

## ВЫВОДЫ

1. Средние показатели костной прочности у детей младшего школьного возраста находятся в пределах физиологических колебаний, определённых по России.

2. Потребление кальция с пищей соответствует 50–70% возрастной нормы, что характеризуется как умеренный дефицит. Выраженный дефицит кальция (потребление менее 50%) практически не встречается, однако рекомендуемое количество кальция в сутки употребляют лишь 11% детей, их доля на селе меньше (7,2%), чем в городе (15%).

3. Недостаток поступления с пищей кальция, вероятно, отражается на темпах роста детей: низкорослых детей меньше в городе, чем в сельской местности.

4. Количество потребляемых в день детьми молока и молочных продуктов существенно ниже возрастных норм.

## ЛИТЕРАТУРА

1. Доскин В.А., Келлер Х., Мураенко Н.М. и др. Морфофункциональные константы детского организма. Справочник. — М.: Медицина, 1997. — 288 с.

2. Рожинская Л.Я. Руководство по остеопорозу / Под ред. Л.И. Беневоленской. — М.: Бином. Лаборатория знаний, 2003. — С. 261–287.

3. Шеплягина Л.А., Рымарчук Г.В., Самохина Е.О. и др. Костная прочность у детей: известные и неизвестные факты. Методические рекомендации. — М., 2010. — 13 с.

4. Optimal Calcium Intake. Consensus development panel on optimal calcium intake // JAMA. — 1994. — Vol. 272. — P. 1942–1948.

УДК 519.254: 614.2: 612.087: 616-036.22

НО09

## ОПЫТ ПРИМЕНЕНИЯ ГРАФИЧЕСКОГО ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА В СРЕДЕ R ДЛЯ АНАЛИЗА КЛИНИЧЕСКИХ И ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Татьяна Ивановна Долгих, Дмитрий Александрович Сербяев, Герман Викторович Чекмарев\*,  
Татьяна Владимировна Кадцына

Омская государственная медицинская академия

## Реферат

**Цель.** Разработка программного продукта для анализа медицинских данных и представления показателей здравоохранения.

**Методы.** Разработана программа R\_MED — интерфейс для решения некоторых типовых задач анализа экспериментальных, клинико-лабораторных и эпидемиологических данных с использованием возможностей системы R.

**Результаты.** Функционально программа состоит из следующих блоков: «Загрузка данных», «Настройка», «Базовый расчёт», «Data Mining», «Представление показателей здравоохранения». Упрощение интерфейса достигается также включением в программу только тех методов анализа данных, которые наиболее часто необходимы при анализе медицинских данных. Так, в «Базовый расчёт» включены следующие процедуры статистического анализа: описательные статистики для количественных переменных, таблицы частот, гистограммы и бокс-графики, корреляционная матрица эллипсов (Пирсон), количественные корреляции Пирсона, порядковые корреляции Спирмена, корреляционная матрица эллипсов (Спирмен), диаграммы рассеяния на плоскости, определение различий в двух независимых групповых выборках с применением критериев Стьюдента и Манна-Уитни, дисперсионный анализ ANOVA. В «Настройке» пользователь может выбрать для анализа набор переменных и наблюдений, изменить набор свойств любой переменной, добавить, удалить, переименовать переменную, а также опционально настроить режим загрузки данных, базового расчёта и вывода. В программе предусмотрена возможность визуализации данных с помощью блока «Представление показателей здравоохранения» в разрезе «территория», «Год», «показатель». Территориальные разрезы могут дифференцироваться по трём уровням: муниципальный, субъектовый, окружной (федеральный округ). Для построения карт в системе R\_MED хранятся векторные данные по всем трём территориальным уровням, включая федеральные округа Российской Федерации.

**Вывод.** Первоначально разработанная для решения исследовательских задач по прогнозированию риска развития патологии программа R\_MED при соответствующей настройке также может быть использована в других клинико-диагностических и эпидемиологических исследованиях с целью мониторинга социально значимых заболеваний и деятельности служб здравоохранения, а также при подготовке годовых статистических отчётов, вплоть до уровня региона.

**Ключевые слова:** организация здравоохранения, анализ данных, статистика, система R, графический пользовательский интерфейс, представление показателей здравоохранения, Data Mining.

**EXPERIENCE IN DEVELOPING GRAPHICAL USER INTERFACE TO R PROGRAMMING LANGUAGE FOR CLINICAL AND EXPERIMENTAL DATA ANALYSIS** T.I. Dolgikh, D.A. Serbaev, G.V. Chekmarev, T.V. Kadcyna. Omsk State Medical Academy, Omsk, Russia. **Aim.** To develop the software product for of medical data analysis and public health indicators presentation. **Methods.** R\_MED software — an interface for typical experimental, clinical and laboratory, epidemiologic analysis using the R system opportunities — was developed. **Results.** Functionally, the program consists of the following blocks: «Load Data», «Settings», «Basic calculations», «Data Mining», «Presentation of health indicators».

Адрес для переписки: german\_1984@inbox.ru

Interface simplifying is also achieved by the inclusion of only those methods that are most often required in medical data analysis. So, the «Basic calculation» unit includes the following statistical calculations: descriptive statistics for quantitative variables, frequency tables, bar charts and box plots, Pearson's correlation matrix, Pearson's linear correlation, Spearman's rank correlation, Spearman's correlation matrix, 2D scatter plots, defining the difference in two independent sample groups using Student's test and the Mann-Whitney test, analysis of variance (ANOVA). In the «Settings» unit, a user can choose a set of variables and observations for analysis, to change the set of features for any value, to add, delete, rename the variable, and optionally customize the «Load Data» mode, the basic calculation, and data output. The program provides the ability to visualize data using «Presentation of health indicators» block in the context of territory, year and variant. Territorial cuts can be differentiated into three levels: municipal, regional, district (Federal District) level. To construct the maps, vector data on all 3 spatial levels are stored in the R\_MED system, including the Federal Districts of Russian Federation. **Conclusion.** Originally developed for research problems solving of pathology risk forecasting, the R\_MED program, if configured properly, can also be used in other clinical diagnostic and epidemiological studies to monitor problems of socially significant diseases and of health services, as well as in the preparation of annual statistical reports, including the regional level. **Keywords:** healthcare management, data analysis, statistics, R system, graphical user interface, health indicators presentation, Data Mining.

При выполнении научно-исследовательских работ необходимо высокое качество анализа данных. Наряду с этим при решении проблемы целесообразна оценка показателей в динамике, а при проведении эпидемиологического и социально-гигиенического мониторинга для оценки состояния здоровья населения в разрезах «территория-год-показатель» требуется программа, сочетающая в себе функции анализа данных и представления показателей с простым графическим пользовательским интерфейсом.

Большие возможности для этого предоставляет распространяемый на условиях стандартной общественной лицензии GNU программный продукт R (среда и язык программирования), предназначенный для статистического анализа данных [9]. Мощный язык программирования R, включающий более 4000 библиотек функций в исходном коде, в которых реализованы современные алгоритмы анализа и визуализации данных, позволяет создавать типовые исследовательские рабочие места, удовлетворяющие перечисленным выше требованиям. Создание графического пользовательского интерфейса в R по сути означает выбор из имеющегося в R подмножества функций для решения определённого круга задач и создание программы, обеспечивающей их комфортный для конкретного пользователя вызов, анализ и вывод результатов.

В центральной научно-исследовательской лаборатории ГБОУ ВПО «Омская государственная медицинская академия» Минздрава России была разработана программа R\_MED – интерфейс для решения некоторых типовых задач анализа экспериментальных, клинико-лабораторных и эпидемиологических данных с использованием возможностей системы R.

Функционально программа состоит из следующих блоков: «Загрузка данных», «Настройка», «Базовый расчёт», «Data Mining», «Представление показателей здравоохранения».

Основные решения, обеспечивающие интеллектуализацию системы, и, следовательно, упрощение пользовательского интерфейса, таковы:

– интеграция большого количества процедур статистического анализа в один расчётный блок «Базовый расчёт»: описательные статистики для количественных переменных, таблицы частот,

гистограммы и бокс-графики, корреляционная матрица эллипсов (Пирсон), количественные корреляции Пирсона, порядковые корреляции Спирмена, корреляционная матрица эллипсов (Спирмен), диаграммы рассеяния на плоскости, определение различий в двух независимых групповых выборках с применением критериев Стьюдента и Манна-Уитни, дисперсионный анализ ANOVA (рис. 1);

– возможность переменной иметь несколько типов (свойств): числовая, категориальная, упорядоченная, бинарная, целевая (зависимая), ключевая, группирующая, дата, год, территория, табличная (количество случаев); это позволяет выполнить «Базовый расчёт» или любую процедуру «Data Mining» без дополнительной настройки.

Упрощение интерфейса достигается также включением в программу только тех методов анализа данных, которые наиболее часто необходимы при анализе медицинских данных.

В «Настройке» пользователь может выбрать для анализа набор переменных и наблюдений, изменить набор свойств любой переменной, добавить, удалить, переименовать переменную, а также опционально настроить режим загрузки данных, базового расчёта и вывода. Загрузка данных происходит из файла базы данных, сформированной пользователем в программе Microsoft Excel. Объём выходных данных существенно зависит от задаваемого пользователем критического значения  $p$  ( $p_{кр}$ ). В процедурах с использованием критерия статистической значимости выдаются только те результаты, для которых  $p \leq p_{кр}$ .

В «Базовом расчёте» объединены следующие аналитические процедуры: анализ частот и пропусков для категориальных переменных, получение основных статистических количественных характеристик, определение возможных статистических связей между показателями и группами данных, визуализация данных. Для решения этих задач проводятся предварительный анализ, корреляционный анализ, анализ таблиц сопряжённости, проверка статистических гипотез, визуализация данных. Затем отбираются результаты с заданной в настройке степенью статистической значимости. Анализируемая выборка сначала в автоматическом режиме проверяется

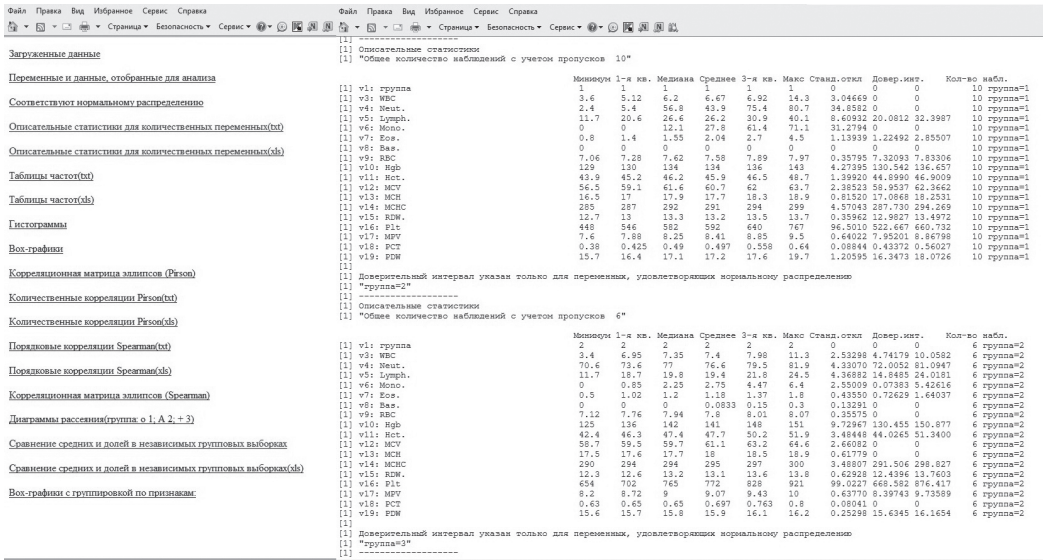


Рис. 1. Пример вывода результатов проведения базового расчёта при анализе экспериментальных данных общего анализа крови крыс.

на соответствие нормальному распределению (с использованием теста Шапиро), по результату проверки также автоматически определяются применяемые статистические методы (параметрические или непараметрические).

Групповые сравнения проводятся между группами, сформированными по значениям группирующей переменной. Если задано несколько группирующих переменных, то группы формируются путём комбинации значений различных группирующих переменных. Если группирующая переменная не задана, то группы для сравнения формируются из дихотомических переменных (имеющих только два значения). Предусмотрен режим создания группирующих переменных из переменных, имеющих более двух значений. Чтобы избежать ошибок, связанных с проблемой множественных сравнений, кроме значения p, выводится количество сравнений. Отчёты выдаются в текстовые и графические файлы, доступ к которым осуществляется через web-браузер [14]. Объединение множества аналитических статистических процедур в один расчёт также помогает обеспечить повторяемость исследования с другими данными, так как в этом случае требуется меньше количество исследовательских транзакций.

Блок «Data Mining» включает процедуры классификации, кластеризации, регрессии и др. Практически все процедуры этого блока требуют задания целевой (зависимой) переменной, которая либо означает зависимую переменную в регрессии, либо обозначает класс в задачах классификации.

Для кластерного анализа предлагается два метода: k-means [8] и иерархический кластерный. Для решения задач классификации включены следующие методы: «Линейный дискриминантный анализ», «Классификация по Naive

Bayes», «Метод опорных векторов» [13], «Дерево классификации» [11], «Случайный лес» [7]. Алгоритм «Случайный лес», кроме решения задач классификации, используется для определения значимости переменных для вероятной классификации.

Регрессионный анализ включает линейную регрессию, логистическую регрессию, авторегрессионный анализ. В линейной регрессии дополнительно автоматически отбирается наиболее оптимальный для регрессии набор независимых переменных по методу AIC (информационный критерий Акаике) [13]. В логистической регрессии одновременно после получения регрессионного уравнения применяется ROC-анализ [1, 3]. В авторегрессионном анализе, используемом для получения прогнозных значений, программой выбирается оптимальная длина временного ряда, используемого в прогнозе [2, 5].

Основные элементы «Представления показателей здравоохранения» — территория, год, показатель. Чаще всего территориальные разрезы могут дифференцироваться по трём уровням: муниципальный, субъектовый, окружной (федеральный округ) (рис. 2–4). Для построения карт в системе R\_MED хранятся векторные данные по всем трём территориальным уровням, включая федеральные округа Российской Федерации. Для отображения карт использовались функции библиотек R [6, 10, 12].

На любом территориальном уровне программа выдаёт таблицы, гистограммы, динамику показателя, карты (в том числе в интерактивном режиме), гистограмму показателей со статистически значимым ростом или убыванием показателя за выбранный период и др.

Включена кластеризация территорий с отображением кластеров на карте, гистограмм и в текстовой таблице. Предусмотрены сборка и вы-

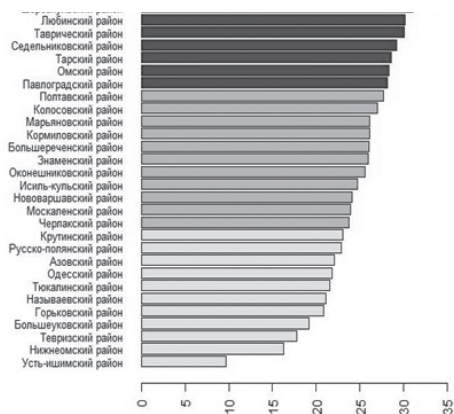


Рис. 2. Гистограмма средней заболеваемости раком желудка в районах Омской области за период 1996–2008 гг. с разбивкой районов на три кластера.

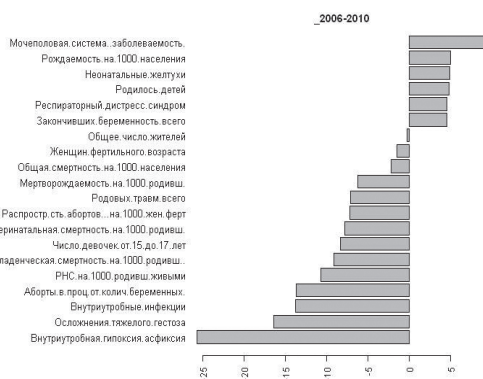


Рис. 3. Гистограмма статистически значимого роста/уменьшения значения показателей, связанных с родопомощением, по Омской области за период с 2006 по 2010 гг.



Рис. 4. Распределение первичной онкологической заболеваемости в 2010 г. по областям Российской Федерации.

числение значений показателей от территории нижнего уровня к территории верхнего уровня с учётом количества населения.

Таким образом, разработанная нами система позволяет создавать тематические исследовательские рабочие места для анализа данных и представления показателей здравоохранения с простым и удобным интерфейсом.

**ВЫВОД**

Первоначально разработанная для решения исследовательских задач по прогнозированию риска развития патологии программа R\_MED при соответствующей настройке также может быть использована в других клинико-диагностических, эпидемиологических исследованиях с целью мониторинга социально значимых заболеваний и деятельности служб здравоохранения, а также при подготовке годовых статистических отчётов, вплоть до уровня региона.

*Работа выполнена в рамках реализации Государственного задания Министерства здравоохранения Российской Федерации по теме: «Изучение*

*роли персистенции антигенов герпесвирусов и токсоплазм in vivo в развитии перинатальной патологии с оценкой нейроиммунных взаимодействий в системе "мать-плод"» (номер государственной регистрации №012001251438).*

**ЛИТЕРАТУРА**

1. Bowman E., Crawford G., Alexander G. et al. Rpanel: simple interactive controls for R functions using the tcltk package // J. Statistical Software. – 2007. – Vol. 17, N 9. – P. 1–18.
2. Carstensen M., Plummer E., Laara E. et al. Epi: a package for statistical analysis in epidemiology. R package version 1.1.34. – <http://CRAN.R-project.org/package=Epi> (дата обращения: 04.12.2012).
3. Chongsuvivatwong V. Epicalc: epidemiological calculator. R package version 2.14.1.6. – <http://CRAN.R-project.org/package=epicalc> (дата обращения: 04.12.2012).
4. Fox J. The R commander: a basic statistics graphical user interface to R // J. Statistical Software. – 2005. – Vol. 14, N 9. – P. 1–42.
5. Hyndman R.J., Razbash S. Drew Schmidt forecasting functions for time series and linear models. R package version 3.20. – <http://CRAN.R-project.org/package=forecast> (дата обращения: 04.12.2012).



6. Ihaka R, Murrell P, Hornik K. et al. Color space manipulation. R package version 1.1-1. — <http://CRAN.R-project.org/package=colspace> (дата обращения: 04.12.2012).

7. Liaw W.M. Classification and regression by random forest // R News. — 2002. — Vol. 2, N 3. — P. 18-22.

8. Maechler M, Rousseeuw P, Struyf A. et al. Cluster analysis basics and extensions. R package version 1.14.3. — <http://cran.r-project.org/web/packages/cluster/cluster.pdf> (дата обращения: 04.12.2012).

9. R Core Team. R: a language and environment for statistical computing. R foundation for statistical computing. — Vienna, Austria. — <http://www.R-project.org/> (дата обращения: 04.12.2012).

10. Richard A, Becker R, Wilks A.R. et al. Draw geographical maps. R package version 2.2-5. — <http://CRAN.R-project.org/package=maps> (дата обращения: 04.12.2012).

11. Ripley B. Classification and regression trees. R package version 1.0-29. — <http://cran.r-project.org/web/packages/tree/tree.pdf> (дата обращения: 04.12.2012), <http://CRAN.R-project.org/package=maptree> (дата обращения: 04.12.2012).

12. Roger S.B., Edzer P.J., Virgilio G.-R. Applied spatial data analysis with R. — Springer, NY, 2008. — 376 p.

13. Venables W.N., Ripley B.D. Modern applied statistics with S. Fourth Edition. — Springer, New York, 2002. — 495 p.

14. Verzani J.G. Widgets API for building toolkit-independent, interactive GUIs. Based on the iwidgets code of Simon Urbanek, suggestions by Simon Urbanek, Philippe Grosjean and Michael Lawrence gWidgets. R package version 0.0-52. — <http://CRAN.R-project.org/package=gWidgets> (дата обращения: 06.03.2013).

УДК 615.277.3: 616.62-006.6-037: 616.345-006.6-037: 616.152

НО10

## ПРОГНОСТИЧЕСКИЕ ФАКТОРЫ ЭФФЕКТИВНОСТИ ПРОТИВООПУХОЛЕВОЙ ТЕРАПИИ

Елена Ивановна Ерлыкина<sup>1</sup>, Татьяна Викторовна Копытова<sup>1</sup>, Анна Валерьевна Алясова<sup>1</sup>, Игорь Георгиевич Терентьев<sup>1</sup>, Татьяна Николаевна Горшкова<sup>2</sup>, Владимир Георгиевич Пименов<sup>3</sup>, Илья Игоревич Евдокимов<sup>3</sup>, Лариса Михайловна Обухова<sup>1\*</sup>

<sup>1</sup>Нижегородская государственная медицинская академия,

<sup>2</sup>Приволжский окружной медицинский центр, г. Нижний Новгород,

<sup>3</sup>Институт химии высококислотных веществ им. Г.Г. Девярых, г. Нижний Новгород

### Реферат

**Цель.** Оценка взаимосвязи биохимических показателей плазмы крови у больных солидными опухолями до и после проведения первого курса химиотерапии с наличием объективного ответа на лечение.

**Методы.** Изучена плазма крови 14 больных с рецидивом рака до лечения и после проведения первого курса специфической химиотерапии карбоплатином, метотрексатом, винбластином (при раке мочевого пузыря), а также иринотеканом, кальция фолинатом и фторурацилом (при раке кишечника). Первая группа включала пациентов с рецидивом рака толстой кишки: 3 мужчин и 4 женщины в возрасте 57–62 лет. Во вторую группу вошли пациенты с рецидивом рака мочевого пузыря: 5 мужчин и 2 женщины в возрасте 48–64 лет. В плазме крови оценивали свободнорадикальную активность, окислительную модификацию белков, степень эндогенной интоксикации, уровень макро- и микроэлементов.

**Результаты.** При достижении объективного эффекта химиотерапии у больных после первого курса лечения отмечено увеличение уровня  $\alpha_1$ - и  $\gamma$ -глобулинов, фосфора, цинка, лития и железа в плазме крови. В случае последующего прогрессирования заболевания была выявлена противоположная динамика данных показателей. При наличии объективного эффекта полихимиотерапии у больных, закончивших первый курс лечения, наблюдали значимое возрастание степени эндогенной интоксикации и снижение свободнорадикальной активности плазмы крови. При последующем прогрессировании опухолевого роста выявлена тенденция к увеличению свободнорадикальной активности плазмы крови после первого курса лечения, сопровождающаяся активацией суммарной окислительной модификации белков.

**Вывод.** Повышение содержания  $\alpha_1$ - и  $\gamma$ -глобулинов, фосфора в плазме крови после первого курса полихимиотерапии может быть использовано в качестве прогностического фактора эффективности противоопухолевой терапии; в качестве дополнительных аналитических факторов можно применять повышение степени эндогенной интоксикации, рост концентрации меди, железа, цинка, лития в плазме крови.

**Ключевые слова:** рак мочевого пузыря, рак кишечника, прогноз эффективности химиотерапии,  $\alpha_1$ -глобулин,  $\gamma$ -глобулин, фосфор.

**PROGNOSTIC FACTORS FOR ANTI-CANCER THERAPY EFFICIENCY** E.I. Erylykina<sup>1</sup>, T.V. Kopytova<sup>1</sup>, A.V. Alyasova<sup>1</sup>, I.G. Terentiev<sup>1</sup>, T.N. Gorshkova<sup>2</sup>, V.G. Pimenov<sup>3</sup>, I.L. Evdokimov<sup>3</sup>, L.M. Obukhova<sup>1</sup>. *1*Nizhny Novgorod State Medical Academy, Nizhny Novgorod, Russia, *2*Volga District Medical Centre, Nizhny Novgorod, Russia, *3*G.G. Devyatykh Institute of Chemistry of High-Purity Substances, Nizhny Novgorod, Russia. **Aim.** To investigate the relationship between blood plasma biochemical indicators in patients with solid tumors before a treatment and after the first course of chemotherapy with objective treatment response. **Methods.** Blood plasma samples taken from 14 patients with cancer relapse were studied before the treatment and after the first course of specific chemotherapy (carboplatin, methotrexate, vinblastine in patients with urine bladder cancer and irinotecan, leukovorin and fluorouracil in patients with colorectal cancer). The first group included patients with colorectal cancer relapse (males – 3, females – 4) aged 57–62 years. The