

*О.А. КОЗИНА*, канд. техн. наук, ХНУРЭ (г. Харьков),  
*А.И. ПОВОРОЗНЮК*, канд. техн. наук, НТУ «ХПИ» (г. Харьков)

## **АНАЛИЗ ИНФОРМАЦИОННОЙ ПОЛЕЗНОСТИ ПРИМЕНЕНИЯ АВТОМАТИЗИРОВАННЫХ ЛАБОРАТОРНЫХ СИСТЕМ В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ**

У статті математично доведена залежність інформаційної корисності функціонування лабораторних систем від типу діагностичної невпевненості. Наведена модель та результати застосування показника інформаційної цілі використання лабораторних систем для двох типів діагностичних ситуацій.

In article dependence of information utility of functioning of laboratory systems on type of diagnostic uncertainty is numerically shown. The model and results of using of a parameter of the information purpose of application of laboratory systems for two types of diagnostic situations is given.

**Постановка проблемы.** Автоматизированные лабораторные системы (АЛС) на сегодняшний день представляют собой сложные компьютерные робототехнические комплексы, предназначенные для проведения большого количества разнообразных лабораторных анализов биопроб с максимальной эффективностью и скоростью. При этом АЛС функционально представляют собой отдельные блоки или участки анализаторов (биохимические, иммунологические, гематологические и т.д.), соединенные лабораторной информационной системой (ЛИС). Автоматизация захватывает все этапы проведения лабораторных исследований: доизмерительную – т.е. пробоподготовительную, аналитическую – т.е. измерительно-вычислительную, а также постаналитическую [1]. В обязанности ЛИС может входить как диспетчирование, управление и контроль над ходом выполнения измерений, так и формирование отчетов, графиков, таблиц, а также преобразование и пересылка данных клиницистам [2]. При этом ЛИС может объединять аналитические блоки как одного производителя, так и разных. Такая сложная структура лабораторных комплексов делает проблему анализа эффективности их функционирования достаточно сложной и не однозначной. Анализ эффективности функционирования АЛС сводится к формированию показателей, с помощью которых можно было бы сравнивать и, следовательно, выбирать АЛС, максимально подходящую для данного конкретного лечебно-диагностического учреждения. Было бы не правильно утверждать, что какие-то современные АЛС не подходят к задачам лабораторных исследований. Речь идет о формировании показателей эффективности, способных выделить АЛС с избыточным количеством полезной информации, с целью снизить дополнительные расходы на ее получение. Для этого необходимо знать как информационные потребности лаборатории, так и информационные

возможности АЛС, планируемой к использованию в данных конкретных условиях.

**Анализ литературы.** Анализу информационной эффективности применения АЛС в конкретных лечебно-диагностических учреждениях посвящено значительно меньше научных работ, чем анализу экономической эффективности [3, 4]. Однако последнее время ведущие производители автоматизированного оборудования для медицинских лабораторий стали уделять внимание информационному маркетингу и предлагать АЛС в зависимости от конкретных потребностей потребителя, т.е. клиники или больницы [5, 6, 7]. В тоже время в большинстве случаев такое «подстраивание» под конкретные нужды оканчивается сравнением потенциальной производительности предлагаемых систем и имеющимся объемом работ. Исходя из этого, фирмы-производители АЛС условно делят весь рынок на маленькие, средние и большие АЛС. Однако при таком анализе остается не рассмотренным назначение проводимых лабораторных анализов, т.е. для каких целей проводится большое, среднее или маленькое количество анализов в данной конкретной лаборатории.

**Цель статьи.** Целью данной работы является анализ и оценка влияния клинико-диагностических условий на информационную эффективность применения АЛС.

**Раздел 1. Влияние диагностической неопределенности на модель информационного показателя.** Функционирование АЛС полностью подчинено процессу постановки диагноза (или системе принятия решения в клинической ситуации). В качестве количественной меры информационной полезности функционирования АЛС можно использовать разницу между информационным параметром цели функционирования АЛС  $L(E)$  и информационным параметром результата  $R(E)$  исследования концентрации некоторого вещества  $E$  в биопробе [8].

Информационный параметр результата  $R(E)$  исследования концентрации вещества  $E$  в биопробе полностью зависит от физико-химического метода проведения исследования концентрации вещества  $E$  и технической реализации всех фаз проведения теста и численно представляет собой количество информации, которое содержится в выходной величине АЛС об анализируемом веществе  $E$  [9]. Определение численного значения параметра  $R(E)$  представляет собой сложную инженерную задачу, т.к. требует анализа законов распределения вероятностей всех погрешностей и ошибок, возникающих в процессе трансформации полезной информации из формы биопробы в численную форму концентрации вещества  $E$ . Другими словами, расчет этого параметра должен производиться самим производителем АЛС или любого функционально законченного блока, из которого потом с помощью ЛИС собирается автоматизированная система.

Диагностическую значимость  $L(E)$  численно можно определить как количество информации, содержащееся в веществе  $E$  об известной группе диагнозов  $D$  [9]:

$$L(E) = - \sum_{x=1}^K \sum_{j=1}^K p(E_x | D_j) \cdot p(D_j) \cdot \ln \frac{p(E_x | D_j)}{\sum_{j=1}^K p(E_x | D_j) \cdot p(D_j)}, \quad (1)$$

где  $K$  – число диагнозов, которые необходимо различить, и соответственно число диапазонов, на которые необходимо разбить весь интервал значений концентрации вещества  $E$  в биопrobe ( $E_{min}, E_{max}$ );  $p(E_x | D_j)$  – вероятность того, что текущее значение концентрации вещества  $E_x \in (E_j, E_{j+1})$  при диагнозе  $D_j$ ;  $p(D_j)$  – априорное значение вероятности диагноза  $D_j$ .

В зависимости от типа клинической ситуации, для решения которой используется АЛС, разобьем все клинические условия применения АЛС на два вида: условия с первичной и вторичной диагностической неопределенностью. Системы принятия диагностического решения в условиях первичной неопределенности применяют в тех случаях, когда необходимо различить  $K$  диагнозов, например, при первичном обращении к врачу по поводу  $i$ -го диагноза из группы  $K$ , или при профилактическом осмотре. При этом полную группу случайных несовместных событий могут составлять такие типы состояний-диагнозов:  $D_1$  – патология 1;  $D_2$  – “серая” зона;  $D_3$  – патология 2. Процедуры постановки диагноза при вторичной неопределенности используются при подтвержденном первичном состоянии-диагнозе. В этом случае необходимо принять решение о степени выраженности или развития патологического процесса, о ходе лечения.

**Раздел 2. Информационный параметр цели применения АЛС в условиях первичной диагностической неопределенности.** Рассмотрим группу из 5 различных состояний:  $D_1$  – патология 1,  $D_2$  – “серая” зона 1,  $D_3$  – норма,  $D_4$  – “серая” зона 2,  $D_5$  – патология 2. Первичная диагностическая неопределенность с точки зрения типа плотности распределения вероятности может характеризоваться такими ситуациями: а) вероятность только одного патологического состояния значительно больше вероятности любого другого диагноза (см. рис.1, кривая 1, 3); б) разница между вероятностями любых соседних диагнозов не более 0,3 с предельным случаем равной вероятности каждого диагноза (см. рис.1, кривая 2, 4). Будем рассматривать только одномерные законы распределения вероятностей. Для случая а) вероятность одной из патологий (например  $D_1$ ) должна быть не менее 0,625. Для формирования  $p(D_j)$  в случае б) следует учесть центральную симметрию

закона и условие нормировки  $\sum_{j=1}^K p(D_j) = 1$ .

При вычислении диагностической значимости  $L(E)$  лабораторного параметра  $E$  по формуле (1) для  $j = 1, \dots, 5$  при первичной диагностической неопределенности клинической ситуации под вероятностью  $p(E_x|D_j)$  следует понимать вероятность того, что при наличии состояния  $D_j$  параметр  $E_x$  принадлежит диапазону  $[E_j, E_{j+1}]$ .

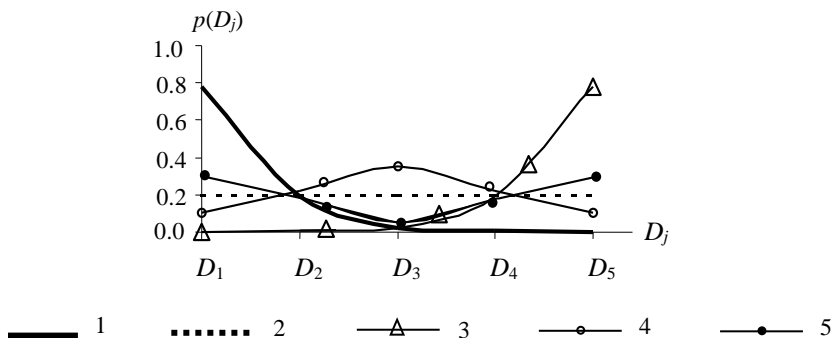


Рис. 1. Распространенные распределения вероятностей патологических состояний

Особое внимание в таком случае следует уделять выбору точки разделения [3], т.к. от определения порога концентрации, превышение которого считается достаточным основанием для оценки патологического состояния, зависят условные вероятности  $p(E_x|D_j)$ , а значит, и диагностическая значимость теста. Например, если задачей скрининга на диабет является выявление максимального количества людей с повышенной концентрацией глюкозы в венозной крови через час после стандартного завтрака, то при точке разделения 5,55 ммоль/л условная вероятность наличия концентрации глюкозы более 5,55 ммоль/л у пациентов с патологией будет составлять 0,48. В тоже время при точке разделения 8,88 ммоль/л эта вероятность будет равна 0,98 [3].

**Раздел 3. Информационный параметр цели применения АЛС в условиях вторичной диагностической неопределенности.** Вторичная неопределенность проявляется при подтвержденном диагнозе, поэтому, в качестве исходной величины при вычислении «параметра цели»  $L(E)$  выступают единицы выраженности  $g_i$  клинического состояния  $D_j$ . Вероятность степени выраженности  $p(g_i)$  определяется также как и  $p(D_j)$  при первичной неопределенности с той разницей, что в качестве диапазона  $[E_{\min}; E_{\max}]$  следует использовать диапазон  $[E_j; E_{j+1}]$  для данного состояния

$D_j$ . При дифференциальной диагностике или при прогнозировании течения заболевания у конкретного пациента на величину лабораторного параметра оказывает влияние внутрииндивидуальная вариация, т.е. для оценки результатов анализов используют данные о таких же анализах, но полученные у этого же пациента ранее. Параметры отдельного организма колеблются в значительно более узком диапазоне, чем в группе, т.е. внутрииндивидуальная вариация значительно меньше межиндивидуальной. При вторичной диагностической неопределенности на  $L(E)$  влияют и условия среды обитания, и физиологическое состояние пациента при измерениях, например, время года, суток, перемещения перед обследованием, стрессы, нагрузки, положение тела, диета, в том числе прием алкоголя и кофе, длительность голодания, курение, эндокринный и репродуктивный статус, в том числе менструации и беременность. Вместо точек разделения при этом выступают границы значащих интервалов, т.е. таких интервалов исследуемого лабораторного параметра, которые могут свидетельствовать о динамике параметра, вызванной не внутрииндивидуальной вариацией, а лечением или, наоборот, его отсутствием. Соответственно, в формуле (1) условную вероятность  $p(E_x|D_j)$  следует заменить на вероятность  $p(E_{j,x}|g_i)$  того, что значение параметра  $E$  принадлежит диапазону  $[E_{j,x}; E_{j,x+1}]$  при  $g_i$ -й степени состояния  $D_j$ , т.е.

$$L(E) = - \sum_{i=1}^G \sum_{x=1}^G p(g_i) \cdot p(E_{j,x}|g_i) \cdot \ln \frac{p(E_{j,x}|g_i)}{\sum_{i=1}^G p(g_i) \cdot p(E_{j,x}|g_i)}, \quad (2)$$

где  $G$  – число степеней выраженности лабораторного параметра  $E$ . Если интервалу дискретизации концентрации вещества  $E$ , зависящему от внутрииндивидуальной вариации, поставить в соответствие одну степень выраженности диагноза, то

$$G = \left| \frac{E_j - E_{j+1}}{E_{j,x} - E_{j,x+1}} \right|, \quad (3)$$

где  $E_{j,x}$  – текущее значение параметра  $E$  в  $j$ -м диапазоне. Из этого следует, что вторичная диагностическая неопределенность фактически разбивает физиологически допустимый интервал значений параметра  $E$  на  $G$  непересекающихся участков  $[E_{\min}; E_1]$ ,  $(E_1; E_2]$ , ...,  $(E_G; E_{\max}]$ , присущих тому или иному заранее известному диагнозу-состоянию, для каждого из которых имеется собственное значение информационного показателя эффективности АЛС. Для определения информационной эффективности функционирования АЛС в условиях вторичной диагностической неопределенности при исследовании  $E$  в диапазоне  $[E_{\min}; E_{\max}]$  необходимо

учитывать весовой вклад исследований для каждого диагноза  $D_j$  в общем числе исследований, проводимых данным учреждением по параметру  $E$ .

**Раздел 4. Численный эксперимент.** В качестве исследуемого вещества рассмотрим иммуноглобулин  $IgG1$ , интервал физиологически допустимой концентрации в сыворотке взрослого человека которого равен от 17 мг/литр до 25 г/литр или от  $6,9 \cdot 10^{16}$  до  $8,773 \cdot 10^{19}$  молекул/литр [10]. Рассмотрим ситуацию, когда первоначально необходимо разделить 3 диагноза:  $D_1$  – пониженный уровень  $IgG1$ ,  $D_2$  – нормальная концентрация и  $D_3$  – повышенный уровень  $IgG1$ .

Таблица 1

Параметры гауссовых распределений вероятностей концентрации  $IgG1$   $p(E/D_j)$  при диагнозах  $D_1, D_2, D_3$

Диагноз	$D_1$ , молекул/литр	$D_2$ , молекул/литр	$D_3$ , молекул/литр
Среднее значение концентрации	$1,114 \cdot 10^{19}$	$2,856 \cdot 10^{19}$	$5,747 \cdot 10^{19}$
Средне квадратичное отклонение	$3,6 \cdot 10^{18}$	$4,7 \cdot 10^{18}$	$4,1 \cdot 10^{18}$

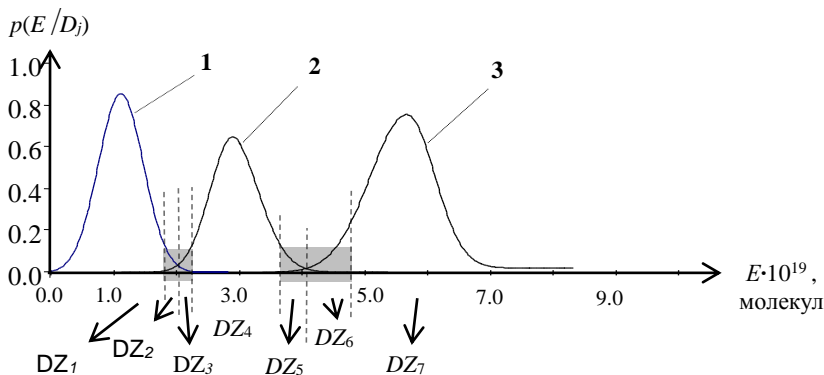


Рис. 2. Распределение условной вероятности концентрации  $IgG1$ :

1 – распределение  $p(E/D_1)$  при диагнозе  $D_1$ , 2 – распределение  $p(E/D_2)$  при диагнозе  $D_2$ , 3 – распределение  $p(E/D_3)$  при диагнозе  $D_3$

С учетом «серых» зон ( $DZ_2, DZ_3, DZ_5, DZ_6$  на рис. 2), весь диапазон значений  $IgG1$  следует разбить на 7 частей:  $DZ_1 - DZ_7$ . В качестве распределений априорных вероятностей  $p(D_j)$  того, что концентрация  $IgG1$

принадлежит одному из поддиапазонов  $DZ_1 - DZ_7$ , возьмем пять распределений, изображенных на рис.1. Результаты вычислений диагностической значимости иммуноглобулина  $IgG1$  по формуле (1) для каждого типа клинической ситуации приведены в табл. 2. Как и следовало ожидать, наибольшую диагностическую значимость значение концентрации  $IgG1$  имеет в клинической ситуации 3-го типа, когда априорные вероятности наличия концентрации в любом поддиапазоне равны, а наименьшую диагностическую значимость – в первой и пятой.

Таблица 2

Параметр цели функционирования ЛКС при различных типах клинической ситуации

Клиническая ситуация	1	2	3	4	5
Диагностическая значимость $L(E)$	0,3713	0,6772	1,0330	0,9514	0,3954

Для вычисления информационного параметра цели функционирования АЛС  $L(E)$  при вторичной диагностической неопределенности рассмотрим случай, когда основная часть лабораторных тестов осуществляется для пациентов с, например, повышенным содержанием иммуноглобулина  $IgG1$ .

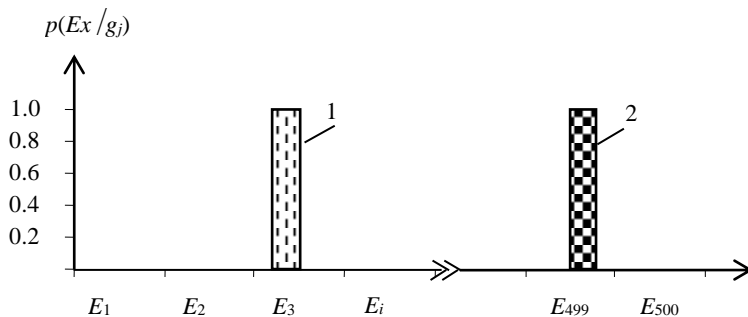


Рис. 3. Примеры распределений вероятности  $p(E_{j,x} | g_i)$  при вторичной диагностической неопределенности:

- 1 – распределение  $p(E_{j,x} | g_3)$  при 3-й выраженности клинического состояния,
- 2 – распределение  $p(E_{j,x} | g_{499})$  при 499-й выраженности клинического состояния

При этом необходимо различить степень влияния проводимых терапевтических процедур, т.е. на сколько изменяется содержание  $IgG1$  в диапазоне  $[4,6768 \cdot 10^{19}; 6,7268 \cdot 10^{19}]$ , если изменение концентрации на  $0,041 \cdot 10^{18}$  молекул/литр имеет диагностическое значение. Другими словами, необходимо различить 500 степеней выраженности патологического состояния  $p(g_i)$ , имеющих равные априорные вероятности появления. В данном случае

каждое значение  $g_i$  соответствует одному значению концентрации  $IgG1$ , а не интервалу значений, и шаг между соседними диагностически значимыми концентрациями  $IgG1$  совпадает с разрешающей способностью анализируемой АЛС ( $0,041 \cdot 10^{18}$  молекул/литр). Поэтому условная вероятность  $p(E_{j,x}|g_i)$  того, что концентрация иммуноглобулина равна  $E_x$  при  $g_i$ -й выраженности клинического состояния, имеет вид, изображенный на рис. 3. В случае, когда  $G = 500$  и априорная вероятность  $p(g_i) = 0,002$ , диагностическая значимость теста, рассчитанная по формуле (2), равна  $L(E) = 6,227$ . В случае, когда  $G = 100$  и  $p(g_i) = 0,01$ , получено  $L(E) = 4,651$ . Сравнение значений информационного параметра цели функционирования АЛС для первичной и вторичной диагностической ситуации показывает, что значения  $L(E)$  при вторичной диагностической неопределенности имеет более высокие значения (не менее, чем в 4 раза больше при любых априорных распределениях), чем в условиях первичной диагностической неопределенности.

**Выводы.** Анализ и сравнение информационной полезности применения одной и той же АЛС в различных лабораториях необходимо обязательно проводить с учетом статистически преобладающей диагностической неопределенности клинических ситуаций, возникающих в конкретном медицинском учреждении, а не только на основании статистики проведенных лабораторных исследований.

**Список литературы:** 1. *Wolman D.M., Kalfoglou A.L., LeRoy L.* Medicare Laboratory Payment Policy: Now and in the Future. – NATIONAL ACADEMY PRESS, Washington, DC, 2000. – P. 214. 2. *Staab T. A., Elling J. W.* ASTM E1989-98 – The New Instrument Control Standard // Journal of the Association for Laboratory Automation. – 1999. – Vol. 4. – № 3. – P. 40–42. 3. *Власов В.В.* Эффективность диагностических исследований. – М.: Медицина, 1988. – 256 с. 4. *Меньшиков В.В., Кадашева О.Г., Делекторская Л.Н. и др.* Методика диагностической оценки клинических лабораторных исследований // Лабораторное дело. – 1988. – № 6. – С. 67–70. 5. *Huscher Cl.* Olympus Laboratory Automation: The Visible Difference // Journal of the Association for Laboratory Automation. – 1999. – Vol. 4. – № 3. – P. 4–5. 6. *Okrongly D.* Taking testing to the next step // IVD Technology. – 2004. – P. 12–15. 7. *Shojania K. G., Duncan B. W., McDonald K. M.* Making Health Care Safer: A Critical Analysis of Patient Safety Practices. – AHRQ Publication. USA, 2001. – 662 p. 8. *Козина О.А.* Метод оценки эффективности лабораторных систем // Медицинская техника. – 2000. – № 1. – С. 43–45. 9. *Козина О.А.* Информационный показатель эффективности функционирования лабораторных компьютерных систем на примере нефелометрических систем: Автореф. дис. канд. тех. наук: 05.11.17 / Харьк. гос. технич. унив-т радиозлектроники. – Харьков, 2000. – 18 с. 10. *Малкина Л.А., Шаханина К.Л.* Уровень иммуноглобулинов (G, A и M) у здоровых взрослых людей // Журнал микробиологии, эпидемиологии и иммунобиологии. – 1975. – № 1. – С. 33–37.

*Поступила в редакцию 18.04.2005*