

**О.А. КОЗИНА**, канд. техн. наук, ХНУРЭ (г. Харьков)

## СИСТЕМА ОПТИМИЗАЦИИ ИНФОРМАЦИОННОЙ ЭФФЕКТИВНОСТИ КОМПЛЕКСОВ ЛАБОРАТОРНОЙ ДИАГНОСТИКИ

Розглянуто принципи застосування моделі інформаційного показника ефективності функціонування комплексів лабораторної діагностики. Аналізуються можливості програмної системи для проведення чисельного експерименту з оцінки діагностичних результатів лабораторних досліджень.

Principles of using of functioning efficient informational index for laboratory diagnostic systems are considered. Abilities of researched software system for numeric experiment of laboratory test results estimation are analyzed.

**Постановка проблемы.** Диагностические комплексы, которые используются в крупных лабораториях, больницах и клиниках во всем мире, часто являются зеркалом исторического развития медицинского учреждения. По этой причине комплексы лабораторной диагностики (КЛД) представляют собой связанные в единый комплекс блоки-модули по проведению специальных типов лабораторных исследований как одного производителя, так и разных. Оценить наиболее важную – медицинскую эффективность функционирования таких КЛД можно путем применения информационного подхода к анализу проблемы. Информационный анализ эффективности КЛД позволяет численно сопоставить результаты лабораторных исследований, проведенных с помощью анализируемого КЛД, с информационной потребностью в результатах такого лабораторного исследования для процесса диагностики у конкретного пациента. При этом информационный результат КЛД определяется только структурой самого КЛД. В то же время информационная потребность в результатах лабораторного исследования определяется как предрасположенностью популяции, которую обслуживает данная клиника, к тому или иному заболеванию, так и теми диагностическими стратегиями, которые практикуют врачи-клиницисты для данной популяции. Поэтому информационная ценность лабораторного исследования, т.е. количество информации, содержащееся в результате лабораторного исследования, которое необходимо для различения клинических состояний у пациента, определяет полезность применения данного КЛД в конкретном лечебном учреждении.

**Анализ литературы.** Проблемам вычислительной диагностики посвящено достаточное количество работ за последние годы [1 – 3]. Этому способствует также популяризация «медицины, основанной на доказательствах». Однако в медицинской научной литературе рассматриваются в основном возможности применения характеристических кривых для разделения популяции на больных и здоровых по критической точке разделения [2, 4]. В технической литературе рассматриваются вопросы создания тех или иных блоков КЛД без учета информационной эффективности разработанных модулей для диагностических потребностей того лечебного учреждения, в котором КЛД будет эксплуатироваться [3, 5]. В известных немногочисленных работах [6 – 8], посвященных сопоставлению информационных потребностей клиницистов в лабораторных анализах с информационными результатами, которые дают КЛД, рассмотрены лишь несколько типов распределений вероятности обнаружения больных для одного типа исследования биопробы.

**Цель статьи.** Создание программной системы, которая численно определяет: на сколько влияют параметры различных распределений вероятности на информационный результат функционирования КЛД.

**Исходные данные для построения системы моделирования.** Сравнение диагностической значимости лабораторного исследования  $L(E)$  с информационным параметром результата  $R(E)$  исследования концентрации вещества  $E$  в биопробе, т.е. с количеством информации, которое содержится в выходной величине КЛД об анализируемом веществе  $E$ , дает объективную численную характеристику целесообразности использования данного КЛД в заданных клинических условиях. В предложенной системе величина параметра результата функционирования КЛД может быть либо задана на основе ранее полученных значений, либо рассчитана. Для расчета необходимы данные о величинах вероятности ошибок или безотказной работы всех функциональных блоков КЛД, которые принимают участие в движении информационных потоков согласно анализируемому лабораторному исследованию. В общем случае это такие блоки как блок пробоподготовки, транспортировки биоматериалов, измерительный и вычислительный блоки, а также модуль управления, ввода-вывода и передачи данных.

Вычисление диагностической значимости лабораторного теста  $L(E)$  проводится в соответствии с математической моделью информационных целей функционирования КЛД при условии первичной диагностической неопределенности. Количество информации, которое содержится в веществе  $E$  относительно известной группы диагнозов  $D$ , численно определяется как [6]:

$$L(E) = - \sum_{x=1}^K \sum_{j=1}^K p_x | D_j | p_j \ln \frac{p_x | D_j | p_j}{\sum_{j=1}^K p_x | D_j | p_j} \quad (1)$$

где  $K$  – количество клинических состояний, которые необходимо различить и соответственно количество диапазонов, на которые необходимо разбить весь интервал значений концентрации вещества  $E$  в биопробе ( $E_{\min}, E_{\max}$ );  $p(E_X/D_j)$  – вероятность того, что текущее значение концентрации вещества  $E_X \in (E_j, E_{j+1})$  при диагнозе-состоянии  $D_j$ ;  $p(D_j)$  – априорное значение вероятности диагноза  $D_j$ . В отличие от большинства характеристических кривых, построенных для случая разделения популяции на больных и здоровых, т.е. только на два различных клинических состояния, в предлагаемой системе можно выбрать любое другое количество различных клинических состояний. Это может быть и двухсторонняя дифференциация состояний: пониженный уровень концентрации вещества в биожидкости, нормальный или референтный уровень для данной популяции и повышенный уровень концентрации. В предлагаемой системе можно добавить и степени выраженности отклонений от референтных интервалов концентраций в качестве отдельных диагностических состояний и тогда  $K$  может быть равно 4 или 5 (например, были выбраны такие состояния: повышенный уровень концентрации, нормальный, слабо пониженный и сильно пониженный). Выбор количества диагнозов-состояний, которые могут быть дифференцированы на результатах данного лабораторного теста и, следовательно, точек разделения представляет собой самую спорную часть всего процесса диагностики. С одной стороны, смещение точки разделения диагнозов влечет за собой изменения в распределении плотности вероятности  $p(E_X/D_j)$ . Поэтому лучше всего применять параметры распределения вероятности, которые были эмпирически получены в результате статистической обработки данных для той популяции, которую обслуживает данное лечебное учреждение. При отсутствии такой статистики можно воспользоваться теоретически возможными параметрами распределений [2]. С другой стороны, положение дискриминантной точки на оси концентраций зависит от преваленса каждого диагноза у данной популяции. Это означает, что если для выбранной популяции существует предрасположенность к какому-либо клиническому состоянию, т.е. преваленс  $p(D_j)$  одного из диагнозов значительно выше уровня остальных априорных значений вероятностей диагнозов, то точка разделения таких состояний должна быть сдвинута в сторону увеличения чувствительности лабораторного теста. Следует также отметить, что как распределение  $p(E_X/D_j)$ , так и распределение  $p(D_j)$  не являются статическими даже для одной и той же популяции. Под воздействием быстро изменяющихся климатических условий и состояний экологии эти распределения вероятностей могут изменяться и поэтому должны периодически проверяться. Предлагаемая система оптимизации позволяет быстро и объективно оценить, например, как изменится полезность использования данного КЛД, если точка разделения выбранных диагнозов-состояний будет смещаться в сторону уменьшения специфичности лабораторного теста на протяжении интервала концентраций  $[E_{\text{реф}}^{\min}, E_{\text{реф}}^{\max}]$  или если преваленс одного из диагнозов увеличится.

**Анализ результатов моделирования.** Численный эксперимент проводился на интервале концентраций иммуноглобулина IgG1 в сыворотке от 0,5 г/литр до 25 г/литр для трех состояний: пониженный уровень концентрации IgG1 – состояние  $D_1$ , нормальный уровень – состояние  $D_2$  и повышенный уровень – состояние  $D_3$ . Априорные вероятности наличия каждого из трех состояний рассматривались для различных клинических задач (см. табл. 1).

Таблица 1

Преваленс состояний

Тип клинической задачи	Априорная вероятность состояния		
	$D_1$	$D_2$	$D_3$
1 – предрасположенность к состоянию $D_1$ , когда понижен уровень иммуноглобулинов	0.68	0.28	0.04
2 – скрининг	0.33	0.34	0.33
3 – предрасположенность к состоянию $D_3$ , когда повышен уровень иммуноглобулинов	0.04	0.28	0.68

Для вычисления диагностической значимости лабораторного теста  $L(E)$  использовалось выражение:

$$L(E) = H(E) + H(D) - H(E, D), \quad (2)$$

где  $H(E)$  – безусловная энтропия концентраций иммуноглобулина;  $H(D)$  – безусловная энтропия наличия состояний  $D_j$ ;  $H(E, D)$  – совместная энтропия.

Распределение вероятности  $p(E, D)$  концентрации  $E_X$  при состоянии  $D_j$  описано нормальным законом со средним значением концентрации  $m$  и средним квадратическим отклонением  $d$  (см. табл. 2).

Таблица 2

№ эксперимента	Распределение вероятности $p(E)$						Распределение вероятности $p(E, D_j)$					
	Клиническая задача 1		Клиническая задача 2		Клиническая задача 3		Состояние $D_1$		Состояние $D_2$		Состояние $D_3$	
	$m$	$d$	min	max	$m$	$D$	$m$	$d$	$m$	$d$	$m$	$d$
1	11	6	5	30	18	8	8	2.5	13	4	26	4

2	11	6	5	30	18	8	15	1.5	18	4	21	2.5
3	11	4	5	30	20	3	15	1.5	18	4	21	2.5

Распределение априорной вероятности  $p(E)$  уровня концентрации  $E_X$  для клинической задачи 1 и 3 описано нормальным законом, а для клинической задачи 2, т.е. для скрининга, равномерным (см. табл. 2). Графически динамика изменений параметров распределений представлена на рис. 1.

Как видно по результатам эксперимента (см. табл. 3), для разделения клинических состояний при априорно равномерной вероятности наличия любого клинического состояния, т.е. при скрининге, необходимо больше информации от КЛД, чем при других клинических задачах.

Таблица 3

Информационная цель  $L(E)$

№ эксперимента	Тип клинической задачи		
	Априорно пониженный уровень концентрации	Скрининг	Априорно повышенный уровень концентрации
1	28.9059	38.7412	34.4544
2	30.2858	38.5173	32.5502
3	29.8871	38.6303	33.0863

Сравнение результатов эксперимента 2 и 3 показывает, что чем больше площадь области наложения априорных распределений вероятности концентраций  $p_1(E)$  и  $p_2(E)$  при прочих равных параметрах вероятностей, тем больше информации должно требоваться от КЛД. Сравнение значений  $L(E)$  при априорной предрасположенности популяции к пониженному содержанию иммуноглобулина и априорно повышенному уровню концентраций для эксперимента 1 и 2 показывает, что увеличение количества информации от КЛД требуется при увеличении разницы между математическими ожиданиями распределений  $p(E)$  и  $p(E, D)$ .

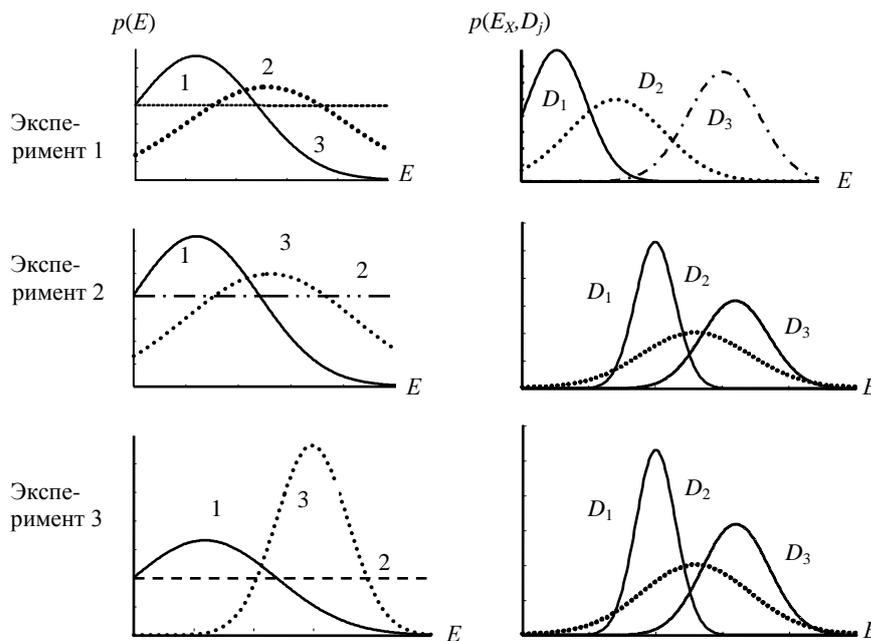


Рис. Динамика изменений параметров распределений

Результаты моделирования для случая разделения только двух клинических состояний: пониженный и нормальный уровень концентрации иммуноглобулина при таких же параметрах безусловных вероятностей и таких же трех клинических задачах представлены в табл. 4.

Таблица 4

№ эксперимента	Параметры нормального распределения $p(E)$	Точка разделения состояний	Информационная цель $L(E)$		
			Предрасположенность к пониженному уровню $D_1$	Равновоятные состояния $D_2$	Предрасположенность к норм. уровню $D_3$
1	$m1 = 9$ $m2 = 16$ $d1 = 2$ $d2 = 2$	11.53	31.1347	42.4566	37.5749
2	$m1 = 8$ $m2 = 17$ $d1 = 2$ $d2 = 3$	11.83	31.2694	42.0733	37.4183
3	$m1 = 9$ $m2 = 16$ $d1 = 2$ $d2 = 3,5$	12.09	30.5669	41.8348	37.1682

4	$m1 = 9$ $m2 = 16$	$d1 = 2,5$ $d2 = 3,5$	12.33	30.5814	41.8289	37.1199
5	$m1 = 9$ $m2 = 16$	$d1 = 4$ $d2 = 5$	12.50	30.9452	41.8246	37.1141

**Выводы.** Использование системы оценки информационной эффективности КЛД позволяет выбрать оптимальное соответствие между реальными потребностями конкретного лечебного учреждения в результатах теста и функциональными возможностями конкретного КЛД. Разработанная система позволит быстро произвести вычисления информационных целей данного теста при изменении клинической ситуации. Для расширения возможностей по оптимизации информационной эффективности КЛД в систему моделирования необходимо добавить дополнительные методики проведения лабораторных тестов для определения информационных показателей результата функционирования КЛД.

**Список литературы:** 1. *Okrongly D.* Taking testing to the next step // IVD Technology. – 2004. – January/February. – P. 12 – 15. 2. *Власов В.В.* Эффективность диагностических исследований. – М.: Медицина, 1988. – 256 с. 3. *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. 4. *Ockene I.S., Matthews C.E.* Variability and Classification Accuracy of Serial High-Sensitivity C-Reactive Protein Measurements in Healthy Adults // Clinical Chemistry. – 2001. – № 47. – P. 444 – 450. 5. *Huscher Cl.* Olympus Laboratory Automation: The Visible Difference // Journal of the Association for Laboratory Automation. – 1999. – Vol. 4. – № 3. – P. 4 – 5. 6. *Козина О.А.* Метод оценки эффективности лабораторных систем // Медицинская техника. – 2000. – № 1. – С. 43 – 45. 7. *Козина О.А., Поворозюк А.И.* Анализ информационной полезности применения автоматизированных систем в лабораторной диагностике // Вісник НТУ "ХПІ". – Харків: НТУ "ХПІ". – 2005. – № 46. – С. 100 – 108. 8. *Lamb DA, Lopinski R, Sun DH, Janowiak D.* Operational effects of total laboratory automation // Clinical Leadersh Managment Review. – 2000. – № 14. – P. 173 – 177.

Поступила в редакцию 20.04.2006