

В.Н. Алдобаев¹, А.А. Масликов², Дядищева В.П.¹, Л.А. Еременко¹, А.А. Мазанова¹,

Расчет характеристик токсикологических зависимостей «доза-эффект» на основе моделирования функций плотности распределения регрессионных коэффициентов.

¹*ФГБУН «Научно-исследовательский центр токсикологии и гигиенической регламентации биопрепаратов Федерального медико-биологического агентства» (ФГБУН НИЦ ТБП ФМБА России), Серпухов, Московская обл.*

²*Международный Университет природы, общества и человека «Дубна», филиал «Протвино», Московская область*

Введение

Побудительным моментом данного исследования явилось отсутствие на сегодняшний день в научной литературе по токсикологии единого строгого подхода к описанию зависимостей «доза-эффект» и вычислению точечных оценок доз (концентраций) в совокупности с соответствующими доверительными интервалами. Это в свою очередь нашло отражение в отсутствии единого согласованного алгоритма обчета соответствующих данных в различных распространяемых среди исследователей программных продуктах.

Авторами была предпринята попытка, используя современные итерационные методы поиска экстремума функции 2-х переменных, путем построения соответствующих функций распределения, определить характерные дозы и их доверительных интервалы посредством точного вычисления математических ожиданий и граничных точек для означенных функций. Для расчетов был использован мощный программный пакет *Mathematica 9*, Wolfram Research Inc. На сегодняшний день эта компания считается одним из наиболее авторитетных производителей программного обеспечения в мире.

Проанализировав ряд коммерческих программных продуктов, содержащих в своем составе модули *probit-regression*, с точки зрения их расчетных возможностей, удобства использования и стоимости, авторы попытались максимально учесть потребности токсикологов и фармакологов, занимающихся поиском новых лекарственных форм.

Материалы и методы

Считается, что идея выравнивания кривой доза-эффект и преобразования ее в прямую линию принадлежит [Gaddum J.H., Bliss C.I., в 1933-34 [1,2]. Позже Bliss C.I. предложил использовать смещенные на 5 единиц по оси ординат нормированные отклонения эффектов. Показатель, получаемый в результате смещения нормированных отклонений по оси ординат, получил название «пробит».

В пробит анализе, при определении характерных эффективных (летальных) доз и доверительных интервалов для них, постулируется линейная зависимость между пробитом и логарифмом дозы.

$$Y = a + bX \quad (1.1)$$

Здесь $X = \lg(D)$, и пробит Y связан с вероятностью P воздействия дозы через формулу:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y-5} \exp(-z^2 / 2) dz \quad (1.2)$$

При этом дисперсия пробита Y не обязана обладать свойством гомогенности (что и наблюдается экспериментально), а следовательно подход стандартного регрессионного анализа для оценки параметров a и b нуждается в модификации. Стандартным способом оценивания параметров известного распределения является метод максимального правдоподобия (Maximum Likelihood), который состоит в максимизации функции правдоподобия [3]:

$$F_{ML}(a, b) = \prod_{i=1}^N \frac{n_i!}{r_i! n_i - r_i!} P_i^{r_i} (1 - P_i)^{n_i - r_i} \quad (1.3)$$

или ее логарифма (в силу монотонности логарифма):

$$L(a, b) = \ln F_{ML} = \sum_{i=1}^N r_i \ln P_i + (n_i - r_i) \ln(1 - P_i) + \text{const} \quad (1.4)$$

Здесь n_i и r_i - соответственно число объектов и число объектов на которых проявился эффект в i -ой группе, P_i - ожидаемая вероятность эффекта, суммирование ведется по N группам, а константа (не-существенная при нахождении максимума) $\text{const} = \sum_{i=1}^N \ln \frac{n_i!}{r_i! n_i - r_i!}$. Необходимым условием

максимума функции L является выполнение условий:

$$\begin{cases} \partial_a L = \frac{\partial L}{\partial a} = 0 \\ \partial_b L = \frac{\partial L}{\partial b} = 0 \end{cases}$$

что в нашем случае приводит к системе нелинейных уравнений для параметров a и b :

$$\begin{cases} \sum_{i=1}^N \frac{n_i(r_i/n_i - P_i)}{P_i(1 - P_i)} Z_i = 0 \\ \sum_{i=1}^N \frac{n_i(r_i/n_i - P_i)}{P_i(1 - P_i)} Z_i x_i = 0 \end{cases} \quad (1.5)$$

здесь Z_i - значение производной ожидаемой вероятности по пробиту (или по параметру a) при логарифмической дозе x_i , ср. (1.2):

$$Z_i = \frac{1}{\sqrt{2\pi}} \exp(-(Y_i - 5)^2 / 2) \quad (1.6)$$

Далее, для анализа и численного решения системы (1.5), можно ее линеаризовать относительно поправок к некоторому начальному приближению: $a = a_0 + \delta a$ и $b = b_0 + \delta b$, путем разложения логарифмической функции правдоподобия L в ряд Тейлора (при этом оставить только линейные по $\delta a, \delta b$ члены) $L = L(a_0, b_0) + \partial_a L(a_0, b_0) \cdot \delta a + \partial_b L(a_0, b_0) \cdot \delta b$ [4]. Тогда, если после дифференцирования пренебречь членами высшего порядка малости, пропорциональными $(r_i / n_i - P_i)$ и учесть, что $\partial_a P = Z$ и $\partial_b P = xZ$, получится система линейных уравнений на $\delta a, \delta b$:

$$\begin{cases} \sum_{i=1}^N \frac{n_i(r_i / n_i - P_i)}{P_i(1 - P_i)} Z_i - \delta a \sum_{i=1}^N \frac{n_i Z_i^2}{P_i(1 - P_i)} - \delta b \sum_{i=1}^N \frac{n_i Z_i^2 x_i}{P_i(1 - P_i)} = 0 \\ \sum_{i=1}^N \frac{n_i(r_i / n_i - P_i)}{P_i(1 - P_i)} Z_i x_i - \delta a \sum_{i=1}^N \frac{n_i Z_i^2 x_i}{P_i(1 - P_i)} - \delta b \sum_{i=1}^N \frac{n_i Z_i^2 x_i^2}{P_i(1 - P_i)} = 0 \end{cases} \quad (1.7)$$

Здесь все значения P_i и Z_i берутся в точке начального приближения a_0, b_0 , повторяющиеся комбинации

$w_i = \frac{n_i Z_i^2}{P_i(1 - P_i)}$ имеют смысл весовых коэффициентов при суммированиях. Тогда, если величину $\delta y_i = \frac{(r_i / n_i) - P_i}{Z_i}$ считать поправкой к пробиту (т.е. определить так называемый *рабочий*

пробит как $y = Y + (p - P) / Z$, где p – наблюдаемая частота), и разделить на сумму весовых коэффициентов, систему (1.7) можно привести к виду:

$$\begin{cases} \delta a + \delta b \cdot \bar{x} = \overline{\delta y} \\ \delta a \cdot \bar{x} + \delta b \cdot \overline{x^2} = \overline{x \cdot \delta y} \end{cases} \quad (1.8)$$

Здесь все средние понимаются как взвешенные:

$$\bar{A} = \frac{\sum_{i=1}^N w_i \cdot A_i}{\sum_{i=1}^N w_i} \quad (1.9)$$

Система уравнений (1.8) имеет структуру хорошо известных уравнений для коэффициентов взвешенной линейной регрессии. Однако, по смыслу, ее решение:

$$\delta b = \frac{\overline{x \cdot \delta y} - \bar{x} \cdot \overline{\delta y}}{\overline{x^2} - \bar{x}^2} \quad (1.10)$$

$$\delta a = \overline{\delta y} - \delta b \cdot \bar{x}$$

дает не значения коэффициентов регрессии, а поправки к ним в первом приближении.

Таким образом, формулы (1.10) в контексте пробит-анализа описывают шаг итерационной процедуры, задающей последовательность оценок параметров a_k и b_k , сходящуюся к точечным оценкам a и b , определяемым методом максимального правдоподобия. Стартовые значения a_0 и b_0 можно вычислять либо по обычным формулам регрессионного анализа, либо с учетом вышеопределенных весовых коэффициентов w . При этом следует отдавать себе отчет, что это будут только исходные приближения, к которым следует применять итерационную процедуру (1.10).

Мы изложили один из возможных итерационных алгоритмов нахождения точки максимума функции правдоподобия применительно к задаче пробит-анализа, который был предложен ещё (Fisher 1922 [3]; Bliss и Fisher 1935 [5]) и впоследствии исследовался на сходимость в разных модификациях (Garwood 1940 [6]).

Однако, при современном уровне развития компьютерной техники и программного обеспечения, вполне возможно работать непосредственно с логарифмической функцией правдоподобия $L(a,b)$ без использования ее приближенного разложения. Современные пакеты программ способны численно находить экстремум функции 2-х переменных за разумное время. Для этого мы использовали широко известный программный пакет *Mathematica 9*.

После получения точечных оценок параметров a и b , становится возможным точечное оценивание любых характерных эффективных (летальных) доз.

$$D_p = 10^{[5+\Phi^{-1}(p-0.5)-a]/b} \quad (1.11)$$

Здесь p – ожидаемая вероятность наступления эффекта, а Φ^{-1} – функция обратная к функции Лапласа.

Далее встает задача получения доверительных интервалов для точечных оценок. Для этого необходимо найти дисперсии $D(a)$ и $D(b)$ коэффициентов. Для точечных оценок, полученных методом максимального правдоподобия, в случае больших выборок и нормального распределения коэффициентов дисперсии определяются квадратичной формой вторых дифференциалов логарифмической функции правдоподобия взятой в точке оценок по уравнениям [3]:

$$\begin{cases} \partial_a^2 L(a,b) = -\frac{1}{1-r^2(a,b)} \cdot \frac{1}{D(a)} \\ \partial_b^2 L(a,b) = -\frac{1}{1-r^2(a,b)} \cdot \frac{1}{D(b)} \\ \partial_a \partial_b L(a,b) = +\frac{1}{1-r^2(a,b)} \cdot \frac{r^2(a,b)}{\sqrt{D(a) \cdot D(b)}} \end{cases} \quad (1.12)$$

Здесь коэффициент корреляции:

$$r(a,b) = \frac{\text{cov}(a,b)}{\sqrt{D(a) \cdot D(b)}}$$

Решение этой системы определяет элементы дисперсионно-ковариационной матрицы:

$$\begin{aligned} D(a) &= \frac{\partial_b^2 L}{(\partial_a \partial_b L)^2 - \partial_a^2 L \cdot \partial_b^2 L} \\ D(b) &= \frac{\partial_a^2 L}{(\partial_a \partial_b L)^2 - \partial_a^2 L \cdot \partial_b^2 L} \\ \text{cov}(a,b) &= -\frac{\partial_a \partial_b L}{(\partial_a \partial_b L)^2 - \partial_a^2 L \cdot \partial_b^2 L} \end{aligned} \quad (1.13)$$

Причем:

$$K = \begin{pmatrix} D(a) & \text{cov}(a,b) \\ \text{cov}(a,b) & D(b) \end{pmatrix} = - \begin{pmatrix} \partial_a^2 L & \partial_a \partial_b L \\ \partial_a \partial_b L & \partial_b^2 L \end{pmatrix}^{-1} \quad (1.14)$$

При заданной логарифмической дозе x_p дисперсия среднего значения пробита определяется формулой:

$$D(Y) = D(a) + 2x_p \operatorname{cov}(a, b) + x_p^2 D(b) \quad (1.15)$$

Далее, не составляет труда стандартным способом построить доверительную область для пробита с требуемой надежностью (уровнем доверия). Однако, обратная задача – построение доверительного интервала для (логарифмической) дозы при фиксированном пробите, требует более аккуратного подхода.

В начальном приближении можно считать, что дисперсия логарифмической дозы будет выглядеть следующим образом:

$$D(X) = \frac{D(Y)}{b^2} = \frac{1}{b^2} [D(a) + 2x_p \operatorname{cov}(a, b) + x_p^2 D(b)] = \frac{\partial_b^2 L - 2x_p (\partial_a \partial_b L) + x_p^2 (\partial_a^2 L)}{b^2 [(\partial_a \partial_b L)^2 - \partial_a^2 L \cdot \partial_b^2 L]} \quad (1.16)$$

Тогда нетрудно записать грубо приближенный доверительный интервал для логарифмической дозы на желаемом уровне значимости α :

$$[X(p) - t(\alpha) \sqrt{D(X)}; X(p) + t(\alpha) \sqrt{D(X)}] \quad (1.17)$$

Поскольку формула (1.16) получена в предположении нормальности распределения коэффициентов и больших выборок, то следует ожидать, что интервал задаваемый выражением (1.17) будет зауженным. Чтобы исправить ситуацию нужно корректно учесть эффективное число степеней свободы. Следует брать критическую точку $t(\alpha)$ из распределения Стьюдента с $(N-2)$ степенями свободы и учесть, что формулы (1.15-1.16) дают заниженную точечную оценку дисперсии, т.е. в формуле (1.17) подразумевается $D(X) = N \cdot D(X) / (N-2)$ (здесь N количество экспериментальных групп).

Далее, точечную оценку логарифмической дозы и края доверительного интервала можно пересчитать в обычную дозу, для этого нужно выполнить потенцирование по основанию 10. Разумеется, доверительный интервал при этом перестанет быть симметричным.

Существуют и другие подходы к построению приближенных доверительных интервалов для характерной дозы. Например, в соответствующем модуле программы SPSS v.15-22 используется приближенная формула, следующая из теоремы Фиеллера (Fieller) [7-9]:

$$x_{L,U} = x + \frac{g}{1-g} \left(x + \frac{\text{cov}(a,b)}{D(b)} \right) \pm \sqrt{\left[D(a) + 2x \text{cov}(a,b) + x^2 D(b) - g \left(D(a) - \frac{\text{cov}^2(a,b)}{D(b)} \right) \right]} \cdot h, \text{ где} \quad (1.18)$$

$x = (y - a)/b$ – значение логарифмической дозы, предсказанное по пробиту, h – фактор гетерогенности связанный со статистикой Пирсона хи-квадрат:

$$h = \frac{\chi^2}{N-2} = \frac{1}{N-2} \sum_{i=1}^N \frac{(r_i - n_i \cdot P_i)^2}{n_i P_i (1 - P_i)} \quad (1.19)$$

фактор $g = ht^2 D(b)/b^2$, где t – критическая точка статистики Стьюдента с числом степеней свободы $N - 2$. Причем, если выполняется критерий Пирсона, то фактор h обращается в единицу, а критическая точка t определяется нормальным распределением.

Если требуется ещё более аккуратная оценка доверительного интервала для дозы равно как и самой дозы, то следует вспомнить, что логарифмическая доза связана с пробитом соотношением: $x = (y - a)/b$. Поскольку случайные величины, построенные из коэффициентов регрессии как отношения: $t_a = (a - M(a)) / S(a)$, $t_b = (b - M(b)) / S(b)$ подчиняются t -статистике Стьюдента (здесь M и S – точечные оценки математического ожидания и среднеквадратического отклонения

коэффициентов), то числитель и знаменатель логарифмической дозы x можно аппроксимировать случайными величинами, распределенными по смещенному и растянутому t -распределению Стьюдента. Причем эти случайные величины, вообще говоря, обладают ненулевой ковариацией (1.13). Следовательно, для корректного точечного и интервального оценивания этого отношения нужно построить соответствующую функцию плотности распределения и найти критические точки. Это удобнее делать в системе координат, где $\text{cov}(a,b) = 0$, что достигается сдвижкой по оси x : $x \rightarrow x - \overline{x_w}$, где

$$\overline{x_w} = -\frac{\text{cov}(a,b)}{D(b)} \quad (1.20)$$

В новых координатах точечная оценка (и дисперсия) коэффициента a , конечно, изменятся: $a \rightarrow a^* = a + b \cdot \overline{x_w}$,

$$D(a^*) = D(a) + 2\overline{x_w} \text{cov}(a,b) + \overline{x_w}^2 D(b) \quad (1.21)$$

а для b эти параметры распределения останутся неизменными.

Таким образом, при фиксированном пробите y (соответствующем интересующей вероятности) мы имеем две независимые случайные величины u и v с t -распределениями Стьюдента:

$$f_1(u) = \frac{\Gamma((n_f + 1) / 2)}{\Gamma(n_f / 2) \cdot S(b) \cdot \sqrt{n_f \pi}} \left[1 + \frac{((u - b) / S(b))^2}{n_f} \right]^{-\frac{(n_f + 1)}{2}}, \text{ где} \quad (1.22)$$

$$f_2(v) = \frac{\Gamma((n_f + 1) / 2)}{\Gamma(n_f / 2) \cdot S(a^*) \cdot \sqrt{n_f \pi}} \left[1 + \frac{((v - (y - a^*)) / S(a^*))^2}{n_f} \right]^{-\frac{(n_f + 1)}{2}}$$

$\Gamma()$ – известная гамма-функция Эйлера, $n_f = N - 2$ – эффективное число степеней свободы,

$S(b) = \sqrt{h \cdot D(b)}$ и $S(a^*) = \sqrt{h \cdot D(a^*)}$ – точечные оценки среднеквадратических отклонений

коэффициентов регрессии в сдвинутой системе координат, поправленные на фактор гетерогенности h (1.19).

Далее нас интересует математическое ожидание отношения $t = v/u$ и доверительный интервал для него. Интегральная функция распределения отношения имеет вид:

$$F(t) = \int_{-\infty}^0 f_1(u) \left[\int_{tu}^{+\infty} f_2(v) dv \right] du + \int_0^{+\infty} f_1(u) \left[\int_{-\infty}^{tu} f_2(v) dv \right] du \quad (1.23)$$

Соответственно, функция плотности распределения отношения будет иметь вид:

$$f(t) = F'(t) = - \int_{-\infty}^0 f_1(u) f_2(tu) u du + \int_0^{+\infty} f_1(u) f_2(tu) u du \quad (1.24)$$

Теперь в рамках программного пакета *Mathematica* мы можем стандартным образом найти математическое ожидание отношения $t = v/u$, которое у нас по смыслу является логарифмической дозой:

$$M(t) = \int_{-\infty}^{+\infty} t \cdot f(t) dt, \quad (1.25)$$

а также найти границы доверительного интервала $[t_1, t_2]$ для отношения при уровне значимости α , численно решив уравнения:

$$F(t_1) = \alpha/2, \quad F(t_2) = 1 - (\alpha/2). \quad (1.26)$$

Кроме строгих точечных оценок эффективных доз и границ их доверительных интервалов, на наш взгляд в практической токсикологии существует проблема корректной оценки кумуляции (привыкания) при хроническом воздействии потенциальных токсикантов наряду с проблемой корректного сравнения эффективности (токсичности) различных субстанций, к примеру, при отборе кандидатов в фармацевтические препараты. На уровне математических подходов эти проблемы едины.

Применительно к первому вопросу, используя подходы (1.23 -1.25), мы можем получить оценки характерных эффективных доз (логарифмов) при множественном и однократном введении, а затем корректно сравнить соответствующие дозы.

Фактически, к этому сводится наиболее распространенная в токсикологии оценка коэффициента кумуляции

$$K_{cum} = \frac{ED_{50}^n}{ED_{50}^1}, \quad (1.27)$$

представляющая собой отношение полуэффективных (полулетальных) доз при множественном и однократном введении.

Если построены функции плотности распределения логарифмов доз при множественном и однократном введении $f_{(n)}(x)$ и $f_{(1)}(x)$ для характерных (например полуэффективных доз) – $x_{(n)} = \lg D_{(n)}$ и $x_{(1)} = \lg D_{(1)}$, то можно вычислить функцию плотности распределения их разности, как соответствующую свертку:

$$f_{(n)-(1)}(r) = \int_{-\infty}^{+\infty} f_{(1)}(x-r) \cdot f_{(n)}(x) dx \quad (1.28)$$

В нашем случае нужно только учесть, что функции $f_{(n)}(x)$ и $f_{(1)}(x)$ получаются в сдвинутых (по-разному) координатах (1.20).

После этого нетрудно посчитать, например, вероятность того, что $D_{(n)} > D_{(1)}$:

$$P(D_{(n)} > D_{(1)}) = \int_0^{+\infty} f_{(n)-(1)}(r) dr \quad (1.29)$$

Далее, для вывода о статистической достоверности нужно использовать либо односторонний критерий (I случай), либо двусторонний критерий (II случай), в зависимости от того возможно ли в конкретной ситуации обратное неравенство или нет. Если, например, мы просто сравниваем две дозы на уровне значимости α (II случай), то посчитанную $P(D_{(n)} > D_{(1)})$ мы должны сравнить с

$\alpha / 2$, и если вероятность окажется меньше, то мы на уровне значимости α должны принять альтернативную гипотезу ($D_{(n)} < D_{(1)}$). В противном случае у нас нет оснований считать, что ($D_{(n)} < D_{(1)}$). Тогда нужно проделать аналогичное сравнение для левой стороны критерия, и либо принять гипотезу ($D_{(n)} > D_{(1)}$) , либо заключить, что $D_{(n)}$ и $D_{(1)}$ неразличимы на уровне значимости α .

В I-ом случае выполняется аналогичное сравнение вероятности с α , но только для одной стороны критерия.

При установлении статистической значимости коэффициента кумуляции (1.27) сравнение соответствующих $D_{(n)}$ и $D_{(1)}$ целесообразно проводить с помощью однохвостового критерия в том случае, когда вариант сверхкумуляции исключён *a priori*.

Описанный подход можно также использовать для сравнения эффективности двух различных субстанций (токсикантов) по эффективным (токсическим) дозам $ED(1)_{50}$ и $ED(2)_{50}$ вызывающим сходный эффект. При этом, по-видимому, нужно руководствоваться соображениями, изложенными выше для II-го случая.

Кроме коэффициента кумуляции (1.27), используя подход, исторически связанный с фамилиями Личфилда-Вилкоксона [10], принято также вычислять коэффициент кумуляции (в данном случае по наблюдению летальных эффектов) по следующей формуле:

$$\bar{K}_{cum} = \frac{LD_{50}^1 \cdot s_1}{LD_{50}^n \cdot s_n} , \quad \text{где} \quad (1.30)$$

$s_{1,n}$, так называемая функция наклона прямой, определяемая, в свою очередь, формулой:

$$s_{1,n} = \frac{1}{2} \left(\frac{LD_{84}^{1,n}}{LD_{50}^{1,n}} + \frac{LD_{50}^{1,n}}{LD_{16}^{1,n}} \right) \quad (1.31)$$

Здесь цифры при дозе (нижний индекс) означают вероятность достижения эффекта в процентах.

Если записать характерные дозы из формулы (1.31) через пробиты, то получим:

$$LD_{16,50,84} = 10^{\left(\frac{4,5,6-a}{b}\right)} \quad (1.32)$$

Далее нетрудно получить, что функция наклона прямой $s_{1,n} = 10^{(1/b_{1,n})}$, т.е. является просто потенцированным котангенсом угла наклона прямой зависимости пробита от логарифма дозы. А формула для коэффициента кумуляции (1.30) может быть теперь переписана в укороченном виде:

$$\bar{K}_{cum} = \frac{LD_{84}^1}{LD_{84}^n} \quad (1.33)$$

Таким образом, статистическая достоверность коэффициента \bar{K}_{cum} может быть также установлена с использованием подхода (1.28-1.29), который предложен для решения аналогичного вопроса в случае K_{cum} (1.27).

Для установления соответствия в расчетах эффективных доз и их доверительных интервалов, а также статистической достоверности коэффициентов кумуляции в разных программных продуктах нами был проанализирован ряд общедоступных, рекомендуемых для этих целей программ (включая наши программные файлы в среде *Mathematica* 9 (<http://www.wolfram.com/>):

StatsDirect (<http://www.statsdirect.com/>);

StatPlus (Статистика+) (автор – Алексей Симачёв, <http://www.statplus.net.ua/>);

LdP Line (автор — Ehab Bakr, <http://www.ehabsoft.com/>);

IBM SPSS Statistics 22 (<http://www-01.ibm.com/software/analytics/spss/>);

Statistica 10 (<http://www.statsoft.ru>).

Исходные данные для расчета эффективных концентраций и их доверительных интервалов были получены экспериментальным путем на модели *Daphnia magna Straus* по показателю летальности с использованием стандартного для данного вида токсиканта в соответствии с [11]. В качестве токсиканта служил бихромат калия, х.ч., по ГОСТ 4220-75, он же был использован для контроля чувствительности используемых тест-организмов и оперативного контроля приемлемости результатов измерений. Диапазон исследованных концентраций бихромата калия составлял

0,3 - 2,6 мг/см³. Количество экспериментальных точек регрессии равнялось пяти (для концентраций 0,3; 1,0; 1,5; 2,0 и 2,6 мг/см³). Повторность опыта для каждой исследованной концентрации была 6-ти кратной. В случае концентрации, близкой к LC₅₀, повторность опыта была 24-х кратной. Смертность рачков учитывалась в течение 24 часов. Каждая отдельная выборка, из которой определялась частота эффекта, составляла 30 особей.

Для оценки достоверности установленных коэффициентов кумуляции дополнительно к исходным, полученным на модели *Daphnia magna Straus*, были сгенерированы данные указывавшие на сверхкумуляцию по токсическому эффекту и выраженную кумуляцию в соответствии с (1.33).

Результаты и обсуждение

Все входные данные были обработаны вышеуказанными программами, и результаты вычислений сведены в табл.1-2.

Таблица 1. Точечные оценки эффективных доз (LD_n) и границ их доверительны интервалов (CI_n), выполненные с помощью различных программных продуктов

	LD ₅	LD ₁₀	LD ₅₀	LD ₈₄	CI ₅	CI ₁₀	CI ₅₀	CI ₈₄
<i>Mathematica 9</i>	0,3877	0,5125	1,3711	2,9549	0,8530	0,9825	1,9512	9,1848
					0,0473	0,0947	0,9129	1,8260
StatsDirect	0,4400	0,5665	1,3818	2,7599	1,2192	1,2968	1,8479	4,9169
					0,1511	0,2380	1,0295	1,5870
StatPlus	0,4399	0,5665	1,3818	2,7599	–	–	1,9549	–
					–	–	0,7121	–
LdP Line	0,4396	0,5661	1,3817	2,7614	–	–	–	–
IBM SPSS	0,440	0,567	1,382	2,842**	0,778	0,907	1,963	24,413

Statistics 22					0,008	0,024	0,791	1,988
Statistica 10*	-	-	-	-	-	-	-	-
					-	-	-	-

*- соответствующий модуль Statistica 10 выдает только значения коэффициентов a и b , которые до третьего знака после запятой совпадают с аналогичными данными из программ: IBM SPSS Statistics 22, Mathematica 9, StatPlus, StatsDirect; значение a при этом выдается для пробит-регрессии не смещенной на 5 единиц.

** - в выходной таблице с оценками эффективных доз и их доверительных интервалов нет частоты 0,84; в данном случае она заменена на ближайшее значение – 0,85.

Таблица 2. Оценки коэффициентов кумуляции и их статистической достоверности, выполненные с помощью различных программных продуктов

	K_{cum}	\bar{K}_{cum}	$P(D_{(n)} > D_{(1)})$	$\bar{P}(D_{(n)} < D_{(1)})$
Mathematica 9	0,6858	1,6416	0,0309	0,9704
StatsDirect	—	—	—	—
StatPlus	—	1,5472	—	—
LdP Line	—	—	—	—
IBM SPSS Statistics 22	—	—	—	—
Statistica 10	—	—	—	—

Как следует из анализа табл. 1-2 программный продукт Statistica 10 оказался наименее приспособлен для нужд токсикологов в плане автоматической выдачи необходимой им информации. Хотя необходимо отметить, что в сети в свободном доступе встречается макрос, который дополняет основную программу и обеспечивает выдачу токсикологических параметров в объеме, сходном со StatPlus. Макрос был написан для более ранних версий Statistica, в связи с чем существуют определенные проблемы с совместимостью.

Если судить по объему выдаваемой полезной информации, среди «чужих» программных продуктов лидирует IBM SPSS Statistics 22, но его ограничения связаны с отсутствием возможности автоматических расчетов коэффициентов кумуляции. Эта функция наличествует только в StatPlus, но ограничена выдачей только одного коэффициента – \bar{K}_{cum} . Кроме того программа не рассчитывает доверительные интервалы для эффективных доз за исключением CI_{50} .

Программа StatsDirect напротив выдает доверительные интервалы для любых эффективных доз, но в ней также как и в IBM SPSS Statistics 22 не заложена опция автоматического расчета коэффициентов кумуляции.

И наконец, не один из программных продуктов, за исключением *Mathematica 9*, не оценивает статистическую достоверность установленных коэффициентов кумуляции. А с учетом тех обстоятельств, что в программных файлах *Mathematica 9* заложены более строгие математические подходы для получения точечных оценок эффективных доз и границ их доверительных интервалов, а также бесплатности распространения, преимущества данного программного продукта очевидны.

Программные модули *Mathematica 9* находятся в свободном доступе на сайте ФГБУН НИЦ ТБП ФМБА России. Для тестирования модули можно скачать по ссылке <http://toxicbio.ru/rus/activity/publications/> в виде архива вместе с инструкцией по их использованию.

Выводы

1. Для выполнения пробит-анализа в качестве базового был использован метод функции максимального правдоподобия Фишера, для нахождения максимума которой использовались современные итерационные процедуры поиска экстремума функции 2-х переменных. Непоправленные дисперсии коэффициентов регрессии вычислялись точно, через квадратичную форму 2-го дифференциала в точке максимума.
2. Для вычисления точечных оценок эффективных доз и их доверительных интервалов впервые был использован подход, основанный на рассмотрении коэффициентов регрессии как случайных величин, аппроксимированных t -распределениями Стьюдента (сдвинутыми и деформированными) с учетом поправок на фактор гетерогенности. Характерные логарифмические дозы, являющиеся отношениями таких случайных величин, определялись через построение соответствующих функций распределения с последующим точным вычислением математических ожиданий и граничных точек.
3. Впервые на программном уровне была реализована возможность корректно попарно сравнивать характерные дозы схожего эффекта с использованием аппарата функций плотности распределения вероятности.
4. В рамках пакета *Mathematica 9* было создано несколько программных модулей, позволяющих оценивать статистическую достоверность 2-х общепринятых в токсикологии коэффициентов кумуляции, основанных на соответствующих отношениях ED_{50} и ED_{84} и рассчитывать математическое ожидание любой характерной логарифмической дозы в совокупности с границами доверительного интервала. Программные модули вместе с инструкцией по их использованию находятся в свободном доступе на сайте ФГБУН НИЦ ТБП ФМБА России.

Литература

1. Gaddum J.H. Reports on biological standards. III. Methods of biological assay depending on a quantal response. // Spec. Rep. Ser. Med. Res. Coun., Lond. 1933. No. 183.
2. Bliss C.I. The method of probits. // Science. 1934. V. 79. P. 38--39, 409--410.
3. Fisher R. A. On the Mathematical Foundations of Theoretical Statistics. // Phil. Trans. R. Soc. Lond. A. 1922. V. 222. P. 309—368.
4. Finney D.J. Probit analysis. A statistical treatment of the sigmoid response curve. // Cambridge at the university press. 1947.
5. Fisher R. A. Appendix to Bliss C.I. The case of zero survivors. // Ann. Appl. Biol. 1935. V. 22. P. 164—165.
6. Garwood F. The application of maximum likelihood to dosage-mortality curves. // Biometrika. 1940. V. 32. P. 46—58.
7. Irwin J.O. On the calculation of the error of biological assays. // J. Hyg. Camb. 1943. V. 43. P. 121—128.
8. Fieller E.C. A fundamental formula in the statistics of biological assay, and some applications. // Quart. J. Pharm. 1944. V. 17. P. 117—123.
9. Fieller E.C. Some problems in interval estimation. // J. Royal Stat. Soc. B. 1954. V. 16. No. 2. P. 175—185.
10. Litchfield J.T., Wilcoxon F.A., A simplified method of evaluating dose-effect experiments. // J. Pharmacol. 1949. V. 96. P. 99—113.
11. Методика определения токсичности воды и водных вытяжек из почв, осадков сточных вод, отходов по смертности и изменению плодовитости дафний; ФР.1.39.2007.03222.