрана труда, № 1, 2012 г., с.50 – 51.

[14] **Россия в цифрах 2010** /Краткий статистический сборник.-М.: Росстат.-2010.-560 с.