

# Алгоритм идентификации параметров модели прогнозирования индекса моторики<sup>1</sup>

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

**Аннотация.** Рассматривается алгоритм идентификации параметров модели прогнозирования уровня двигательного развития – индекса моторики при последовательном накоплении результатов обследований. В ходе апробации алгоритма на выборке наблюдений за группой пациентов с церебральным параличом исследуется влияние увеличения объема исходных данных на точность индивидуальных прогнозов.

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

## Введение

Математические методы и информационные технологии занимают неотъемлемую часть современных биомедицинских исследований. Наряду с физико-механическим моделированием физиологических систем [2, 4, 10], интерес представляет изучение статистических закономерностей протекания заболеваний у пациентов с аналогичными диагнозами. Вероятностно-статистическое моделирование в медицине охватывает большое количество областей, среди которых: кардиология [13, 1], пульмонология [6, 3], неврология [11], физиология двигательных систем [5] и др.

Существенным недостатком статистических прогнозных моделей является то, что они справедливы для некоторого осредненного объекта и дают представление только о развитии исследуемого процесса в среднем. Для более точного прогнозирования необходима идентификация параметров статистической модели с использованием данных, характеризующих особенности конкретного объекта. При этом из множества неизвестных выделяется ряд величин, характе-

ризующих групповой тренд и вычисляемых в результате осреднения по большой совокупности объектов. Оставшиеся параметры определяют отличительные особенности конкретной реализации рассматриваемого случайного процесса. Получение достоверных персональных прогнозов требует постоянной корректировки индивидуальных моделей по мере поступления новой информации об исследуемом объекте. При автоматизации процессов обработки и анализа данных необходима алгоритмизация используемых методов и методик.

Практическим применением настоящего подхода в медицине является прогнозирование двигательного развития у детей с церебральным параличом. В качестве комплексного показателя, определяющего интеграционную оценку различных навыков движения, рассматривается индекс моторики [7]. При алгоритмизации метода индивидуального прогнозирования двигательного развития интерес представляет также исследование влияния объема исходных данных на точность результатов и определение оптимального соотношения периода обучения моделей и горизонта прогнозирования.

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ (грант № 10-04-96096-р\_урал\_a).

## 1. Алгоритм идентификации параметров при последовательном накоплении данных

### Шаг 0. Определение среднего тренда

Установление зависимости среднего индекса моторики от возраста производится с использованием метода последовательного выделения экспоненциальных слагаемых [9, 12, 14]. Осреднение производится по реализациям случайного процесса развития двигательных навыков.

Для применения метода необходимо наличие представительной выборки наблюдений за одновозрастной группой пациентов с аналогичными диагнозами. Обследования должны производиться через примерно одинаковые промежутки времени у каждого из пациентов. В ходе обследования экспертами должны оцениваться различные группы двигательных навыков, в результате суммирования которых вычисляется индекс моторики [7]. Значения данного показателя у детей с церебральными параличами в одном и том же возрасте, как правило, существенно отличаются [8, Рис. 3]. В указанной работе коэффициент вариации индекса моторики внутри сечения случайного процесса развития двигательных навыков у группы пациентов в среднем составил 49 %.

Период наблюдения группы пациентов должен покрывать интервал обучения математической модели и интервал прогнозирования.

Используемый метод восстановления зависимости среднего индекса моторики от возраста является параметрическим и предполагает аппроксимацию табличных значений набором экспоненциальных базисных функций:

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

где  $a_0$ ,  $a_q$ ,  $\lambda_q$  ( $q = \overline{1, Q}$ ) – параметры, определяемые в результате применения метода.

### Шаг 1.1. Построение одномерной модели на основе единичного наблюдения

Рассмотрим первичное обращение пациента. В условиях существенной вариации индекса моторики для группы пациентов прогнозирование на основе среднего тренда, вычисленного на предыдущем шаге, является малоэффектив-

ным. Для повышения достоверности прогноза необходимо уточнение групповой зависимости с использованием имеющихся индивидуальных данных [14].

Тренд индивидуального индекса моторики ищется в виде монотонно возрастающей ограниченной сверху функции:

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

где  $J$  – условный номер пациента,  $A_0^{(J)}$ ,  $A_q^{(J)}$  ( $q = \overline{1, Q}$ ) – искомые индивидуальные коэффициенты;  $\lambda_q$  ( $q = \overline{1, Q}$ ) – известные величины, определенные при выделении среднего тренда  $f(t)$ .

Условие монотонного возрастания может быть наложено, поскольку в рамках настоящей работы исследуется только процесс накопления двигательных навыков, случаи утраты двигательных способностей не рассматриваются. Ограниченность индекса моторики объясняется способом его вычисления (от 0 до 60).

Индивидуальные коэффициенты определяются при решении выпуклой задачи условной оптимизации:

Найти  $A_q^{(J)} \Big|_{\text{Шаг 1.1}}$  ( $q = \overline{0, Q}$ ):

$$\begin{cases} (F^{(J)}(t_1) - y_1^{(J)})^2 \rightarrow \min, \\ A_q^{(J)} \geq 0 \quad (q = \overline{1, Q}), \\ F^{(J)}(t_1) \leq y_{\max}, \end{cases} \quad (3)$$

где  $y_1^{(J)}$  – величина реального индекса моторики у  $J$  пациента при первичном обследовании;  $t_1$  – возраст пациента при первичном обследовании;  $y_{\max}$  – наибольшее возможное значение индекса моторики.

Первое выражение системы (3) представляет собой условие минимума суммы квадратов отклонений модельной функции от реальных значений индекса моторики, второе – условие монотонного возрастания индекса моторики, третье – условие ограниченности индекса моторики.

Возможность применения модели (3) на практике была показана в работе [14] на приме-

ре группы пациентов с детским церебральным параличом. Исследуемая в работе выборка, объемом 157 наблюдений, была сформирована в результате нерегулярных обращений пациентов в реабилитационное учреждение в течение девяти лет. Удалось выделить 5 человек, постоянно наблюдаемых на протяжении всего рассматриваемого периода. Полученные таким образом динамические данные использовались для обучения и проверки моделей. Коэффициент детерминации при прогнозировании составил в среднем 0,84.

**Шаг 1.2. Переход к многомерной модели на основе единичного наблюдения**

Полученная на предыдущем шаге модель может быть уточнена при наличии сведений о состоянии здоровья родителей и близких родственников пациента до его зачатия, а также характеристик пренатального и перинатального периодов его жизни. Перечисленные группы факторов определяют родовые и дородовые начальные условия развития пациента [8].

Для учета их влияния в модели вводится обобщенный фактор:

$$\Phi = \sum_r \delta_r P_r, \tag{4}$$

где  $P_r$  ( $r = \overline{1, R}$ ) – факторы риска во время беременности и при рождении, оказывающие наиболее значимое влияние на индекс моторики;  $\delta_r$  – коэффициент корреляции  $P_r$  с индексом моторики.

Принимается гипотеза о том, что факторы  $P_r$  определяют задержку начала двигательного развития и, как следствие, влияют на величину свободного члена  $A_0^{(J)}$ .

С помощью линейной аппроксимации зависимости реального индекса моторики в начальный момент времени от величины обобщенного фактора родовых и дородовых условий для группы пациентов вычисляется угловой коэффициент  $\mu$ , определяющий, насколько сильно начальный индекс моторики изменяется в связи с влиянием факторов  $P_r$ .

Полученный коэффициент  $\mu$  используется для корректировки  $A_0^{(J)}$  с учетом влияния

родовых и дородовых факторов  $A_0^{(J)} = A_0^{(J)}|_{\Phi=0} + \mu\Phi^{(J)}$ .

За величину индекса моторики в начальный момент времени при отсутствии влияния родовых и дородовых факторов риска  $A_0^{(J)}|_{\Phi=0}$  принимается значение  $A_0^{(J)}|_{\text{Шаг 1.1}}$ , вычисленное на шаге 1.1 при решении задачи (3).

Полученное значение  $A_0^{(J)}|_{\text{Шаг 1.2}} = A_0^{(J)}|_{\text{Шаг 1.1}} + \mu\Phi^{(J)}$  подставляется в систему (3) вместо  $A_0^{(J)}$  и коэффициент  $A_0^{(J)}$  исключается из числа неизвестных. Значения всех остальных индивидуальных коэффициентов  $A_q^{(J)}|_{\text{Шаг 1.2}}$  ( $q = \overline{1, Q}$ ) находятся из решения задачи оптимизации (3), аналогичной рассмотренной на шаге 1.1, но имеющей на 1 неизвестную меньше.

**Шаг 2.1. Уточнение одномерной модели на основе двух индивидуальных наблюдений**

Исходными данными прогнозной модели при вторичном обследовании являются характеристики среднего индекса моторики, индивидуальные родовые и дородовые условия и два последовательных значения индекса моторики у исследуемого пациента.

На данном шаге повторяются все процедуры шага 1.1 с учетом дополнительных исходных данных и решается задача оптимизации:

$$\begin{cases} \text{Найти } A_q^{(J)}|_{\text{Шаг 2.1}} \quad (q = \overline{0, Q}): \\ \sum_{\gamma=1}^2 (F^{(J)}(t_\gamma) - y_\gamma^{(J)})^2 \rightarrow \min, \\ A_q^{(J)} \geq 0 \quad (q = \overline{1, Q}), \\ F^{(J)}(t_2) \leq y_{\max}, \end{cases} \tag{5}$$

**Шаг 2.2. Переход к многомерной модели на основе двух индивидуальных наблюдений**

Выполняемые на данном шаге процедуры аналогичны процедурам шага 1.2. Вычисляется значение индекса моторики в начальный момент времени  $A_0^{(J)}|_{\text{Шаг 2.2}} = A_0^{(J)}|_{\text{Шаг 2.1}} + \mu\Phi^{(J)}$ , с использованием которого находятся величины

$A_q^{(J)} \Big|_{\text{Шаг 2.2}} (q = \overline{1, Q})$ . Значение слагаемого  $\mu\Phi^{(J)}$  не нуждается в пересчете на этом и последующих шагах.

**Шаг N.1 Уточнение одномерной модели на основе N индивидуальных наблюдений**

Решается задача:

$$\begin{cases} \text{Найти } A_q^{(J)} \Big|_{\text{Шаг N.1}} (q = \overline{0, Q}): \\ \sum_{\gamma=1}^{\Theta_0} (F^{(J)}(t_\gamma) - y_\gamma^{(J)})^2 \rightarrow \min, \\ A_q^{(J)} \geq 0 (q = \overline{1, Q}), \\ F^{(J)}(t_\Theta) \leq y_{\max}, \end{cases} \quad (6)$$

где  $\Theta_0 = N$  – количество значений индекса моторики на интервале основания прогноза  $\tau_0$ ;  $y_\gamma^{(J)}$  – величина реального индекса моторики у J пациента в момент времени  $t_\gamma \in \tau_0$ ;  $t_\Theta$  – наиболее отдаленный момент времени, на который строится прогноз;  $y_{\max}$  – наибольшее возможное значение индекса моторики.

**Шаг N.2 Переход к многомерной модели на основе N индивидуальных наблюдений**

Решается задача:

$$\begin{cases} \text{Найти } A_q^{(J)} \Big|_{\text{Шаг N.2}} (q = \overline{1, Q}): \\ \sum_{\gamma=1}^{\Theta_0} \left( A_0^{(J)} \Big|_{\text{Шаг N.1}} + \mu\Phi^{(J)} + \sum_{q=1}^Q A_q^{(J)}(1 - e^{-\lambda_q t_\gamma}) - y_\gamma^{(J)} \right)^2 \rightarrow \min, \\ A_q^{(J)} \geq 0 (q = \overline{1, Q}), \\ A_0^{(J)} \Big|_{\text{Шаг N.1}} + \mu\Phi^{(J)} + \sum_{q=1}^Q A_q^{(J)}(1 - e^{-\lambda_q t_\Theta}) \leq y_{\max} \end{cases} \quad (7)$$

**2. Влияние объема исходных данных на точность индивидуальных прогнозов**

Рассмотрим в качестве примера задачу идентификации параметров модели прогнозирования индекса моторики в процессе наблюдений за пациентом J\* с детским церебральным параличом в течение первых 9 лет жизни.

В Табл. 1 представлены результаты наблюдений,  $t_\gamma$  – возраст пациента в момент обследования в месяцах,  $y_\gamma^{(J*)}$  – значение индекса моторики, соответствующее данному возрасту.

Средний тренд индекса моторики, вычисленный для группы пациентов с ДЦП с использованием процедур шага 0 (раздел 1), задается выражением [14]:

$$f(t) \Big|_{\text{Шаг 0}} = -16,49 + 66,87(1 - e^{-0,01t}) + 28,39(1 - e^{-0,09t}). \quad (8)$$

Одномерная модель прогнозирования на основе единичного наблюдения определяется в соответствии с шагом 1.1 алгоритма:

$$F^{(J^*)}(t) \Big|_{\text{Шаг 1.1}} = 0,82 + 0,48(1 - e^{-0,01t}) + 2,90(1 - e^{-0,09t}). \quad (9)$$

Если предположить, что во время первичного обследования врачу также стало известно о наличии травмирующих факторов в момент рождения и факторов, негативно сказывающихся на развитии больного в период беременности, это позволит построить многомерную модель индивидуального прогнозирования.

При расчете величины обобщенного фактора родовых и дородовых условий  $\Phi^{(J*)}$  в рамках шага 1.2 рассматриваются факторы риска во время беременности и при рождении  $P_r (r = \overline{1, R})$ , оказывающие наиболее значимое влияние на индекс моторики (Табл. 2). При определении коэффициентов корреляции  $\delta_r$  использовалась выборка объемом 157 наблюдений [8]. В Табл. 3 приведены индивидуальные значения родовых и дородовых факторов в следующих цифровых обозначениях: 0 – «нет», 1 – «да».

Обобщенный фактор родовых и дородовых условий с учетом коэффициентов корреляции, представленных в Табл. 2, и индивидуальных данных в Табл. 3 принимает значение:  $\Phi^{(J*)} = 0,67$ .

Коэффициент  $\mu$  для группы пациентов, полученный методом наименьших квадратов, составляет 16,68.

Многомерная модель прогнозирования на основе единичного наблюдения с учетом данных родовых и дородовых факторов принимает вид:

$$F^{(J^*)}(t) \Big|_{\text{Шаг 1.2}} = -10,35 + 25,12(1 - e^{-0,01t}) + 16,05(1 - e^{-0,09t}). \tag{10}$$

Табл. 1. Индивидуальные данные индекса моторики у пациента  $J^*$

$t_\gamma$ , мес.	15	30	45	60	75	90	105
$y_\gamma^{(J^*)}$	3.00	7.00	13.25	13.25	15.50	19.25	25.00

Табл. 2. Множество значимых родовых и дородовых факторов

$P_r$	Наименование	$\delta_r$
$P_1$	Наличие признаков внутриутробного инфицирования	-0.50
$P_2$	Наличие признаков внутриутробной гипоксии	-0.39
$P_3$	Экстремально низкая масса при рождении	-0.28

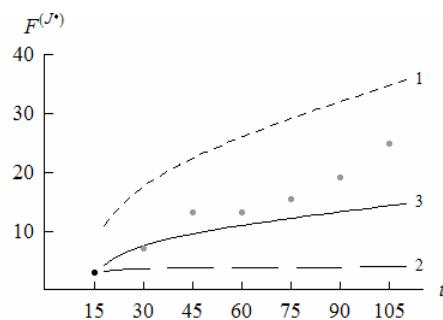
Табл. 3. Индивидуальные данные родовых и дородовых факторов у пациента  $J^*$

$P_1^{(J^*)}$	$P_2^{(J^*)}$	$P_3^{(J^*)}$
0	1	1

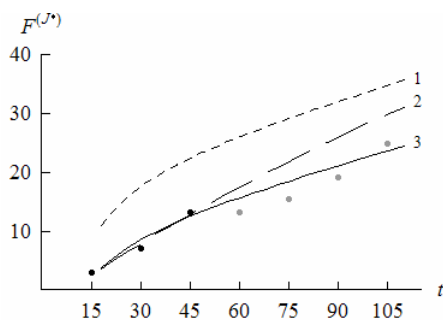
На Рис. 1 (а) построены графики  $f(t) \Big|_{\text{Шаг 0}}$ ,  $F^{(J^*)}(t) \Big|_{\text{Шаг 1.1}}$  и  $F^{(J^*)}(t) \Big|_{\text{Шаг 1.2}}$ . Как видно из Рис. 1 прогноз на основе среднего тренда (пунктирная линия, короткие штрихи), как и прогноз с использованием одномерной модели (пунктирная линия, длинные штрихи), очень плохо согласуются с контрольными значениями индекса моторики (серые точки). Переход к многомерной модели (сплошная линия) позволяет существенно улучшить качество прогнозирования. Однако данная модель не достаточно точно предсказывает уровень развития двигательных навыков в будущем.

Наилучшего качества прогнозирования удастся добиться, применяя многомерную модель на основе трех наблюдений (Рис. 1 (б)). При дальнейшем увеличении количества наблюдений результаты прогнозирования с использованием одномерной и многомерной моделей постепенно сближаются (Рис. 1 (в)).

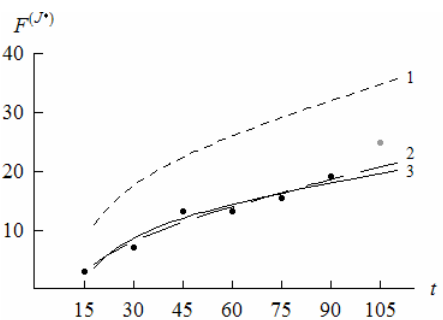
Для оценки точности моделей используем среднюю ошибку прогнозирования:



(а) прогнозирование на основе единичного наблюдения



(б) прогнозирование на основе трех наблюдений



(в) прогнозирование на основе шести наблюдений

Рис. 1. Индивидуальные значения (точки) и прогнозы индекса моторики

- 1 – средний тренд
- 2 – одномерная модель индивидуального прогнозирования
- 3 – многомерная модель индивидуального прогнозирования

$$S_F^{(J)} = \frac{1}{\Theta - \Theta_0} \sum_{\gamma=\Theta_0+1}^{\Theta} \left| \frac{F^{(J)}(t_\gamma) - y_\gamma^{(J)}}{y_\gamma^{(J)}} \right|, \tag{11}$$

где  $t \in [t_1; t_\Theta]$  – рассматриваемый временной интервал, включающий интервал основания прогноза и интервал прогнозирования;  $\Theta_0$  – количество наблюдений на интервале основания прогноза;  $\Theta$  – общее количество наблюдений;  $F^{(J)}(t_\gamma)$  – модельные данные;  $y_\gamma^{(J)}$  – реальные данные.

На Рис. 2 построены графики средней ошибки прогнозирования для группы пациентов с детским церебральным параличом, в которую входит пациент  $J^*$ .

Как видно из Рис. 2, прогнозы, полученные с применением моделей индивидуального прогнозирования, в целом точнее по сравнению с прогнозами по среднему тренду. Для индивидуальных прогнозных моделей характерно существенное уменьшение ошибки прогнозирования (более чем на 75 %) уже при небольшом количестве исходных данных (три наблюдения), тогда как ошибка прогнозирования по среднему тренду с ростом объема исходных данных уменьшается незначительно.

Начиная с момента времени  $t = 45$  мес. средняя ошибка прогнозирования для индивидуальных моделей снижается до 0,14 (одномерная модель) и 0,11 (многомерная модель). При дальнейшем увеличении интервала основания прогноза ошибка прогнозирования практически не изменяется.

При величине общего интервала наблюдения в 90 мес. ( $t \in [15; 105]$ ) момент времени  $t = 45$  мес. соответствует соотношению интервала основания прогноза к интервалу прогнозирования: 34 % к 66 %. Возможность получения достоверных прогнозов в условиях, когда период обучения модели меньше контрольного интервала, является отличительной чертой индивидуальных прогнозных моделей.

## Заключение

Разработан алгоритм идентификации параметров модели прогнозирования уровня двигательного развития – индекса моторики при последовательном накоплении результатов обследований.

Апробация алгоритма на выборке наблюдений за группой пациентов с церебральным параличом показала уменьшение средней ошибки прогнозирования более чем на 75 % при увеличении периода обучения моделей с 1 % до 34 % от величины общего интервала исследования. Так как дальнейшее расширение интервала основания прогноза не приводит к существенному изменению точности результатов, полученное соотношение может считаться наилучшим.

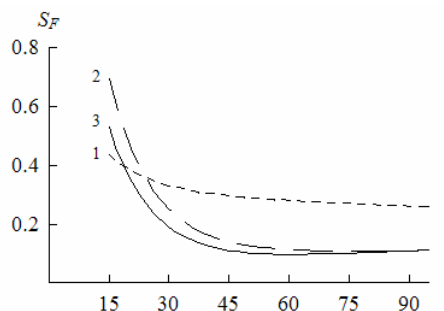


Рис. 2. Средняя ошибка прогнозирования

- 1 – средний тренд
- 2 – одномерная модель индивидуального прогнозирования
- 3 – многомерная модель индивидуального прогнозирования

Значение ошибки для данного периода обучения у многомерной модели прогнозирования индекса моторики на 21 % меньше, чем у одномерной модели.

Процесс уточнения индивидуальных моделей с ростом объема исходной информации должен сопровождаться пополнением базы данных всех обращений пациентов с аналогичными диагнозами, на основе которых должны также постоянно корректироваться групповые параметры моделей.

## Литература

1. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. Classification and Regression Trees. – Wadsworth, Belmont, CA, 1984. – 358 pp.
2. Fung Y.C. Biomechanics. Mechanical Properties of Living Tissues. Second Edition. Springer-Verlag. New York, 1993.
3. Keene O.N., Calverley P.M.A., Jones P.W. et al. Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited // Eur. Respir. J. -2008.-Vol.32.-P. 17-24.
4. Nagatomi Jiro. Mechanobiology, Handbook. CRC Press Taylor & Fransis Group, 2011.
5. Wildemann A.V., Tashkinov A.A., Bronnikov V.A. Locomotor Development Prediction Based on Statistical Model Parameters Identification // Computational and Mathematical Methods in Medicine. – Vol. 2012. - doi:10.1155/2012/548208.
6. Безруков Н.С. Еремин Е.Л., Перельман Ю.М. Автоматизированная система диагностики заболеваний легких // Проблемы управления. – 2007. – № 5. – С. 75–80.
7. Бронников В.А. Интегративная деятельность мозга в условиях дизнейроонтогенеза и при формировании функциональных систем двигательного анализатора у детей со спастическими формами церебральных параличей (перспективное исследование): автореф. дис... д-ра мед. наук. – Пермь, 2005. – 45 с.

8. Вильдеман А.В., Ташкинов А.А., Бронников В.А. Многомерный метод индивидуального прогнозирования индекса моторики // Информационные технологии и вычислительные системы. – 2010. – № 3. – С. 79 – 85.
9. Еремин Ю.А., Радченко В.П., Самарин Ю.П. Расчет индивидуальных деформационных свойств элементов конструкций в условиях ползучести // Машиноведение. – 1984. – № 1. – С. 67 – 72.
10. Няшин Ю.И., Лохов А.В. Основы биомеханики: Учебное пособие. – Пермь.: ПГТУ, 2008. – 209 с.
11. Реброва О.Ю. Применение методов интеллектуального анализа данных для решения задачи медицинской диагностики // Новости искусственного интеллекта. – 2004. – № 3. – С. 76 – 80.
12. Самарин Ю.П. Построение экспоненциальных аппроксимаций для кривых ползучести методом последовательного выделения экспоненциальных слагаемых // Проблемы прочности. – 1974. – № 9. – С. 24 – 27.
13. Суфиянов В.Г. Разработка адаптивных статистических моделей классификации и прогнозирования: автореф. дис... канд. физ.-мат. наук. – Пермь, 2004. – 16 с.
14. 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. – Т. 14, № 2 (48). – С. 69 – 78.

**Вильдеман Александр Валерьевич.** Директор Центра математического моделирования медико-социальных систем и процессов Пермского национального исследовательского политехнического университета. Окончил Пермский государственный технический университет в 2008 году. Кандидат физико-математических наук. Автор 27 научных работ. Область научных интересов: многомерный статистический анализ, математические модели прогнозирования, информационно-аналитические системы. E-mail: wildemann@mail.ru

**Ташкинов Анатолий Александрович.** Ректор Пермского национального исследовательского политехнического университета. Окончил Пермский политехнический институт в 1979 году. Доктор физико-математических наук, профессор. Автор более 230 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, 2 монографий. Область научных интересов: реабилитация инвалидов, качество жизни, детский церебральный паралич. E-mail: dolphinrehab@rambler.ru