

Многомерный метод индивидуального прогнозирования индекса моторики¹

А.В. Вильдеман, А.А. Ташкинов, В.А. Бронников

Аннотация. Изучаются проблемы прогнозирования двигательного развития у больных детским церебральным параличом (ДЦП) на основе многомерного статистического анализа. Предлагается метод индивидуального прогнозирования индекса моторики у больного ДЦП, обобщенный на многомерный случай. Показано, что учет в моделях прогнозирования влияния родовых и дородовых факторов риска позволяет улучшить достоверность прогнозов по сравнению с одномерными моделями.

Ключевые слова: математическое моделирование, индивидуальное прогнозирование, многомерный статистический анализ, медико-социальные системы.

Введение

Анализ состояния и поведения реальных объектов во многих случаях связан с применением статистических моделей и методов для исследования закономерностей и определения трендовых зависимостей. Изучаемые явления и процессы, как правило, протекают в условиях многофакторности и требуют от исследователя использования аппарата многомерного статистического анализа.

Одним из направлений моделирования систем и процессов является разработка математических моделей прогнозирования, описывающих взаимосвязи между ключевыми факторами области приложения. Важное практическое значение при создании прогнозных моделей имеет учет индивидуальных характеристик конкретного объекта, позволяющий скорректировать параметры групповой модели и получить более достоверный индивидуальный прогноз.

Особое место среди статистических исследований занимает моделирование медико-социальных систем, находящееся на стыке информационных технологий, математических

методов и современных подходов в медицине и социальной реабилитации.

Широкий круг приложений методов статистического анализа в медицине охватывает такие области как кардиология [8], пульмонология [1], неврология [6], хирургия [3].

Настоящее исследование направлено на создание математических моделей индивидуального прогнозирования состояния больного с врожденными нарушениями двигательных функций с учетом влияния родовых и дородовых факторов риска.

1. Исходные данные

Объектом исследования являлся больной детским церебральным параличом (ДЦП), состояние которого характеризовалось рядом дородовых, родовых и текущих факторов. Всего рассматривалось 62 фактора. Анализируемая база данных содержала сведения о 73 больных. Каждый больной наблюдался от одного до нескольких раз в течение различных интервалов времени. Общий объем статистической выборки при этом составил 157 наблюдений.

¹ Работа выполнена при финансовой поддержке РФФИ (грант № 10-04-96096-р_урал_a).

Данные были получены в ходе совместных исследований Пермского государственного технического университета с Краевым центром комплексной реабилитации инвалидов (г. Пермь).

Исследуемая выборка формировалась в результате нерегулярных обращений пациентов с диагнозом ДЦП в медико-социальное учреждение в течение десяти лет. При этом, несмотря на достаточно большое количество больных, обследованных на протяжении данного промежутка времени, удалось выделить только 5 человек, постоянно наблюдаемых в течение всего рассматриваемого периода. В связи с этим предлагаемые модели индивидуального прогнозирования иллюстрировались на примере данной группы больных, а для вычисления средних параметров использовалась вся исходная выборка.

В качестве прогнозируемого показателя рассматривался индекс моторики Y – величина, характеризующая интегральный уровень развития двигательных навыков. Исходя из способа вычисления, Y изменяется в интервале от 0 до 60. При этом верхней границе соответствует наилучшее двигательное развитие, нижней – наихудшее.

Изменение индекса моторики с возрастом у группы больных представляет собой случайный процесс $Y(t)$. Динамика изменения индекса моторики для j больного из группы является реализацией $y^{(j)}(t)$ случайного процесса $Y(t)$.

2. Одномерные модели индивидуального прогнозирования

В соответствии с подходом индивидуального прогнозирования, используемым в задачах моделирования поведения технических объектов [4], построение прогнозных моделей начинается с аппроксимации зависимости среднего прогнозируемого показателя от времени.

Для определения среднего (по реализациям) индекса моторики применялся метод последовательного выделения экспоненциальных слагаемых [7], в основе которого лежит предположение о наличии математической модели зависимости, заданной с точностью до параметров:

$$f(t, \theta) = a_0 + \sum_{q=1}^Q a_q (1 - e^{-\lambda_q t}), \quad (1)$$

где $f(t, \theta)$ – групповой тренд, аппроксимирующий зависимость среднего индекса моторики от возраста;

$\theta = (a_0 \ a_1 \ \dots \ a_q \ \dots \ a_Q \ \lambda_1 \ \dots \ \lambda_q \ \dots \ \lambda_Q)^T$ – вектор неизвестных параметров.

При построении моделей индивидуального прогнозирования вместо выражения (1) рассматривалась случайная функция:

$$Y(t) = A_0 + \sum_{q=1}^Q A_q(t)(1 - e^{-\lambda_q t}), \quad (2)$$

где $A_q(t) = A_q + A'_q(t)$;

A_q – случайные величины, для которых:

$$\langle A_q \rangle = a_q;$$

$A'_q(t)$ – центрированные случайные функции,

для которых: $\langle A'_q(t) \rangle = 0$;

$\langle \cdot \rangle$ – знак осреднения по реализациям.

Таким образом, функция (2) представима в виде:

$$Y(t) = F(t) + \xi(t), \quad (3)$$

где $F(t) = A_0 + \sum_{q=1}^Q A_q (1 - e^{-\lambda_q t})$ – главная составляющая индекса моторики;

$\xi(t) = \sum_{q=1}^Q A'_q(t)(1 - e^{-\lambda_q t})$ – наложенный на главную составляющую «шум», действием которого в силу малости по сравнению с $F(t)$ можно пренебречь.

В соответствии с принятыми предположениями тренд индивидуального индекса моторики, описывающий динамику двигательного развития у j -го пациента, представлялся в виде:

$$F^{(j)}(t) = A_0^{(j)} + \sum_{q=1}^Q A_q^{(j)} (1 - e^{-\lambda_q t}), \quad (4)$$

где $A_0^{(j)}$ – свободный член, определяющий начальный индекс моторики у j -го больного; в рамках настоящего исследования данная величина считалась неизвестной, поскольку все пациенты в выборке начали наблюдаться спустя некоторое время после рождения;

$1 - e^{-\lambda_q t}$ – экспоненциальные базисные функции;

$A_q^{(j)}$ ($q = \overline{1, Q}$) – неизвестные коэффициенты базисных функций;

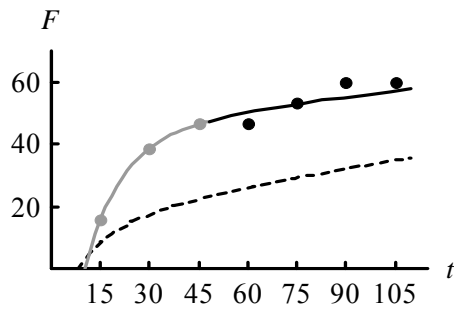


Рис. 1. Индивидуальные замеры (точки), индивидуальный прогноз (сплошная линия) и групповой тренд (пунктирная линия) индекса моторики. Серые точки – данные, на которых строился прогноз, черные точки – контрольные измерения

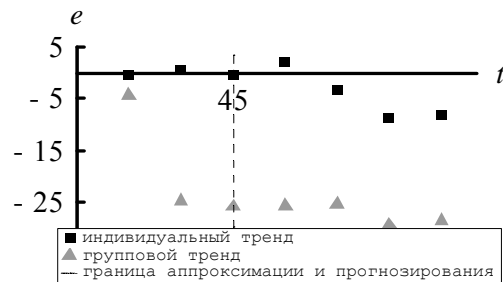


Рис. 2. Отклонения модельных значений от реальных замеров индекса моторики. Слева от пунктирной линии – отклонения при аппроксимации известных данных, справа – отклонения при прогнозировании неизвестных значений

λ_q ($q = \overline{1, Q}$) – известные константы базисных функций, определенные при выделении группового тренда $f(t, \theta)$.

Задача индивидуального прогнозирования заключается в том, чтобы по ряду измерений состояния больного на начальном интервале протекания заболевания τ_0 спрогнозировать развитие его состояния в будущем.

Для определения неизвестных $A_q^{(j)}$ использовался метод наименьших квадратов на интервале τ_0 , а также условия монотонного возрастания $F^{(j)}(t)$: $A_q^{(j)} \geq 0$ ($q = \overline{1, Q}$).

В результате были получены индивидуальные прогнозы индекса моторики для каждого больного из группы. В качестве исходных данных рассматривался определенный ранее групповой тренд и ряд индивидуальных замеров индекса моторики в течение первых четырех лет жизни. При этом прогнозы строились на ближайшие пять лет. Была показана эффективность применения метода для случая, когда интервал основания прогноза меньше интервала, на который осуществляется прогнозирование [9].

На Рис. 1 в качестве примера приведены графики группового тренда и индивидуального индекса моторики для больного, динамика двигательного развития которого наиболее сильно отличается от среднегрупповой. На Рис. 2 построены соответствующие графики отклонений модельных значений от реальных замеров индекса моторики $e(t)$; t – возраст в месяцах.

Следует отметить, что отклонения данных индивидуальной модели от реальных значений индекса моторики делятся на две части: отклонения, полученные на исходных данных и отражающие погрешность аппроксимации (слева от пунктирной линии), и отклонения, полученные в результате проверки прогнозной модели на контрольных данных, отражающие погрешность прогнозирования (справа от пунктирной линии).

Из Рис. 1 и Рис. 2 наглядно следует, что индивидуальная зависимость с большей степенью точности аппроксимирует эмпирические данные по сравнению с групповым трендом.

3. Обобщение метода индивидуального прогнозирования на многомерный случай

3.1. Основы построения метода

Во время наблюдений за пациентами, помимо значений индекса моторики, определялись также значения ряда факторов, отражающих комплексное состояние больного.

Методом деревьев классификации было определено множество факторов P_r ($r = \overline{1, R}$), характеризующих родовое и дородовое состояние и оказывающих значимое влияние на последующее развитие двигательных навыков у ребенка с церебральным параличом [10].

В Табл. 1 приведены наиболее значимые родовые и дородовые факторы риска и соответствующие им коэффициенты корреляции с индексом моторики.

Табл. 1. Множество значимых факторов, характеризующих родовое и дородовое состояние больного

Обозначение	Наименование	Коэффициент корреляции с индексом моторики, γ_r
X36	Наличие признаков внутриутробного инфицирования	- 0,50
X37	Наличие признаков внутриутробной гипоксии	- 0,39
X33	Низкая масса при рождении	- 0,28

Отобранные факторы P_r принимали одно из двух значений: 0 – «нет» или 1 – «да».

В настоящей работе была предпринята попытка учесть влияние этих факторов на развитие индекса моторики с целью уточнения полученных ранее одномерных моделей индивидуального прогнозирования.

Для полученной зависимости индивидуального индекса моторики от возраста (Рис. 1) характерно отрицательное значение свободного члена $A_0^{(j)}$, представляющего собой величину индекса моторики в момент времени $t = 0$.

По определению индекс моторики является неотрицательной величиной. Поэтому его динамику будем рассматривать только в положительной области, а отрицательные значения свободного члена будем трактовать как задержку начала двигательного развития.

В то время как для здорового человека $A_0^{(j)}$, как правило, положительно, т.е. ребенок в момент рождения уже обладает начальными моторными навыками, для больного ДЦП в зависимости от тяжести заболевания процесс двигательного развития может начаться с запозданием в момент времени, отличный от $t = 0$ [2].

В работе [11] было показано, что факторы родового и дородового состояния больного оказывают достаточно сильное влияние на процесс моторного развития в последующие годы. В частности было обнаружено, что в возрасте старше 7 лет на уровне развития двигательных способностей во многом сказывается экстремально низкая масса больного при рождении. Показано, что наличие признаков внутриутробного кислородного голодания или признаков внутриутробного инфицирования больного особенно негативно

сказывается при формировании двигательных навыков в возрасте до 4 лет.

О негативном влиянии факторов P_r говорит также знак «-» перед коэффициентами корреляции в Табл. 1.

Таким образом, родовые и дородовые факторы риска формируют начальные условия развития ребенка с церебральным параличом. В связи с этим представляется обоснованным выдвижение гипотезы о том, что факторы P_r , определяющие задержку начала двигательного развития, влияют на величину свободного члена $A_0^{(j)}$.

Введем обобщенный фактор родовых и дородовых условий:

$$\Phi = \sum_r \gamma_r P_r. \tag{5}$$

В качестве весовых коэффициентов γ_r в выражении (5) в силу линейности негативного влияния P_r представляется оправданным использование коэффициентов корреляции, характеризующих степень тесноты линейной связи между индексом моторики и P_r . Таким образом, если обозначить $P_r^{(j)}$ – индивидуальное значение фактора P_r , то величина обобщенного фактора родовых и дородовых условий для j -го больного будет определяться выражением $\Phi^{(j)} = \sum_r \gamma_r P_r^{(j)}$.

При создании одномерных моделей индивидуального прогнозирования определение $A_0^{(j)}$ выполнялось без учета негативного влияния родовых и дородовых факторов. Таким образом, может быть выдвинута гипотеза о том, что найденное в результате решения одномерной задачи значение $A_0^{(j)}$ представляет собой величину свободного члена, соответствующую отсутствию травмирующих факторов в момент рождения и факторов, негативно сказывающихся на развитии ребенка в период беременности $A_0^{(j)}|_{\Phi=0}$.

Зависимость $A_0^{(j)}$ от $\Phi^{(j)}$ для каждого больного в группе также будем искать в линейном виде:

$$A_0^{(j)} = A_0^{(j)}|_{\Phi=0} + \mu \Phi^{(j)}, \tag{6}$$

где μ – пока неизвестный угловой коэффициент линейной зависимости, определяющий,

насколько сильно начальный индекс моторики изменяется под влиянием факторов риска P_r .

Выражение (6), таким образом, представляет собой индивидуальную зависимость. Однако определение значения μ для индивидуального больного не представляется возможным, поскольку для конкретного человека невозможно воспроизвести реализации начального индекса моторики, соответствующие различным родовым и дородовым условиям. Следовательно, μ может быть определен только на основе данных группы пациентов.

Поиск углового коэффициента μ может быть выполнен следующим образом. Если нанести на диаграмму разброса реальные данные начального индекса моторики $y^{(j)}|_{t=0}$ и обобщенного фактора родовых и дородовых условий $\Phi^{(j)}$ для группы пациентов, то каждой точке на диаграмме будет соответствовать отдельный больной, состояние которого определяет пара значений $(y^{(j)}|_{t=0}, \Phi^{(j)})$. Полученные таким образом точки могут быть аппроксимированы линейной зависимостью, найденной по методу наименьших квадратов:

$$\hat{y}^{(j)}|_{t=0} = \nu + \mu\Phi^{(j)}, \quad (7)$$

где μ – искомый коэффициент.

Если величины индекса моторики при рождении у группы больных неизвестны, то вместо $y^{(j)}|_{t=0}$ могут быть использованы реальные данные индекса моторики, соответствующие моменту времени, наиболее близкому к $t = 0$.

3.2. Общий алгоритм метода

Таким образом, предлагаемый многомерный метод индивидуального прогнозирования основывается на следующем алгоритме.

1. На первом этапе методом последовательного выделения экспоненциальных слагаемых вычисляются параметры группового тренда индекса моторики λ_q ($q = \overline{1, Q}$).

2. Величина индекса моторики в начальный момент времени при отсутствии влияния родовых и дородовых факторов риска $A_0^{(j)}|_{\phi=0}$ нахо-

дится в результате решения задачи оптимизации с ограничениями типа неравенств:

$$\begin{cases} \sum_{i=1}^{n_0} \left(A_0^{(j)}|_{\phi=0} + \sum_{q=1}^Q A_q^{(j)} (1 - e^{-\lambda_q t_i}) - y_i^{(j)} \right)^2 \rightarrow \min, \\ A_q^{(j)} \geq 0 \quad (q = \overline{1, Q}) \end{cases}, \quad (8)$$

где $A_0^{(j)}|_{\phi=0}$ и $A_q^{(j)}$ ($q = \overline{1, Q}$) – неизвестные величины, доставляющие минимум в задаче (8);

n_0 – количество измерений на интервале τ_0 ;

$y_i^{(j)}$ – величина реального индекса моторики у j -го пациента в i момент времени;

$t_i \in \tau_0$.

Задача (8) может быть решена численно, например, одним из градиентных методов, либо аналитически по теореме Куна-Таккера [5].

3. С помощью линейной аппроксимации зависимости реального индекса моторики в начальный момент времени от величины обобщенного фактора родовых и дородовых условий для группы больных вычисляется угловой коэффициент μ .

4. Полученный коэффициент μ используется для корректировки $A_0^{(j)}$ с учетом влияния факторов P_r по формуле (6).

5. Остальные индивидуальные коэффициенты $A_q^{(j)}$ ($q = \overline{1, Q}$) находятся из решения задачи оптимизации, аналогичной рассмотренной ранее, с использованием вычисленного на предыдущем шаге скорректированного $A_0^{(j)}$:

$$\begin{cases} \sum_{i=1}^{n_0} \left(A_0^{(j)} + \sum_{q=1}^Q A_q^{(j)} (1 - e^{-\lambda_q t_i}) - y_i^{(j)} \right)^2 \rightarrow \min, \\ A_q^{(j)} \geq 0 \quad (q = \overline{1, Q}) \end{cases}. \quad (9)$$

3.3. Проверка достоверности многомерных моделей

Для иллюстрации предлагаемого метода были построены индивидуальные прогнозы индекса моторики для каждого из пациентов в группе (Рис. 3).

При вычислении групповых параметров λ_r и μ использовались данные всей исходной выборки из 157 наблюдений.

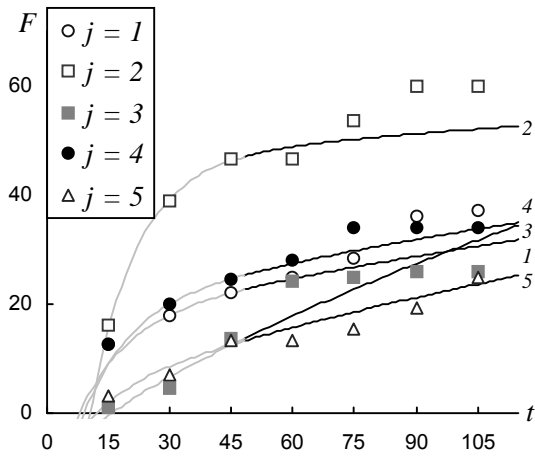


Рис. 3. Индивидуальные замеры и индивидуальные прогнозы индекса моторики, полученные с учетом влияния родовых и дородовых факторов риска

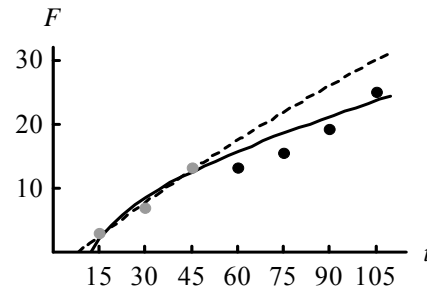


Рис. 4. Индивидуальные замеры (точки), индивидуальный прогноз индекса моторики для больного $j=5$ с учетом влияния родовых и дородовых факторов (сплошная линия) и индивидуальный прогноз без учета этого влияния (пунктирная линия)

В Табл. 2 приведены для сравнения коэффициенты детерминации R^2 , вычисленные в результате применения одномерного и многомерного методов индивидуального прогнозирования.

Табл. 2. Сравнительный анализ достоверности моделей прогнозирования

Номер реализации j	1	2	3	4	5	Итого
Одномерная модель (R^2)	0,84	0,96	0,85	0,95	0,60	0,84
Многомерная модель (R^2)	0,84	0,89	0,88	0,95	0,92	0,90

В результате учета влияния родовых и дородовых факторов величина коэффициента R^2 , характеризующего близость модели к реальным данным, увеличилась в среднем для группы больных с 0,84 до 0,90 (Табл. 2). При этом несколько ухудшился R^2 для пациента $j=2$: с 0,96 до 0,89, однако существенно улучшилась достоверность прогнозирования для больного $j=5$: с 0,60 до 0,92 (Рис. 4).

Заключение

В работе рассмотрен подход многомерного статистического анализа, учитывающий при создании моделей прогнозирования индивидуальные характеристики конкретного объекта.

Предложенный метод индивидуального прогнозирования, обобщенный на многомерный случай, позволяет увеличить степень достовер-

ности прогнозных моделей, что было продемонстрировано на примере группы больных детским церебральным параличом.

Полученные математические модели с достаточной точностью описывают изменение индивидуального индекса моторики у больного ДЦП с возрастом в случае, когда интервал основания прогноза меньше периода, на который строится прогноз.

Показано, что многомерные модели индивидуального прогнозирования, учитывающие влияние факторов риска в момент рождения и период беременности, лучшим образом предсказывают отрицательную динамику развития двигательных навыков, что играет важную роль при применении их на практике.

Литература

1. Безруков Н.С. Еремин Е.Л., Перельман Ю.М. Автоматизированная система диагностики заболеваний легких // Проблемы управления. – 2007. – № 5. – С. 75–80.
2. Бронников В.А., Лебедева А.Ю., Мавликаева Ю.А. Возможности прогнозирования двигательного развития детей со спастическими формами церебральных параличей // Актуальные проблемы реабилитации детей, страдающих церебральным параличом: материалы второго международного конгресса. – 2007. – С. 33–35.
3. Драгун И.А. Автоматизированная система количественной оценки операционного риска: автореф. дис... канд. тех. наук. – Барнаул, 2006. – 22 с.
4. Еремин Ю.А., Радченко В.П., Самарин Ю.П. Расчет индивидуальных деформационных свойств элементов конструкций в условиях ползучести // Машиноведение. – 1984. – № 1. – С. 67–72.

5. Кузнецов А.В., Сакович В.А., Холод Н.И. Высшая математика: Математическое программирование. – Минск: «Вышэйшая школа», 1994. – 286 с.
6. Реброва О.Ю. Применение методов интеллектуального анализа данных для решения задачи медицинской диагностики // Новости искусственного интеллекта. – 2004. – № 3. – С. 76–80.
7. Самарин Ю.П. Построение экспоненциальных аппроксимаций для кривых ползучести методом последовательного выделения экспоненциальных слагаемых // Проблемы прочности. – 1974. – № 9. – С. 24 – 27.
8. Суфиянов В.Г. Разработка адаптивных статистических моделей классификации и прогнозирования: автореф. дис... канд. физ.-мат. наук. – Пермь, 2004. – 16 с.
9. Ташкинов А.А., Вильдеман А.В., Бронников В.А. Индивидуальное прогнозирование двигательного развития у больных детским церебральным параличом на основе подходов статистического анализа // Российский журнал биомеханики. – 2010. – № 2 (в печати).
10. Ташкинов А.А., Вильдеман А.В., Бронников В.А. Модели классификации в задачах прогнозирования двигательного развития у детей с церебральным параличом // Системный анализ и управление в биомедицинских системах. – 2010. – Том 9. – № 1. – С. 142 – 149.
11. Ташкинов А.А., Вильдеман А.В., Бронников В.А. Применение метода деревьев классификации к прогнозированию уровня развития моторики у больных с нарушениями двигательных функций // Российский журнал биомеханики. – 2008. – Т. 12, № 4 (42). – С. 84 – 95.

Вильдеман Александр Валерьевич. Директор Центра математического моделирования медико-социальных систем и процессов Пермского государственного технического университета. Окончил Пермский государственный технический университет в 2008 году. Автор 17 научных работ. Область научных интересов: многомерный статистический анализ, математические модели прогнозирования, информационно-аналитические системы. E-mail: wildemann@mail.ru.

Ташкинов Анатолий Александрович. Первый проректор Пермского государственного технического университета. Окончил Пермский политехнический институт в 1979 году. Доктор физико-математических наук, профессор. Автор 227 печатных работ, 11 монографий. Область научных интересов: методы решения стохастических краевых задач, математическое моделирование процессов в структурно-неоднородных средах, многомерный статистический анализ. E-mail: tash@pstu.ru.

Бронников Владимир Анатольевич. Директор-врач Центра комплексной реабилитации инвалидов (г. Пермь). Окончил Пермский государственный медицинский институт в 1989 году. Доктор медицинских наук, профессор. Автор 122 печатных работ и двух монографий. Область научных интересов: реабилитация инвалидов, качество жизни, детский церебральный паралич. E-mail: dolphinrehab@rambler.ru