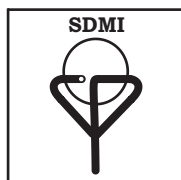


# INFORMATICA MEDICA SLOVENICA

- 1 Hiter in preprost algoritem za razdvoumljanje simbolov genov
- 9 Enhanced Electrostatic AChE Activity of Abnormally Hydrophobic Environment in Alzheimer's Plaques
- 15 Izzivi prenove spletišča Medicinske fakultete v Ljubljani
- 21 Analiza podatkov o ocenjevanju z Lestvico funkcijske neodvisnosti na Inštitutu Republike Slovenije za rehabilitacijo
- 33 Farmakogenomika in informacijski izzivi
- 41 Vabilo na kongres MI 2008 z mednarodno udeležbo: Od e-zdravja k zdravju



Revija Slovenskega društva za medicinsko informatiko  
Informatica Medica Slovenica  
LETNIK 13, ŠTEVILKA 1  
ISSN 1318-2129  
ISSN 1318-2145 on line edition  
<http://ims.mf.uni-lj.si>

## GLAVNI UREDNIK

Janez Stare

## SOUREDNIKA

Jure Dimec  
Blaž Zupan

## TEHNIČNI UREDNIK

Peter Juvan

## UREDNIŠKI ODBOR

Gregor Anderluh  
Valentin Fidler  
Emil Hudomalj  
Brane Leskošek  
Marjan Mihelin  
Mojca Paulin  
Borut Peterlin  
Uroš Petrovič  
Vladislav Rajkovič  
Gaj Vidmar

## BIVŠA GLAVNA UREDNIKA

Martin Bigec  
Peter Kokol

## O REVIJI

Informatica Medica Slovenica je interdisciplinarna strokovna revija, ki objavlja prispevke s področja medicinske informatike, informatike v zdravstvu in zdravstveni negi, ter bioinformatike. Revija objavlja strokovne prispevke, znanstvene razprave, poročila o aplikacijah ter uvajanju informatike na področjih medicine in zdravstva, pregledne članke in poročila. Še posebej so dobrodošli prispevki, ki obravnavajo nove in aktualne teme iz naštetih področij.

Informatica Medica Slovenica je strokovna revija Slovenskega društva za medicinsko informatiko. Revija je dostopna na naslovu <http://ims.mf.uni-lj.si>. Avtorji člankov naj svoje prispevke v elektronski obliki pošiljajo glavnemu uredniku po elektronski pošti na naslov [janez.stare@mf.uni-lj.si](mailto:janez.stare@mf.uni-lj.si). Revijo prejemajo vsi člani društva. Informacije o članstvu v društvu oziroma o naročanju na revijo so dostopne na tajništvu društva ([Drago Rudel](mailto:Drago.Rudel@mf.uni-lj.si), [drago.rudel@mf.uni-lj.si](mailto:drago.rudel@mf.uni-lj.si)).

## VSEBINA

*Izvirna znanstvena članka*

- 1 **Andrej Kastrin, Dimitar Hristovski**  
Hiter in preprost algoritem za razdvoumljanje simbolov genov
- 9 **Rakesh Sharma**  
Enhanced Electrostatic AChE Activity of Abnormally Hydrophobic Environment in Alzheimer's Plaques

*Strokovni članki*

- 15 **Branimir Leskošek, Andrej Lesjak**  
Izzivi prenove spletišča Medicinske fakultete v Ljubljani
- 21 **Gaj Vidmar, Helena Burger, Črt Marinček, Robert Cugelj**  
Analiza podatkov o ocenjevanju z Lestvico funkcijske neodvisnosti na Inštitutu Republike Slovenije za rehabilitacijo
- 33 **Polonca Ferik, Branimir Leskošek**  
Farmakogenomika in informacijski izzivi

*Bilten SDMI*

- 41 Vabilo na kongres MI 2008 z mednarodno udeležbo: Od e-zdravja k zdravju

Izvorni znanstveni članek ■

## Hiter in preprost algoritem za razdvoumljanje simbolov genov

## A fast and simple document classification algorithm for gene symbol disambiguation

---

Instituciji avtorjev: Inštitut za medicinsko genetiko, Univerzitetni klinični center (AK), Inštitut za biomedicinsko informatiko, Medicinska fakulteta, Univerza v Ljubljani (DH).

Kontaktna oseba: Andrej Kastrin, Inštitut za medicinsko genetiko, Univerzitetni klinični center, Šljajmerjeva 3, 1000 Ljubljana. email: andrej.kastrin@guest.arnes.si.

**Andrej Kastrin, Dimitar Hristovski**

**Izveček.** Razdvoumljanje simbolov genov je raziskovalno zelo aktualno področje. Največji problem predstavlja ločevanje med besednimi simboli, ki označujejo gene oz. njihove produkte ter simboli, ki se nanašajo na ostale biomedicinske koncepte (npr. CT, MR). V članku predstavimo hiter in preprost pristop k razdvoumljanju, ki temelji na razvrščanju MEDLINE® zapisov v genetsko in negenetsko domeno, ob predpostavki, da se simboli v genetski domeni najverjetneje nanašajo na gene. Algoritem sloni na statistični primerjavi domensko reprezentativnih korpusov zgrajenih na osnovi MeSH® deskriptorjev. Metoda je jasno razumljiva, preprosta za implementacijo ter računsko nezahtevna. Točnost razvrstitve, merjena na validacijski množici podatkov, je znašala 0,91. Algoritem je implementiran kot pomožni sistem za razdvoumljanje simbolov genov v sistemu za odkrivanje dejanskih ali potencialnih zakonitosti iz bibliografskih podatkov BITOLA.

**Abstract.** Gene symbol disambiguation is an important problem for biomedical text mining systems. When detecting gene symbols in MEDLINE® citations one of the biggest challenges is the fact that many gene symbols also denote other, more general biomedical concepts (e.g. CT, MR). Our approach to this problem is first to classify the citations into genetic and non-genetic domains and then to detect gene symbols only in the genetic domain. We used ontological information provided by Medical Subject Headings (MeSH®) for this classification task. The proposed algorithm is fast and is able to process the full MEDLINE distribution in a few hours. It achieves predictive accuracy of 0,91. The algorithm is currently implemented in the BITOLA literature-based discovery support system.

■ **Infor Med Slov:** 2008; 13(1): 1-8

## Uvod

Redno sledenje novim znanstvenim spoznanjem, rezultatom raziskav in razvoju tehnologij je celo za izkušenega raziskovalca z dobro informacijsko podporo zahteven zalogaj. Bibliografske zbirke rastejo iz dneva v dan; na področju biomedicine najpogosteje uporabljena zbirka MEDLINE® trenutno obsega že več kot 16 milijonov zapisov. Za učinkovit priklic relevantnih zapisov zato potrebujemo ustrezno zmogljive iskalnike. Uveljavljenim načinom poizvedovanja po tekstovnih podatkovnih zbirkah se je v zadnjem desetletju pridružilo področje z literaturo podprtega odkrivanja zakonitosti iz podatkov (*angl.* literature-based discovery), katerega namen je iskanje novih in potencialno uporabnih zakonitosti na osnovi implicitnih relacij med posameznimi bibliografskimi zapisi. Oče ideje je Don R. Swanson, ki je na ta način odkril, da ribje olje lahko služi kot učinkovito zdravilo pri zdravljenju Raynaudovega sindroma.<sup>1</sup> Skupaj s sodelavci je ponudil tudi orodje za odkrivanje potencialno zanimivih relacij v obliki prosto dostopne spletne storitve Arrowsmith.<sup>2</sup> Inovativen koncept so povzeli tudi drugi raziskovalci in ga na različne načine implementirali v svojih iskalnih sistemih.

BITOLA je interaktivni sistem za podporo odkrivanju dejanskih ali potencialnih zakonitosti iz bibliografskih podatkov na področju biomedicine in je plod domačega znanja.<sup>3</sup> Relacije med biomedicinskimi koncepti so opisane s pomočjo asociacijskih pravil. Kljub temu, da omogoča posplošeno odkrivanje novega znanja znotraj celotne biomedicinske domene, je zlasti uporaben za opisovanje relacij znanja med posameznimi geni in fenotipom oz. boleznimi.

V trenutni implementaciji sistem tekstovne simbole, ki se potencialno nanašajo na gene, ekstrahira iz naslovov in povzetkov posameznih MEDLINE zapisov. Problem pa predstavlja učinkovito ločevanje med simboli, ki dejansko označujejo gene oz. njihove produkte ter simboli, ki se nanašajo na ostale biomedicinske koncepte.

Ideja je, da najprej razvrstimo zapise na genetsko in negenetsko domeno,<sup>4</sup> pri čemer lahko upravičeno pričakujemo, da se bodo v genetski domeni simboli najverjetneje nanašali na posamezne gene. Z izrazom genetska domena označujemo množico MEDLINE zapisov, v kateri je verjetnost pojavljanja simbolov genov večja kot v katerikoli drugi množici zapisov.

Nalogo lahko opišemo kot problem razvrščanja MEDLINE zapisov na genetsko in negenetsko domeno na osnovi vsebine posameznega zapisa. Formalno imamo torej množico domen  $C = \{c_1, c_2\}$  ter množico zapisov  $D = \{d_1, d_2, \dots, d_n\}$ , kjer vsakemu paru  $(c_i, d_j; i = 1, 2 \text{ in } 1 \leq j \leq n)$  priredimo vrednost indikatorske spremenljivke 0 ali 1, odvisno od tega ali zapis  $d_j$  pripada domeni  $c_i$  ali ne:

$$F: C \times D \rightarrow \{0, 1\}.$$

Funkcijo  $F$ , ki priredi vrednost indikatorski spremenljivki, imenujemo klasifikator.<sup>5</sup>

V literaturi najdemo vrsto različnih pristopov k razvrščanju podobnih problemskih domen (npr. metoda podpornih vektorjev,  $k$  najbližjih sosedov, naivni Bayesov klasifikator, itd.), med katerimi so ene bolj, druge manj uspešne. Večina do sedaj uporabljenih klasifikatorjev delno ali v celoti temelji na tehnologiji nadzorovanega učenja, kjer nov zapis razvrstimo na podlagi znanja, ki smo ga pridobili na osnovi množice učnih primerov.

V članku predstavimo hiter in preprost pristop k razvrščanju zapisov, ki temelji na metodi statistične primerjave dveh domensko reprezentativnih korpusov. Metoda je jasno razumljiva, preprosta za implementacijo ter računsko nezahtevna. Preizkusi kažejo, da lahko celotno distribucijo MEDLINE bibliografske zbirke s predlagano metodo sprocesiramo v pičlih nekaj urah.

## Dosedanje raziskave

Razdvoumljanje večpomenskih besed (*angl.* word sense disambiguation) je proces iskanja ustrezne

interpretacije za pojavitev določene besede v danem kontekstu na osnovi množice možnih interpretacij, ki jih tej besedi lahko pripišemo.<sup>6</sup> Odličen pregled nad tem zelo širokim področjem ponujata Agirre in Edmonds.<sup>7</sup> Biomedicinska domena je z dvoumnimi besedami zelo bogata. Weber s sodelavci<sup>8</sup> npr. poroča, da metatezaver UMLS vsebuje kar 7400 dvoumnih konceptov.

V zadnjih letih je zelo aktualno, tako raziskovalno kot tudi aplikativno, področje razdvoumljanja simbolov genov (*angl.* gene symbol disambiguation).<sup>9</sup> Če za primer vzamemo poved<sup>10</sup> "The inverse association between MR and VEGFR-2 expression in carcinoma suggest a potential tumor-suppressive function for MR", je naloga sistema za razdvoumljanje, da poda oceno ali se simbol MR nanaša na gen za mineralokortikoidni receptor ali na slikanje z magnetno resonanco.

Obravava dvoumnih simbolov genov je z vidika računalniškega procesiranja kompleksen proces. Dosedanje raziskave na tem področju kažejo, da je dvoumnost problematična iz večih razlogov:<sup>11</sup> (i) simbol gena se lahko nanaša na različne organizme, (ii) simbol lahko označuje gen ali nek drug biomedicinski koncept, (iii) simbol lahko označuje gen, protein oz. nek drug biološki produkt ali (iv) pa se nanaša celo na različne gene.

Kljub temu, da se s razdvoumljanjem simbolov genov ukvarja več raziskovalnih skupin, problem avtomatskega in predvsem učinkovitega procesiranja še vedno ostaja odprt. Weber<sup>8</sup> je sestavil obsežen korpus biomedicinskih zapisov namenjen vrednotenju novih algoritmov in metod. Savova in sodelavci<sup>12</sup> so problem dvoumnih simbolov poskusili rešiti z uporabo metode razvrščanja v skupine. Schijvenaars in sodelavci<sup>13</sup> so razvili hiter algoritem razvrščanja, ki zahteva zelo malo učinkovitih podatkov. Humphreyeva<sup>4,14</sup> je problem dvoumnosti rešila z računanjem moči povezanosti med posameznimi besedami in širokopomenskimi deskriptorji MeSH® (Journal Descriptor Indexing). Liu je s sodelavci<sup>15</sup> predstavil dvostopenjsko metodo, kjer za vsak dvoumen simbol najprej zgradimo referenčni

reprezentativni korpus, ki ga nato na drugi stopnji uporabimo za gradnjo klasifikatorja. Xu in sodelavci<sup>11</sup> so na osnovi različnih podatkovnih virov vsakemu simbolu gena priredili reprezentativni meta-profil, s čimer so dosegli boljšo diskriminativnost simbolov med seboj. Nedavno je Farkas<sup>16</sup> objavil zanimivo študijo, v kateri se je problema dvoumnosti lotil z analizo omrežij citiranosti posameznih avtorjev.

V članku najprej predstavimo teoretični okvir in gradnjo predlaganega klasifikatorja. Nato predstavimo rezultate eksperimentalnega dela. Končamo s kratko razpravo in idejami za nadaljnje delo.

## Metoda

### Gradnja reprezentativnih korpusov

Vsak MEDLINE zapis je označen z množico deskriptorjev MeSH, s katerimi opišemo vsebino posameznega zapisa. Iz obsežnega geslovnika MeSH, ki v letu 2008 obsega 24.767 deskriptorjev, izkušeni ocenjevalci vsakemu zapisu v povprečju priredijo 12 MeSH deskriptorjev.

Algoritem poleg korpusa genetske domene zahteva tudi podatke referenčnega korpusa. Referenčni korpus smo zgradili na osnovi polne distribucije bibliografske zbirke MEDLINE (MEDLINE Baseline Repository), ki je do konca leta 2007 vsebovala 16.880.015 zapisov. Distribucija je na voljo v XML zapisu, zato smo zaradi lažje obdelave podatkov s pomočjo razčlenjevalnika prebrali relevantne elemente ter jih prepisali v relacijski podatkovni model (t.j. ena vrstica za vsak deskriptor MeSH v vsakem MEDLINE zapisu). S preštevanjem smo potem izdelali frekvenčno porazdelitev deskriptorjev v referenčnem korpusu.

Korpus genetske domene smo pripravili na osnovi referenčnega korpusa, pri čemer smo upoštevali le tiste zapise, ki so reprezentativni za področje genetike. Pri tem smo si pomagali z datoteko "gene2pubmed" z Entrezovega<sup>17</sup> spletnega

skladišča, ki obsega seznam MEDLINE indentifikatorjev zapisov v katerih se pojavljajo simboli genov. S preštevanjem smo potem izdelali frekvenčno porazdelitev deskriptorjev v korpusu genetske domene.

### Algoritem razvrščanja

Vsakemu deskriptorju MeSH smo priredili vrednost hi-kvadrat statistike ter testirali domnevo o statistično značilni povezanosti opazovanih frekvenc v domeni genetskega in referenčnega korpusa. Hi-kvadrat statistiko smo izračunali po enačbi:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}; E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

kjer je  $O_i$  opazovana frekvenca,  $E_i$  pričakovana frekvenca,  $N_i$  pa skupna frekvenca MeSH deskriptorjev, bodisi v korpusu genetske domene bodisi v referenčnem korpusu. Ničelna hipoteza je bila, da opazovani frekvenci danega deskriptorja v domeni genetskega in referenčnega korpusa nista statistično značilno soodvisni.

V naslednjem koraku smo vsakemu deskriptorju pripisali pozitiven predznak indikatorske spremenljivke (*Ind*), če je bila relativna opazovana frekvenca v korpusu genetske domene večja kot relativna opazovana frekvenca deskriptorja v referenčnem korpusu in obratno, deskriptor je dobil negativen predznak indikatorske spremenljivke, če je bila relativna frekvenca opazovane frekvence v referenčnem korpusu večja kot v korpusu genetske domene.

Algoritem za razvrščanja zapisov potrebuje dva vhodna podatka:

1. Tabela frekvenčnih profilov z vsemi deskriptorji, pri katerih je razlika med korpusoma statistično značilno različna, z označenim predznakom indikatorske spremenljivke.

2. Množico zapisov, ki jih je potrebno razvrstiti.

Deskriptorje, katerih relativna zastopanost je v celotni bibliografski zbirki zelo visoka (npr. Humans, Animals, Mice, itd.) in s tega vidika ne pripomorejo k večji diskriminativnosti algoritma, smo predhodno odstranili iz procesa razvrščanja. Seznam praznih deskriptorjev smo povzeli po uradnem MEDLINE seznamu. Tabela frekvenčnih profilov indeksira frekvence deskriptorjev v korpusu genetske domene ter referenčnem korpusu, hi-kvadrat vrednosti razlik med njima ter vrednost indikatorske spremenljivke. Proces indeksiranja je prikazan na Sliki 1, primer tabele frekvenčnih profilov za množico vzorčnih deskriptorjev pa v Tabeli 1.

Algoritem začne z branjem MEDLINE zapisa  $z$  in piše vrednost skupnega odločitvenega dosežka po shemi:

*Dosežek* ( $z$ ) = 0

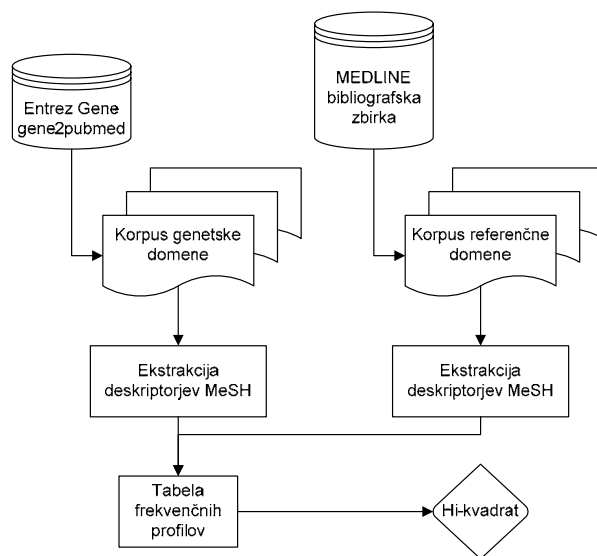
za vsak deskriptor MeSH  $m$

če *Ind*( $m$ ) pozitiven indikator

*Dosežek* ( $z$ ) = *Dosežek* ( $z$ ) + 1

sicer *Ind*( $m$ ) negativen indikator

*Dosežek* ( $z$ ) = *Dosežek* ( $z$ ) - 1



**Slika 1.** Diagram poteka priprave tabele frekvenčnih profilov na osnovi virov informacij.

**Tabela 1.** Struktura tabele frekvenčnih profilov.

<i>MeSH Deskriptor</i>	$O_B$	$O_G$	$E_B$	$E_G$	$\chi^2$	<i>Ind</i>
Amyotrophic Lateral Sclerosis	6133	224	6196	161	25	+
Animals	3615845	133463	3654231	95077	15901	+
Computer Simulation	36544	326	35935	935	407	-
Elasticity	16272	48	15906	414	332	-
Genes, Recessive	10948	1269	11907	310	3047	+
Guanine Nucleotide Exchange Factors	1568	685	2196	57	7080	+
Humans	8544765	108109	8433449	219425	57941	-
Mice	691495	77905	749889	19511	179314	+
Models, Chemical	29588	608	29430	766	33	-
Models, Molecular	68781	5425	72324	1882	6845	+
Models, Statistical	19630	60	19191	499	397	-
Mutation	178885	23627	197377	5135	68316	+
Polymers	26724	170	26212	682	394	-
Stress, Mechanical	25706	320	25366	660	180	-
Surface Tension	2976	12	2912	76	55	-

Pojasnilo:  $O_B$  – opazovana frekvenca v korpusu referenčne domene;  $E_G$  – pričakovana frekvenca v korpusu genetske domene;  $\chi^2$  – hi-kvadrat statistika; *Ind* – indikatorska spremenljivka.

Na izhodu dobimo množico zapisov, katerim je pripisana vrednost odločitvenega dosežka. Višji odločitveni dosežek imajo zapisi, pri katerih obstaja večja verjetnost, da pripadajo genetski domeni.

### Ocenjevanje natančnosti razvrščanja

Pražno vrednost, ki ločuje med zapisi obeh domen, ocenimo na testni množici zapisov, natančnost razvrstitve pa potem preverimo na neodvisni validacijski množici zapisov. Iz bibliografske zbirke MEDLINE smo najprej naključno izbrali vzorec 200 zapisov; prvo polovico zapisov smo uvrstili v testno, drugo polovico pa v validacijsko množico. Vsak zapis sta ocenila dva kompetentna ocenjevalca in mu pripisala indikator domene, ki je označeval reprezentativnost zapisa, bodisi za genetsko bodisi za negenetsko domeno. Skladnost med ocenjevalcema smo ovrednotili s pomočjo kappa ( $\kappa$ ) koeficienta.<sup>18</sup> V testni množici je znašalo ujemanje med ocenjevalcema  $\kappa = 0,81$ , v validacijski množici pa  $\kappa = 0,91$ . Zapise, pri katerih se je ocena razlikovala, sta ocenjevalca ponovno pregledala in končno oceno podala s soglasjem. Obe množici sta prosto dostopni pri avtorjih.

Empirično pražno vrednost odločitvenega dosežka, ki ločuje med zapisi genetske in negenetske domene smo določili z optimizacijo točnosti razvrstitve (*angl.* accuracy). Točnost (*Acc*) razvrstitve izračunamo kot razmerje med vsoto pravih razvrstitev ( $TP + TN$ ) in skupnim številom zapisov, ki jih razvrščamo ( $TP + FP + FN + TN$ ). *TP* (*angl.* true positive) označuje število pravilno razvrščenih pozitivnih zapisov (zapisi poročajo o genetiki in so razvrščeni v domeno genetike); *FP* (*angl.* false positive) označuje število napačno razvrščenih negativnih zapisov (zapisi so razvrščeni v domeno genetike, dejansko pa poročajo o nečem drugem); *FN* (*angl.* false negative) se nanaša na število napačno razvrščenih pozitivnih zapisov (zapisi so napačno razvrščeni v domeno, ki ni reprezentativna za genetiko), *TN* (*angl.* true negative) pa označuje število pravilno razvrščenih negativnih zapisov (zapisi so pravilno razvrščeni v domeno, ki ni reprezentativna za genetiko). Pražno vrednost lahko določimo tudi poljubno in s tem spreminjamo razmerje med napačno pozitivnimi in napačno negativnimi zapisi. Poleg natančnosti razvrstitve smo izračunali še priklic (*angl.* recall), natančnost (*angl.* precision) ter povprečno mero priklica in natančnosti *F*. Priklic predstavlja razmerje med pravilno razvrščenimi zapisi in

množico vseh zapisov v testni domeni, ki pripadajo domeni genetike. Natančnost pa predstavlja razmerje med pravilno razvrščenimi zapisi in vsemi zapisi, ki jih je algoritem razvrstil v domeno genetike. Priklic (*Rec*) in natančnost (*Pre*) izračunamo po naslednjih enačbah:

$$Rec = \frac{TP}{TP + FN} \text{ in } Pre = \frac{TP}{TP + FP} .$$

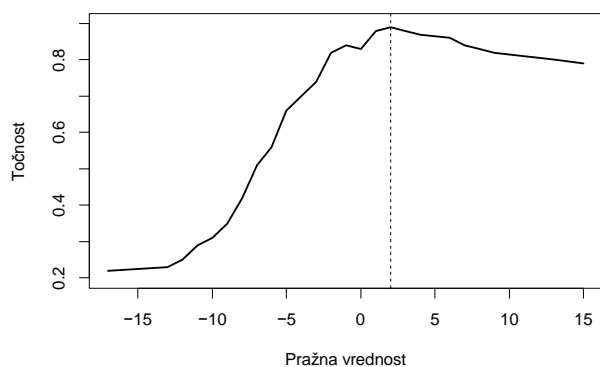
*F* statistiko izračunamo kot harmonično sredino priklica in natančnosti po enačbi:

$$F = \frac{2 \times Rec \times Pre}{Rec + Pre} .$$

Razpon vseh mer je v intervalu [0,1], pri čemer 1 označuje maksimalno kvaliteto klasifikatorja.

## Rezultati

Na Sliki 2 je prikazana točnost razvrstitve v odvisnosti od prazne vrednosti oz. porazdelitve odločitvenega dosežka. V našem primeru smo kot prazno vrednost definirali  $\theta = 2$ , pri kateri je bila točnost razvrstitve maksimalna. Prazne vrednosti med poskusi nismo več spreminjali. Vse zapise, pri katerih je bila vrednost odločitvenega dosežka  $\theta \geq 2$ , smo uvrstili v genetsko domeno.



**Slika 2.** Točnost razvrstitve zapisov v odvisnosti od prazne vrednosti na testni množici zapisov.

Proces razvrščanja si oglejmo na primeru. Izbrali smo dva MEDLINE zapisa, katerih naslova sta

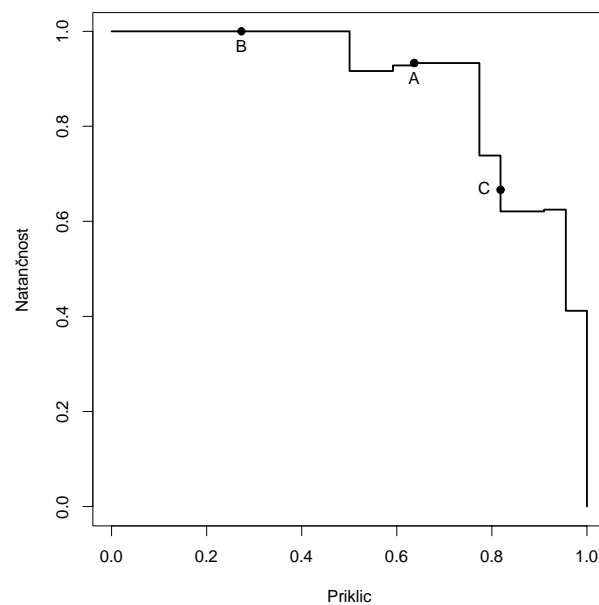
“Strain-dependent localization, microscopic deformations, and macroscopic normal tensions in model polymer networks” (*PMID*: 15697942) in “Recessive motor neuron diseases: mutations in the ALS2 gene and molecular pathogenesis for the upper motor neurodegeneration” (*PMID*: 15651293). Prvi zapis vsebuje osem, drugi pa sedem deskriptorjev MeSH. Trije deskriptorji v drugem zapisu so prazni (Animals, Humans in Mice), zato jih izločimo iz nadaljnjega procesiranja. Glede na vsak deskriptor povečamo oz. zmanjšamo vrednost odločitvenega dosežka skladno z vrednostjo indikatorske spremenljivke (Tabela 1). Končni rezultat sta dva odločitvena dosežka:

*Dosežek (PMID: 15697942)* = -6

*Dosežek (PMID: 15651293)* = 4.

Glede na zgoraj definirano prazno vrednost lahko prvi zapis uvrstimo v negenetsko domeno, drugi zapis pa v genetsko domeno.

Učinkovitost klasifikatorja smo preverili na validacijski množici zapisov. Na Sliki 3 so prikazane vrednosti priklica in natančnosti za različne prazne vrednosti parametra  $\theta$ .



**Slika 3.** Graf odnosa med priklicem in natančnostjo razvrstitve na validacijski množici zapisov.



Povedena metoda je dosegla  $Acc = 0,91$  točnost razvrstitve, pri čemer je priklic znašal  $Rec = 0,64$ , natančnost pa  $Pre = 0,93$ . Harmonična sredina priklica in natančnosti je znašala  $F = 0,76$ .

Graf poteka skozi dve točki. Točka (0,1) označuje klasifikator, ki ne prepozna genetsko relevantnih zapisov. Vsi negenetski zapisi so v tem primeru razvrščeni pravilno, genetski zapisi pa napačno. V točki (1,0) klasifikator vse zapise razvrsti kot genetsko relevantne. Genetski zapisi so v tem primeru razvrščeni pravilno, vsi negenetski zapisi pa napačno. Optimalno točnost oz. optimalno razmerje med priklicem in natančnostjo je v našem primeru klasifikator dosegel v točki A. Če prazno vrednost povečamo ( $\theta_B = 5$ ), bomo zapise razvrstili z večjo natančnostjo ( $Pre = 1,00$ ), na račun katere pa se bo zmanjšal priklic ( $Rec = 0,27$ ). Do obratnega učinka pride v točki C, kjer prazno vrednost zmanjšamo ( $\theta_C = -1$ ). Dokumenti so v tem primeru razvrščeni z manjšo natančnostjo ( $Pre = 0,67$ ), priklic pa je višji ( $Rec = 0,82$ ).

## Zaključki

V članku smo predstavili preprost in hiter algoritem za razvrščanje MEDLINE zapisov na osnovi deskriptorjev MeSH. Trenutno je algoritem implementiran kot pomožni sistem za razdvoumljanje simbolov genov v interaktivnem sistemu za podporo odkrivanju zakonitosti iz bibliografskih podatkov BITOLA. Eksperimentalni rezultati potrjujejo razmeroma visoko napovedno točnost klasifikatorja ( $Acc = 0,91$ ). V odvisnosti od namena poizvedovanja ter željenega razmerja med priklicem in natančnostjo, lahko prazno vrednost tudi spreminjamo.

V nadaljnjih raziskavah bomo preizkusili klasifikatorje z vključitvijo kompleksnejših prediktorskih spremenljivk: (i) kvalifikatorjev MeSH, ki podrobneje omejuje vsebinski obseg posameznega deskriptorja, (ii) parov deskriptor/kvalifikator ter (iii) prostim besedilom naslova in povzetka MEDLINE zapisa. V gradnji je

tudi obsežnejši označeni korpus zapisov, ki bo omogočal bolj zanesljivo in veljavno vrednotenje uporabljenih klasifikatorjev.

## Zahvala

Avtorja se zahvaljujeva Susanne M. Humphrey in Thomasu C. Rindfleschu za koristne napotke in vzpodbudne komentarje. Raziskovalna sredstva je zagotovila Javna agencija Republike Slovenije za raziskovalno dejavnost (J3-7411).

## Literatura

1. Swanson DR: Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect Biol Med* 1986; 30(1): 7-18.
2. Swanson DR, Smalheiser NR: An Interactive System for Finding Complementary Literatures: A Stimulus to Scientific Discovery. *Artif Intell* 1997; 91(2): 183-203.
3. Hristovski D, Stare J, Peterlin B, et al.: Supporting discovery in medicine by association rule mining in MEDLINE and UMLS. *Medinfo* 2001; 10(2): 1344-1348.
4. Hristovski D, Peterlin B, Mitchell JA, et al.: Using literature-based discovery to identify disease candidate genes. *Int J Med Inform* 2005; 74(2-4): 289-298.
5. Feldman R, Sanger J: *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge 2006: Cambridge University Press.
6. Manning CD, Schuetze H: *Foundations of statistical natural language processing*. Cambridge 2003: MIT Press.
7. Agiree E, Edmonds P (ur.): *Word sense disambiguation: Algorithms and applications*. Berlin 2006: Springer.
8. Weeber M, Schijvenaars BJ, Van Mulligen EM, et al.: Ambiguity of human gene symbols in LocusLink and MEDLINE: creating an inventory and a disambiguation test collection. *Proc AMIA Symp* 2003; 704-708.
9. Chen L, Liu H, Friedman C: Gene name ambiguity of eukaryotic nomenclatures. *Bioinformatics* 2005; 21(2): 248-256.
10. Di Fabio F, Alvarado C, Majdan A, et al.: Underexpression of mineralocorticoid receptor in colorectal carcinomas and association with

- VEGFR-2 overexpression. *J Gastrointest Surg* 2007; 11(11): 1521-1528.
11. Xu H, Fan JW, Hripcsak G, et al.: Gene symbol disambiguation using knowledge-based profiles. *Bioinformatics* 2007; 23(8): 1015-1022.
  12. Savova G, Pedersen T, Purandare A, et al.: Resolving ambiguities in biomedical text with unsupervised clustering approaches. Research Report UMSI 2005/80 and CB Number 2005/21; Minneapolis, Minnesota 2005: University of Minnesota Supercomputing Institute.
  13. Schijvenaars BJ, Mons B, Weeber M, et al.: Thesaurus-based disambiguation of gene symbols. *BMC Bioinformatics* 2005; 6: 149.
  14. Humphrey SM, Rogers WJ, Kilicoglu H, et al.: Word sense disambiguation by selecting the best semantic type based on Journal Descriptor Indexing: preliminary experiment. *J Am Soc Inform Sci Tech* 2006; 57(1): 96-113.
  15. Liu H, Lussier YA, Friedman C: Disambiguating ambiguous biomedical terms in biomedical narrative text: an unsupervised method. *J Biomed Inform* 2001; 34(4): 249-261.
  16. Farkas R: The strength of co-authorship in gene name disambiguation. *BMC Bioinformatics* 2008; 9: 69.
  17. Maglott D, Ostell J, Pruitt KD, et al.: Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2007; 35(Database issue): D26-31.
  18. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20(1): 37-46.

Research Paper ■

# Enhanced Electrostatic AChE Activity of Abnormally Hydrophobic Environment in Alzheimer's Plaques

**Rakesh Sharma**

**Abstract.** Computational techniques based on Brownian dynamics simulation speculate the electrostatic and kinetic properties of Acetylcholinesterase (AChE) in the range of dielectric constant values 60-78. Typically the enzyme specific to Alzheimer's Disease has specificity constant ( $k_{cat}/K_m$ ) of AChE at both dielectrics constants. AChE is distributed heterogeneously within the senile plaques and neurofibrillary tangles (NFTs) of Alzheimer's Disease (AD). The role of decreased brain AChE has been highlighted as responsible for specific Magnetic Resonance Spectroscopic pattern profiles in AD. These findings strengthen the suggestion that the altered AChE in AD may be associated with specific MR spectral peak patterns. This phenomenon may account for the reduced levels of the ACh neurotransmitter in the cerebral tissue of Alzheimer's disease (AD) patients. These results can be useful in clinical trials on AD lesions.

■ **Infor Med Slov:** 2008; 13(1): 9-14

---

Author's institution: Department of Chemical & Biomedical Engineering, FAMU-FSU College of Engineering, Tallahassee, Florida, USA.

Contact person: Rakesh Sharma, 901 West Jefferson Street, Jeffwood B-7, Tallahassee, FL 32304, USA. email: rs2010@columbia.edu.

## Introduction

### Alzheimer's disease

Alzheimer's disease (AD) has manifestations of senile plaques and neurofibrillary tangles (NFTs) in the cerebral cortex involving hippocampus of Alzheimer's brains. Many studies have shown that the levels of the reduced ACh neurotransmitter in AD brain had been curiosity in recent past. In this direction, promising success is claimed in drug therapeutic trials based on AChE inhibitors. In AD presumably five neurological cell groups are commonly seen around the cortex rich with AD lesions associated with the low AChE enzyme. Other important problem of AD therapeutics was solved by use of higher concentrations of AChE inhibitor, such as tacrine, physostigmine, and BW284C51, to inhibit AChE within AD lesions. Simultaneously, elevated levels of the ACh substrate also inhibit AChE activity. Serotonin, 5-hydroxytryptophan, carboxypeptidase inhibitor, and bacitracin have proven good choice to effectively inhibit cholinesterase activity within plaques and tangles, but fail to alter the AChE activity in normal tissue at standard physiological conditions.

This paper offers electrostatic explanation of lowered AChE catalytic activity in AD brain tissue over normal tissue with possibility of the protein-rich deposits associated with the onset of AD. The high concentration of protein-rich deposits in plaques and NFTs such as alpha amyloid protein ( $\alpha$ AP), heparan and dermatan sulfate proteoglycans, serum amyloid P component, complement factors, and protein kinase C, has been an active research topic together with the increased hydrophobicity of AChE in AD lesions and abnormally reduced dielectric constant. AChE catalysis has been explained as electrostatic screening mechanism where altered dielectric conditions correspond to  $\beta$ AP deposition. Possibly, dielectric constant shift in AD tissue also allows Coulombic interactions to permeate longer distances resulting in enhanced enzymatic activity and simultaneously decreased ACh levels. MR

spectral pattern of enhanced choline also supports association with decreased ACh levels in AD.

### Acetylcholinesterase

The acetylcholinesterase enzyme (AChE) has 537 amino acid long polypeptide in the postsynaptic neural membranes of central nervous system and neuromuscular junctions by a glycosylphosphatidylinositol linkage. AChE catalyzes the hydrolysis of the acetylcholine (ACh) substrate neurotransmitter at cholinergic synapses. AChE hydrolysis results in the termination of impulse transmission.

The determination of the three-dimensional structure of AChE dimer enzyme comprises 12-stranded mixed  $\beta$ -sheet surrounded by 14  $\alpha$ -helices. These subunits assemble through disulfide linkage and hydrophobic interactions. The enzyme structure shows structural characteristic of AChE as a deep ( $\sim 20\text{\AA}$ ), narrow active site making enzyme's catalytic site Ser<sup>200</sup>, His<sup>440</sup>, and Glu<sup>327</sup> at its base. The walls of this entity are lined with 14 highly conserved aromatic amino acids of active site. Positively charged ACh substrate towards the active site causes low-affinity cation- $\pi$  interactions. Further, amino acid charge distribution over AChE creates an electric field around the enzyme contributing to its enzymatic activity (electrostatic steering mechanism) involving its substrate, ACh. Authors determined that the negative field drives the positively-charged ACh substrate molecule toward the entrance of its active site moiety and increases the catalytic rate of ACh.

## Theory

### Electrostatics

If partial atomic charges are assigned according to the OPLS parameter set at neutral pH, generating a net charge of  $-10e$ ,<sup>1</sup> Coulombic electrostatic forces between the AChE monomer and ACh may be calculated by centering the enzyme in a  $64^3$  grid

lattice with 4 Å spacing and using the following equation to compute the electrostatic force at each point on the grid:

$$F(r) = \frac{e^2}{\epsilon} \sum_{i=1}^N \frac{q_i d_i}{d_i^3} (1 + \kappa d_i) e^{(-\kappa d_i)} \quad (1)$$

where  $e$  is the charge on an electron,  $\epsilon$  is the dielectric constant of the system,  $N$  is the number of atoms in the enzyme,  $q_i$  is the charge of the  $i^{\text{th}}$  enzyme atom,  $d_i$  is the distance between the  $i^{\text{th}}$  enzyme atom and the grid point, and  $\kappa$  is the Debye-Hückel ion screening parameter which approximately accounts for the perturbation of solvent ions on charge-charge interactions between the enzyme and its substrate. This ion screening parameter is defined as

$$\kappa = \sqrt{\frac{2N_A \rho e^2 I}{\epsilon_0 \epsilon k_B T}} \quad (2)$$

where  $N_A$  is Avagadro's number,  $\rho$  is the density of the solvent,  $I$  is the ionic strength of the solvent,  $\epsilon_0$  is the permittivity of vacuum,  $k_B$  is Boltzmann's constant, and  $T$  is the absolute temperature of the system.

### Brownian dynamics

Diffusion-controlled rate constants were obtained by simulating the reaction of AChE with an ensemble of ACh substrate molecules using the theory of Brownian dynamics. This was accomplished by dividing the space around AChE with a spherical surface of radius  $b$  and another with a larger radius  $q$ , such that all electrostatic interactions beyond  $b$  were centrosymmetric and substrate flux was isotropic.

The diffusional trajectory of an ACh molecule was calculated by randomly placing it on the  $b$  sphere and using the following Ermak-McCammon algorithm to compute subsequent motion:

$$r = r^o + \frac{D_{rel} F(r) \Delta t}{k_B T} + S \quad (3)$$

where  $r^o$  and  $r$  are the locations of ACh before and after a trajectory step is taken, respectively,  $D_{rel}$  is the relative diffusion constant,  $\Delta t$  is the length of a trajectory time step, and  $S$  represents a stochastic force due to collisions with solvent molecules having the statistical properties of a Gaussian distribution with zero mean and variance equal to  $2D_{rel}\delta_{ij}\Delta t$ , where  $\delta_{ij}$  is the Kronecker delta.<sup>2</sup> The electrostatic force on the substrate was calculated by interpolating the pre-calculated force values of the nearest grid points on the lattice by using a three-dimensional Taylor's expansion and solving the derivatives with the method of finite differences.<sup>3</sup> A trajectory terminated once the substrate diffused to within a certain distance of the enzyme's active site or beyond the established  $q$  sphere. Numerous trajectories were computed to yield a reaction probability of  $\beta$ , which was utilized in the following expression to solve for the diffusion-controlled rate constant:

$$k = k_D(b)\beta[1 - (1 - \beta)\Omega]^{-1} \quad (4)$$

where  $\Omega$  is the probability of a substrate diffusing past the  $q$  sphere and then returning and  $k_D(b)$  is the steady-state rate constant at which the ACh molecules penetrate the  $b$  sphere.<sup>4</sup> The following expression has been derived to describe this term:

$$k_D(b) = 4\pi D_{rel} \left( \int_b^\infty \frac{e^{\frac{U}{k_B T}}}{r^2} dr \right)^{-1} \quad (5)$$

where  $U$  is the intermolecular potential of mean force and  $r$  is equivalent to the radius of the constructed  $b$  sphere.<sup>5</sup>

### Model system

The use of the *Torpedo californica* AChE crystal structure for studying the pathogenesis of AD is supported by the primary sequence homology of

the electric eel and human enzymes.<sup>6</sup> Molecular modeling was used to determine that the hydrodynamic radius of the AChE monomer and ACh were  $\sim 30$  Å and 4.23 Å, respectively. However, to approximate the active site wall fluctuations of the narrow gorge of AChE, the substrate was modeled as a positively charged sphere with a radius of 2 Å.<sup>7</sup> To verify the dependence of the rate constant on the substrate radius parameter, test simulations were performed at various radii values. All non-hydrogen protein atoms were modeled as spheres with radii of 1.5 Å.<sup>3</sup> In addition, the  $b$  and  $g$  spherical radii were assigned distances of 128 Å and 295 Å, respectively. Reaction criteria were satisfied when the substrate sphere diffused within 15 Å of the OG atom of the Ser<sup>200</sup> catalytic residue. A variable time step algorithm was used with  $\Delta t$  values ranging from 1.25 fs to 800 ps as the substrate approached the outer  $g$  sphere.

Simulations were performed using 8000 trajectories at a temperature of 300 K, an ionic strength of 5 mM, a solvent density of 996.5 kg/m<sup>3</sup>, and dielectrics of 78 and 60. The Brownian dynamics program written for this work was compiled for use on a personal desktop computer, which presented operating system memory allocation problems. This program circumvents this limitation by utilizing a virtual addressing method.<sup>8</sup> All calculations were performed on a Packard Bell personal computer equipped with an Intel Pentium processor.

## Results and discussion

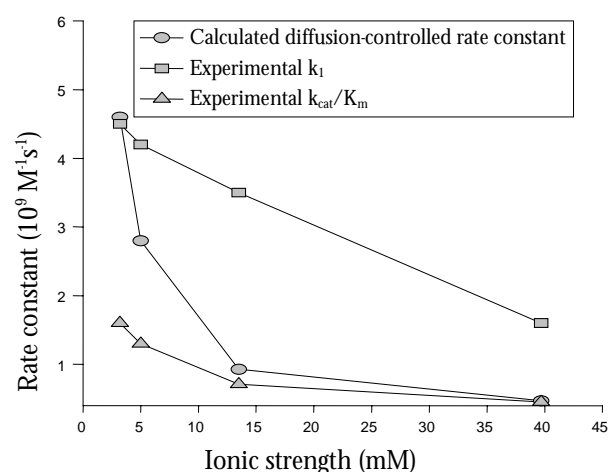
### Ionic strength

Computed rate constants of *Torpedo californica* AChE at various ionic strengths are given in Table 1. These values are compared with experimental bimolecular association constants ( $k_i$ ) and enzymatic specificity constants ( $k_{cat}/K_m$ ) of a related *Electrophorus electricus* AChE enzyme.<sup>9</sup> Since the association constant considers the binding event of the reaction and the specificity

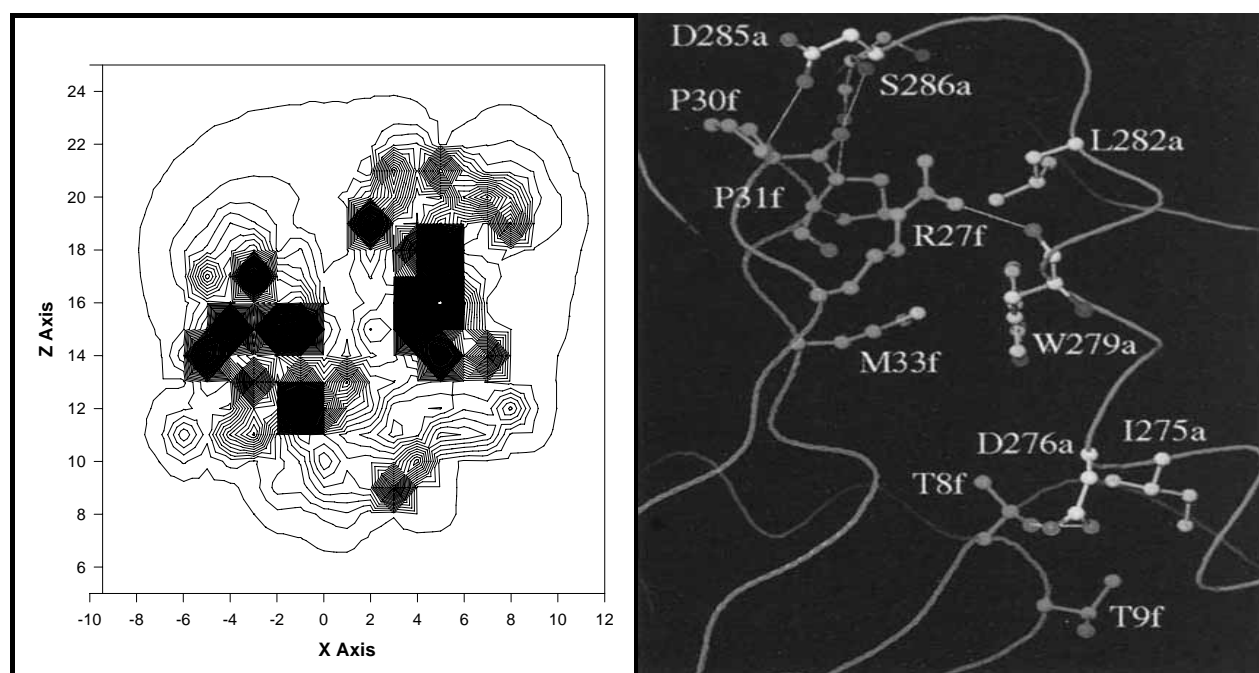
constant describes both binding and subsequent catalytic turnover,  $k_i$  is the theoretical maximum value for the calculated diffusion-controlled rate constant, while  $k_{cat}/K_m$  sets the lower limit on these second-order reactions. As seen in Figure 1, the calculated rate constants found in this work lie between these two extremes throughout the range of ionic strengths tested. This provides encouraging support for the ionic screening approximations used in this work. Furthermore, the decrease in the rate of AChE catalysis with increasing ionic strength provides strong evidence that an electrostatic steering mechanism plays a role in AChE kinetics. The similarity in the negative slope observed for both association and specificity constants indicates that the ligand binding step of the reaction is dependent upon solvent salt concentration.

**Table 1** Effect of Dielectric on Activity of wt AChE. 8000 trajectories were simulated at 300 K, an ionic strength of 5 mM, a solvent density of 996.5 kg/m<sup>3</sup>, and a pH of 7.0.

Dielectric	Rate constant (M <sup>-1</sup> s <sup>-1</sup> )	Standard error
78	$1.3 \times 10^9$	$0.267 \times 10^9$
60	$4.7 \times 10^9$	$0.615 \times 10^9$



**Figure 1** Experimental and calculated rate constants versus ionic strength. Calculations for preliminary ionic strength study used 1000 trajectories at a temperature of 298 K, a dielectric of 78, a solvent density of 997.0 kg/m<sup>3</sup>, and a pH of 7.0.



**Figure 2** Left: Electric field of AChE at an aqueous dielectric ( $\epsilon=78$ ). Active site gorge extends between  $z=14$  and  $z=20$ . Electric field contours =  $5 \times 10^{-12}$  N, x-axis represents the location of the catalytic Ser<sup>200</sup> residue and each grid unit is 4 Å. Right: Interface with interacting residues shown as ball-and-stick models of AChE covering AChE residues 68-90 and AChE residues 271-289 as active (a) and fused (f).

### Substrate radius

To demonstrate the limited accessibility of this enzyme's active site structure, simulations were performed using various substrate radii (Figure 1). It is reasonable to predict increased rate constant values with a reduced substrate radius since the probability of a smaller substrate penetrating the active site gorge and reacting with AChE is higher. The results shown in Figure 2 show this expected trend. These computed rate constants also compare well with the approximated specificity constant of  $4.2 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  for *Electrophorus electricus* AChE at zero ionic strength.<sup>9</sup>

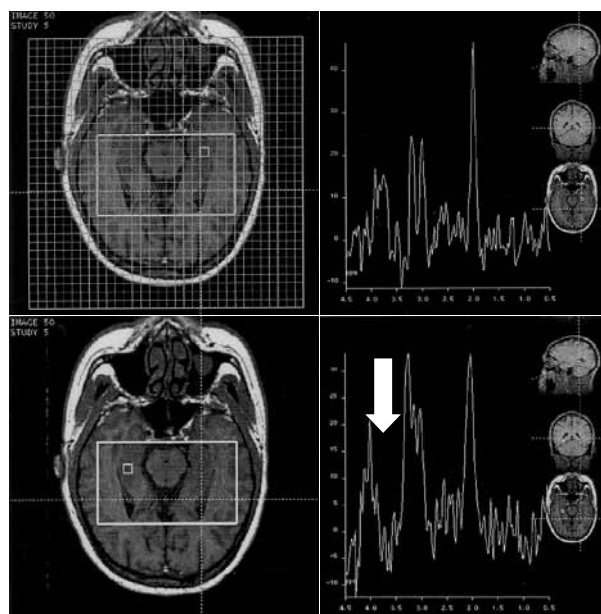
It has been hypothesized that the active site gorge of AChE undergoes conformational fluctuations to allow substrate entry.<sup>7</sup> Further simulation calculations in this work have used a smaller substrate radius (2 Å) than the actual hydrodynamic radius of ACh (4.23 Å) to approximately account for this protein flexibility.

### Reduced dielectric in plaques and NFTs

Simulations of the AChE-ACh reaction at both aqueous ( $\epsilon=78$ ) and hydrophobic ( $\epsilon=60$ ) conditions demonstrate the substantial effect of the AChE environment on its catalytic activity. The results of this study in Table 1 reveal that the abnormally low dielectric medium of the enzyme increases its catalytic rate by a factor of 3.6. These results are experimentally supported by Nolte et al.,<sup>9</sup> who found a rate constant of  $1.6 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  for *Electrophorus electricus* AChE under similar conditions.

These findings are complemented by the visualization of an intensified electric field near the entrance of the AChE active site gorge in the reduced dielectric environment. This observation can be made by comparing the electric field contours around the perimeter of AChE at dielectrics of 78 and 60 (Figure 2). These contour plots illustrate an electrostatic gradient emerging from the gorge entrance and extending along the

enzyme's surface. This effectively enlarges the active site target area, resulting in increased enzyme-substrate association. These electric field calculations along the protein's surface are consistent with the results of Antosiewicz et al.,<sup>7</sup> who have suggested that electrostatic steering is limited to operation near the surface of the enzyme. The contour plots found in this work also offer a plausible explanation for the catalytic rate constant enhancement at lower dielectrics and further suggest that electrostatic attraction is an important component of the AChE mechanism and ultimately its physiological role in the human nervous system.



**Figure 3** A comparison is shown between normal hippocampus MR spectroscopy Choline peak (normal brain peaks) and Alzheimer's Disease abnormal high Choline peak shown with an arrow (in panel at the bottom) due to reduced AChE.

The reduced AChE enzyme activity is also supported by high and localized Choline MR spectroscopy in AD brain as shown in Figure 3. In our previous data, the single voxel point resolved MR spectroscopy suggested strongly the enhanced choline peaks at scan parameters TE = 135 ms, TR= 5000 ms, pointing out the active role of AChE in active plaques and neurofibrillary tangles. However, other dielectric and physical

factors also play role in development of AD lesions and plaques.

## Conclusions

Abnormal Magnetic Resonance spectral peak patterns are associated with low AChE in AD, which is due to enhanced increasing ACh breakdown and surrounding electrostatic field of this enzyme.

## References

1. Jorgensen WL, Tirado-Rives J: The OPLS Potential Functions for Proteins. Energy. Minimizations for Crystals of Cyclic Peptides and Crambin. *J Am Chem Soc* 1988; 110(6):1657-1666.
2. Ermak D, McCammon J: Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phys* 1978; 69: 1352-1360.
3. Allison SA, Bacquet RJ, McCammon JA: Simulation of the Diffusion-Controlled Reaction Between Superoxide and Superoxide Dismutase II: Detailed Models. *Biopolymers* 1988; 27: 251-269.
4. Northrup SH, Allison SA, McCammon JA: Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. *J Chem Phys* 1984; 80(4): 1517-1524.
5. Northrup S, Hynes JT: Short range caging effects for the reactions in solutions, part I: reaction constants and short range caging picture; part II: escape probability and time dependent reactivity. *J Chem Phys* 1979; 71: 871-893.
6. Cygler M, Schrag JD, Sussman JL, et al.: Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Science* 1993; 2, 3: 366-382.
7. Antosiewicz J, McCammon J, Wlodek S, et al.: Simulation of Charge-Mutant Acetylcholinesterases. *Biochemistry* 1995; 34: 4211-4219.
8. Yu Y, Marut C: *Assembly Language Programming and Organisation of IBM PC*. 1992:McGrawhill Inc.
9. Nolte H, Rosenberry T, Neumann E: Effective Charge on Acetylcholinesterase Active Sites Determined from the Ionic Strength Dependence of Association Rate Constants with Cationic Ligands. *Biochemistry* 1980; 19: 3705-3711.



Strokovni članek ■

## Izzivi prenove spletišča Medicinske fakultete v Ljubljani

## Renewal challenges of Faculty of Medicine's website

---

Institucija avtorjev: Inštitut za biomedicinsko informatiko, Medicinska fakulteta, Univerza v Ljubljani.

Kontaktna oseba: Branimir Leskošek, Inštitut za biomedicinsko informatiko, Medicinska fakulteta, Vrazov trg 2, 1000 Ljubljana. enaslov: brane.leskosek@mf.uni-lj.si.

**Branimir Leskošek, Andrej Lesjak**

**Izveček.** Staro spletišče Medicinske fakultete (MF) v Ljubljani, ki je bilo osnovano leta 1993, že dalj časa ni več omogočalo učinkovitega obnavljanja vsebin in videza. Pred približno dvema letoma smo zato začeli s projektom prenove. Izzivi prenove niso bili samo tehnični ampak, kar na začetku ni bilo najbolj jasno, predvsem organizacijski. Glavni organizacijski izziv je bil kako poskrbeti, da bo skoraj 50 organizacijskih enot MF privzelo nov, elektronski, način objave vsebin, ki hkrati tudi bolj jasno razdeli odgovornosti. Novo spletišče obiskovalcem ponuja svež videz ter bolj ažurne informacije, uporabnikom (lokalnim urednikom) pa omogoča vnos vsebin preko sistema za upravljanje vsebin (SUV). V prihodnje želimo omogočiti osebno prilagoditev spletišča ter uvajati nove storitve v skladu z željami vseh uporabnikov. Odzivi so dobrodošli na splet@mf.uni-lj.si.

**Abstract.** The old website of Faculty of Medicine (FM) in Ljubljana was established in 1993 and it doesn't provide the efficient and easy updating according to today's standards. Approximately two years ago the project of website renewal was started. The challenges were not only technical but mainly organizational to the extent that was not foreseen at the beginning. The main challenge was how to make possible for almost 50 organizational units of FM to adopt the new, electronic, way of content publishing and to distribute the publishing responsibilities more clearly. Besides the fresh design the new website offers the visitors an easy access to up-to-date content and allows local editors situated at the organizational units to publish information using the content management system. In the future the development will be focused towards the personalization of website services. Any response is welcome at splet@mf.uni-lj.si.

■ **Infor Med Slov:** 2008; 13(1): 15-20

## Uvod

Staro spletišče Medicinske fakultete v Ljubljani (MF) je bilo postavljeno že leta 1993 kot eno izmed prvih spletišč v Sloveniji. Zaradi zastarele tehnologije in neučinkovitih postopkov obnavljanja vsebin spletišča je bilo zadnjih nekaj let vedno težje slediti zahtevam uporabnikov (obiskovalcem spletišča in avtorjem informacij).

Aktivnosti za prenovo spletišča so sprožile na prvi pogled majhne zahteve v zvezi s poenoteno spletno objavo obvestil za vse organizacijske enote MF. Inštitut za biomedicinsko informatiko Medicinske fakultete v Ljubljani (<http://www.mf.uni-lj.si/ibmi>), ki je bil zadolžen za tehnično vzdrževanje strežnika obstoječega spletišča, je dobil nalogo, da preuči in predlaga možne, po možnosti odprtokodne, rešitve. Med iskanjem odprtokodnih sistemov za upravljanje vsebin (SUV) smo preučili več sistemov kot so npr. Mambo, Joomla, phpNuke, Postnuke, e107, ki so bili pred dvema letoma med boljšimi.<sup>1</sup> Po krajšem preučevanju SUV in vedno novih zahtev s strani uporabnikov smo ugotovili, da je naloga prenove spletišča veliko bolj kompleksna in da z uporabo obstoječih odprtokodnih sistemov naloge ne bo mogoče izvesti.

Učinkovito upravljanje spletišča je možno samo z dobro definiranimi in modernimi postopki pridobivanja, obdelave in objave informacij, ki so ustrezno umeščeni v organizacijsko shemo organizacije. Zato je bila na MF že skoraj pred dvema letoma imenovana projektna skupina za prenovo spletišča MF, ki jo vodi dr. B. Leskošek. Projektna skupina naj bi staro spletišče MF tehnološko prenovila in postavila ustrezne organizacijske postopke za njegovo (predvsem) vsebinsko, tehnično in organizacijsko obnavljanje.

## Namen in cilj

Izdelati tehnično dovršeno in funkcionalno spletišče, ki bo služilo kot ustrezno ogrodje za

dobro pretočnost notranjih in zunanjih informacij v organizaciji.

## Metode

S pomočjo fokusnih skupin smo dobili okvirne zahteve za funkcionalnost spletišča. Bolj izpostavljena zahteva je bila omogočiti lokalnim urednikom organizacijskih enot MF (OE), da sami, v elektronski obliki, sodelujejo pri ustvarjanju vsebin in obveščanju obiskovalcev spletišča ter da lahko enostavno medsebojno komunicirajo.

Projektna skupina je določila zahteve za organizacijsko (glavna organizacijska sprememba je bila, da naj bi se z novim spletiščem informacije vnašale čim bližje mestu nastanka) in tehnično (SUV, tehnične podrobnosti) prenovo. Poleg tega so člani projektna skupine pripravili dokumentacijo za formalno sprejetje novih organizacijskih postopkov in obveščali vodstvo MF o napredovanju projekta.

Strokovne sodelavce smo izbrali preko javnega naročila. Vzpostavili smo redne sestanke tehnične ekipe. V kratkih časovnih intervalih smo preverjali delovanje SUV in skladnost s cilji. Vseskozi smo iskali ravnovesje med zahtevami in možnimi tehničnimi rešitvami. Pri operativnem delovanju tehnične skupine smo se zgledovali po pravilih Demingovega kroga<sup>2</sup> – Načrtuj, Izvedi, Preveri, Ukrepaj in načelih ekstremnega programiranja.<sup>3</sup>

Projekt prenove spletišča MF je bil vseskozi ustrezno umeščen in usklajen z ostalimi prizadevanji na področju informatizacije na Medicinski fakulteti. Spletišče je zasnovano modularno in omogoča vključevanje zunanjih modulov kot so npr. imenik, urnik, koledar, sinhronizacija z zbirko uporabnikov.

Zaradi tehnične in predvsem organizacijske kompleksnosti projekta se je pomembno podaljšal tudi predvideni čas prenove spletišča. Nedokončane naloge smo na vsakem (tehničnem) sestanku razvrstili po prioriteti.

## Rezultati

Novo spletišče MF je komunikacijski center, ki omogoča objavlanje informacij na mestu nastanka brez posrednikov. Tako lahko vsaka organizacijska enota MF (OE) s pomočjo sistema za upravljanje vsebin (SUV) sama pripravlja oz. sodeluje pri pripravi lastnih vsebin v elektronski obliki, brez posrednikov, hitreje in z manj napakami.

Vzpostavili smo uredniški sistem sestavljen iz treh členov: glavni urednik in tehnični urednik (1. člen) sta zadolžena za vsebinsko in tehnično kontrolo dinamičnih vsebin spletišča, lokalni uredniki (2. člen) vnašajo vsebine, ki jih sami zberejo ali jim jih posredujejo drugi avtorji (3. člen). Tak način uporabnikom sistema nalaga tudi večjo odgovornost pri objavi informacij. Vsak izmed uporabnikov sistema nosi vnaprej določene oz. dogovorjene odgovornosti zapisane v internem pravilniku.

Pomembna značilnost novega spletišča je, da se večina spremenljivih informacij (statične in dinamične vsebine) shranjuje v podatkovni zbirki, kar na eni strani omogoča prikaz iste informacije na več različnih mestih spletišča, na drugi strani pa popraviljanje iste informacije le na enem mestu.

Zaradi drugačne vizualne podobe (Slika 1) in organizacije ponuja prenovljeno spletišče MF obiskovalcem čisto novo uporabniško izkušnjo. Poleg moderne in privlačne podobe je z uporabo bližnjic in naprednih možnosti omogočeno lažje in hitrejše iskanje zelenih informacij (npr. študenti samo spremembe predavanj, informacije o študiju, gradivo za študij, učitelji in zaposleni gradiva delovnih teles MF, naključni obiskovalci pa informacije o MF). Vsaka OE ima poenoteno obliko lastne (vstopne) spletne strani.



**Slika 1** Začetna stran novega spletišča MF Ljubljana.

Oblikovna predloga sledi grafični podobi MF Ljubljana in upošteva oblikovna določila za članice Univerze v Ljubljani. Vizualna podoba je podrejena funkcionalnosti vendar vnaša veliko mero svežine z uporabo naključnega izmenjevanja slik v ozadju in uporabo primernih barv ter pisav.

SUV urednikom omogoča pregledno in enostavno urejanje vsebine ter strukture spletišča. Vsebine lahko obogatijo s slikovnim gradivom in dokumenti v standardnih formatih. Možen je vpogled v zgodovino nastajanja vsebine. Uredniki imajo pravice za urejanje delov spletišča za katere so odgovorni. Zagotovljena je komunikacija med uredniki, kar omogoča enostavno medsebojno posvetovanje in sodelovanje pri nastajanju vsebine. Modularna zasnova omogoča hitro nadgrajevanje in vzdrževanje sistema, kar mu daje dolgo življenjsko dobo. Za varnost podatkov in sistema samega je poskrbljeno že na ravni strežnika, prav tako pa sam sistem preverja upravičenost urednikovega dostopa na več nivojih.

## Tehnični del

SUV (Slika 2) lahko uporablja urednik z osnovnim znanjem uporabe računalnika. SUV v celoti deluje preko spletnega vmesnika in poleg urejanja vsebin omogoča še sledljivost, spremljanje statistik ter komunikacijo med uporabniki.

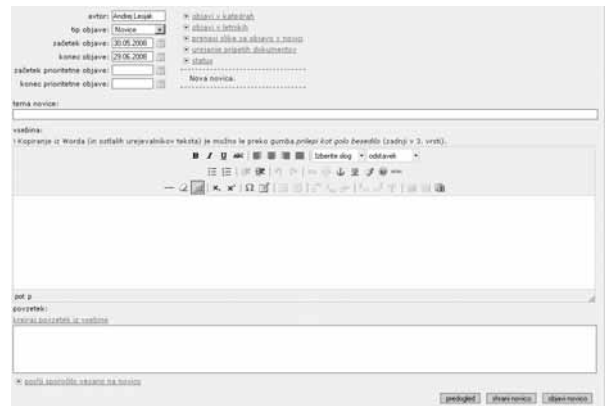


**Slika 2** Uporabniški vmesnik SUV.

SUV omogoča ločevanje podatkov, programske kode in oblikovnih predlog. Prednost ločevanja podatkov je v tem, da je posamezen podatek shranjen le enkrat, vendar je lahko uporabljen večkrat (npr. novica namenjena določeni OE se lahko prikazuje skupaj z vsemi novicami ali pa le pri novicah "matične" OE). Za SUV so uporabljene preizkušene, napredne in odprtokodne tehnologije (npr. AJAX<sup>4</sup>). Sistem deluje na odprtokodnem operacijskem sistemu Linux (CentOS).

## Novice

Modul Novice (Slika 3) je namenjen objavljanju aktualnih dogodkov, obvestil, oglasov, ipd. Objavam lahko določimo čas trajanja objave ter prioriteto. Označimo jih lahko le za določen letnik ali OE, kar omogoča uporabnikom spletišča lažje iskanje relevantnih novic. Urejanje objav poteka s pomočjo WYSIWYG (*What You See Is What You Get*) urejevalnika, kar urednikom omogoča proste roke pri oblikovanju vsebine. Objavam lahko dodamo slikovno gradivo, ter pripne dokumente.



**Slika 3** Obrazec za vnašanje novic z vključenimi naprednimi možnostmi.

## Vsebina

Modul Vsebina je namenjen urejanju strukture ter vsebine spletišča. Uredniki lahko preprosto dodajajo nove strani ter urejajo ali brišejo obstoječe. Vsebino prav tako urejamo z WYSIWYG urejevalnikom. Tudi vsebino lahko obogatimo s slikovnim gradivom in dokumenti.

## Notranji sporočilni sistem (NSS)

Modul NSS (Slika 4) je namenjen komunikaciji med uredniki. V pogovoru lahko sodeluje več urednikov, ki jih izbere avtor začetnega sporočila. Sporočila so lahko vezana na določeno objavo ali vsebino in so prikazana v večnivojski drevesni strukturi.



**Slika 4** Vmesnik za komuniciranje med uporabniki SUV.

### Ostali moduli

Sistem za upravljanje vsebin vsebuje tudi module za objavljane povezav, za urejanje osnovnih podatkov organizacijskih enot, za administracijo uporabnikov, uredniških podatkov in pravic, statistiko ter iskalnik.

### Organizacijski del

Projektna skupina je pripravila organizacijsko in tehnično dokumentacijo v povezavi s spletiščem MF. Organizacijski pravilnik o urejanju spletnih strani, ki ga je potrdil senat MF, določa pravice in dolžnosti posameznih OE v zvezi z vsebinami objavljenimi na spletišču.

Izobraževanje za uporabo SUV za lokalne urednike vseh organizacijskih enot MF smo izvedli konec maja 2008. Že pred izobraževanji smo vsem lokalnim urednikom poslali priročnik za uporabo SUV<sup>5</sup> in jim omogočili testiranje funkcij SUV na testnem strežniku. V pripravo kakovostnega priročnika za uporabo SUV smo vložili veliko truda. Pričakujemo, da bo priročnik v veliko pomoč uporabnikom SUV in da bo pomembno zmanjšal količino potrebne tehnične pomoči.

Izobraževanja so potekala v obliki praktičnega dela s kratkim uvodnim predavanjem. Na koncu izobraževanja je bilo dovolj časa za razpravo in predstavitev naprednejših funkcij SUV. Izobraževanje se je udeležilo preko 80% vseh povabljenih lokalnih urednikov, kar je po našem mnenju zelo dobra udeležba. Uspešno opravljeno izobraževanje je pogoj za pridobitev akreditacije za uporabo SUV.

Zaradi lažjega prehoda na novo spletišče MF in lažjega prenosa vsebin (tudi iz starega spletišča) smo se odločili, da bomo preko poletja 2008 pustili operativno tako staro kot novo spletišče. Novo spletišče MF bo ta čas dostopno na spletnem naslovu <http://www1.mf.uni-lj.si>. Jeseni 2008 ob začetku novega šolskega leta pa bo staro spletišče

ukinjeno, novo pa predstavljeno na ustaljeni naslov <http://www.mf.uni-lj.si>.

### Razprava

Zelo pomemben izziv po zagonu spletišča ostaja stalno upravljanje s kakovostjo spletišča. V okviru upravljanja želimo npr. stalno spremljati delo posameznih lokalnih urednikov (število objav, objava zahtevanih gradiv in informacij: izpitni roki, termini kolokvijev, učna gradiva) ter analizirati in upoštevati povratne informacije obiskovalcev in uporabnikov spletišča z izvajanjem (rednih) delovnih srečanj, anket in mnenj. Načrtno bomo iskali tudi dodatne možnosti optimizacije organizacijskih postopkov.

V prihodnje želimo spletišče nadgraditi tako, da bo

1. obiskovalcem v čim večji meri omogočalo osebno prilagoditev spletnega vmesnika (personalizacija),
2. uporabnikom oz. zaposlenim na MF na voljo poenoten vmesnik do osebnih spletnih strani,
3. možen poenostavljen prenos vsebin iz obstoječih spletišč posameznih OE,
4. skladno z razvojem mobilnih tehnologij omogočeno prikazovanje vsebin na dlančnikih in pametnih telefonih.

Projektna skupina je rešila mnogo organizacijskih in tehničnih zagat ter si pridobila veliko izkušenj. Že med razvojem spletišča je več OE potrebovalo nasvete v zvezi z njihovimi lokalnimi željami pri prenosi podsegmentov spletišča MF (npr. OE Centralna medicinska knjižnica pomagamo pri oblikovanju zahtev), kar kaže na potrebo po dolgoročni sistemski ureditvi vzdrževanja spletišča ne samo s tehničnega ampak tudi z organizacijsko-vsebinskega vidika.

Nadaljnji razvoj spletišča MF bo seveda skladen z željami obiskovalcev/ uporabnikov in s cilji organizacije. Tudi zato ste vsi bralci vabljeni k

obisku spletišča. Vseh konstruktivnih pripomb in opazanj bomo veseli na elektronskem naslovu [splet@mf.uni-lj.si](mailto:splet@mf.uni-lj.si) (ki je hkrati tudi elektronski naslov za tehnično pomoč).

### Literatura

1. OpenSourceCMS - Web Page.  
<http://www.opensourcecms.com>, 2008.
2. Demingov krog - Wikipedia Web Page.  
[http://en.wikipedia.org/wiki/Quality\\_circle](http://en.wikipedia.org/wiki/Quality_circle), 28.05.2008
3. chromatic: *Extreme Programming Pocket Guide*. Sebastopol 2003: O'Reilly Media.
4. Mahemoff M: *Ajax Design Patterns*. Sebastopol 2006: O'Reilly Media.
5. Leskošek B, Lesjak A: Navodila za urejanje spletišča UL MF za lokalne urednike, Uporabniški priročnik. Ljubljana 2008, Medicinska fakulteta Univerze v Ljubljani.

Strokovni članek ■

## Analiza podatkov o ocenjevanju z Lestvico funkcijske neodvisnosti na Inštitutu Republike Slovenije za rehabilitacijo

## Analysis of data on assessment with the Functional Independent Measure at the Institute for Rehabilitation, Republic of Slovenia

---

Institucija avtorjev: Inštitut Republike Slovenije za rehabilitacijo.

Kontaktna oseba: Gaj Vidmar, Inštitut Republike Slovenije za rehabilitacijo, Linhartova 51, 1000 Ljubljana. email: gaj.vidmar@ir-rs.si.

**Gaj Vidmar, Helena Burger, Črt Marinček, Robert Cugelj**

**Izveček.** Lestvica funkcijske neodvisnosti (FIM) je najširše uporabljano orodje za ocenjevanje izida v rehabilitacijski medicini. Sestavljata jo motorična in kognitivna podlestvica s 13 oz. 5 postavkami, ocene so 7-stopenjske. Na našem Inštitutu obvezno ocenjevanje bolnikov poteka od leta 2004. Prispevek obravnava podatke do konca leta 2006. Poleg dosežkov na FIM analiziramo demografske značilnosti bolnikov, trajanje rehabilitacije, čas od okvare, razvrstitev v skupine funkcijskih okvar in razrede primerljivih funkcij, sprejemne in odpustne diagnoze, komplikacije in prekinitve rehabilitacije ter tip sprejema, obravnave in odpusta bolnikov. Preverjamo tudi merske značilnosti FIM ter skušamo izdelati modele za napoved napredka bolnikov. Prispevek poudarja pomen kakovostne vizualizacije podatkov za podporo odločanju in pomen ustrezne analize podatkov za kakovost v zdravstvu.

**Abstract.** Functional Independence Measure (FIM) is the most widely used outcome assessment tool in rehabilitation medicine. It consists of motor and cognitive subscale with 13 and 5 items, respectively, all rated on a 7-point scale. At our Institute, compulsory assessment has been preformed since 2004. The paper addresses the data until the end of 2006. In addition to FIM scores, we analyze patient demographics, duration of rehabilitation, time since impairment, FIC and AN-SNAP classification, admission and discharge diagnosis, complications and treatment cessations, and type of admission, treatment and discharge. We also assess metric characteristics of FIM and try to model patient progress. The paper emphasizes the importance of adequate data visualization for decision support and the importance of professional data analysis for healthcare quality.

■ **Infor Med Slov:** 2008; 13(1): 21-32

## Uvod

Lestvica funkcijske neodvisnosti (Functional Independence Measure, FIM) je najpogosteje in najširše uporabljano orodje za ocenjevanje izida v rehabilitacijski medicini.<sup>1</sup> Temelji na pojmovnem ogrodju Mednarodne klasifikacije okvar, prizadetosti in oviranosti (ICIDH<sup>2</sup>). Sestavljata jo motorična podlestvica s 13 postavkami in kognitivna podlestvica s 5 postavkami, vse postavke pa se ocenjujejo z ocenami od 1 do 7. V Sloveniji se lestvica FIM od leta 2002 uporablja kot referenčni kazalnik kakovosti dela v projektu Zdravniške zbornice Slovenije,<sup>3</sup> na Inštitutu Republike Slovenije za rehabilitacijo (IRSR) pa se je po uvedbi v raziskovalno delo<sup>4</sup> obvezno ocenjevanje bolnikov in sistematično zbiranje podatkov pričelo z letom 2004.<sup>5</sup> Predstavljena analiza zajema podatke do konca leta 2006.

## Metode

Poleg dosežkov na FIM (skupaj in na obeh podlestvicah) smo analizirali demografske značilnosti bolnikov, trajanje rehabilitacije, čas od okvare, razvrstitev v skupine funkcijskih okvar (FIC<sup>6</sup>) in avstralske razrede primerljivih funkcij (ARPF<sup>7</sup>), sprejemne in odpustne diagnoze, komplikacije in prekinitve rehabilitacije, tip obravnave ter obliko sprejema in odpusta bolnikov. Za vse spremenljivke smo izračunali ustrezne opisne statistike in izdelali ustrezne grafične prikaze. Z univariantimi (intervali zaupanja za ocenjene parametre), bivariatnimi

(testi povezanosti in primerjave med skupinami) in multivariatnimi analizami (regresijski modeli, združevanje v skupine) smo ugotavljali značilnosti bolnikov in njihove rehabilitacije s poudarkom na uporabnosti FIM in primerjavi med leti. Analizirali smo tudi merske značilnosti FIM. Vse analize smo izdelali v povezavi z grafičnimi prikazi, ki uveljavljajo načela kakovostne vizualizacije podatkov in s tem olajšujejo odkrivanje zakonitosti. Uporabili smo elektronsko preglednico Microsoft<sup>®</sup> Excel in statistični programski paket SPSS<sup>®</sup> za okolje Windows.



## Izbrani rezultati

Celovita predstavitev vseh rezultatov in njihova poglobljena interpretacija seveda presega dopustni obseg in osnovni namen tega članka. Zato v nadaljevanju predstavljamo izbor rezultatov, ki ilustrira vrste, značilnosti in uporabnost opravljenih analiz.

### Značilnosti bolnikov, primerjane med leti

Tabela 1 prikazuje strukturo spola bolnikov po letih. Ker razlike med leti niso statistično značilne (test  $\chi^2$ :  $p=0,310$ ), je smislen prikaz skupnega deleža s stolpčnim diagramom, ki takoj pokaže, da je moških nekaj manj kot dve tretjini. Že tovrstne preproste kombinacije tabel in diagramov, izdelane z upoštevanjem ustreznih statističnih analiz, bi bistveno izboljšale kakovost rutinskega strokovnega in poslovnega poročanja v zdravstvu.

**Tabela 1** Struktura bolnikov, sprejetih na IRSR v letih 2004-2006, glede na spol.

Spol	Leto odpusta			Skupaj	Grafični prikaz
	2004	2005	2006		
M	771 (59,4%)	743 (56,9%)	831 (59,6%)	2345 <b>(58,7%)</b>	
Ž	528 (40,6%)	562 (43,1%)	563 (40,4%)	1653 <b>(41,3%)</b>	



**Tabela 2** Opisne statistike za starost, čas od okvare in ležalno dobo za leta 2004-2006.

Leto odpusta	Statistika	Starost (leta)	Čas od okvare (meseci)	Bolnišnični dnevi
2004	M (SD)	59,1 (17,6)	37,2 (85,9)	34,0 (26,3)
	Me (razpon)	63,0 (7-94)	6,3 (0-924)	27,0 (0-261)
2005	M (SD)	60,7 (16,8)	31,8 (78,6)	34,5 (27,5)
	Me (razpon)	63,0 (15-95)	5,6 (0-922)	27,0 (0-310)
2006	M (SD)	58,9 (17,6)	35,0 (85,3)	34,7 (31,4)
	Me (razpon)	62,0 (15-97)	6,0 (0-840)	25,0 (0-366)
Skupaj	M (SD)	59,5 (17,4)	34,7 (83,4)	34,4 (28,6)
	Me (razpon)	63,0 (7-97)	6,0 (0-924)	26,0 (0-366)

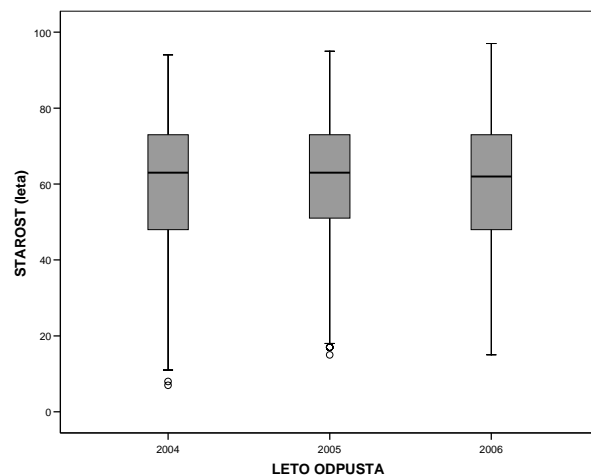
Iz tabele 2 je razvidno, da se tudi starost, čas od okvare do sprejema in čas rehabilitacije na IRSR v obravnavanem obdobju niso opazno spreminjali. Prikaz porazdelitve starosti bolnikov s škatlastimi diagrami kvantilov (slika 1) potrjuje levo asimetričnost porazdelitve, na katero kaže podatek, da je aritmetična sredina manjša od mediane.

Analiza strukture FIC (tabela 3) je pokazala statistično značilne razlike med leti (eksaktna vrednost  $p$  za  $\chi^2$  test povezanosti, ocenjena s ponovnim vzorčenjem, je pod 0,001). Tovrstne analize (z ustreznim prepoznavanjem odstopajočih celic) so uporabne predvsem kot opozorilo oziroma povod za bolj poglobljene analize, ki razjasnijo vzroke ter morebitni klinični oziroma praktični pomen opaženih razlik.

**Ocene FIM ob sprejemu in odpustu ter njihova povezanost z značilnostmi bolnikov**

Morebitne spremembe srednje vrednosti, razpršenosti oziroma druge značilnosti porazdelitve številskih kazalnikov je širšemu krogu uporabnikov analiz najlažje predočiti z ustrežno urejenimi histogrami. Slika 2 tako nazorno kaže pomik porazdelitve ocen po rehabilitaciji v desno v okviru možnega razpona FIM. Sprejemne in odpustne vrednosti oziroma spremembe FIM smo nato

zaradi njihovega osrednjega pomena kot najširše uporabne mere uspešnosti rehabilitacije podrobneje analizirali z različnimi analitičnimi grafikami (slike 3-7) in tabelami (opisne statistike in statistilni testi za spremembo celotne lestvice in obeh podlestvic glede na tip sprejema, ARPF idr., kar ni obravnavano v tem prispevku). Zaradi nazornosti prikazi potrebujejo le zelo kratke komentarje, ki se jih zato da vključiti kar v opise slik.

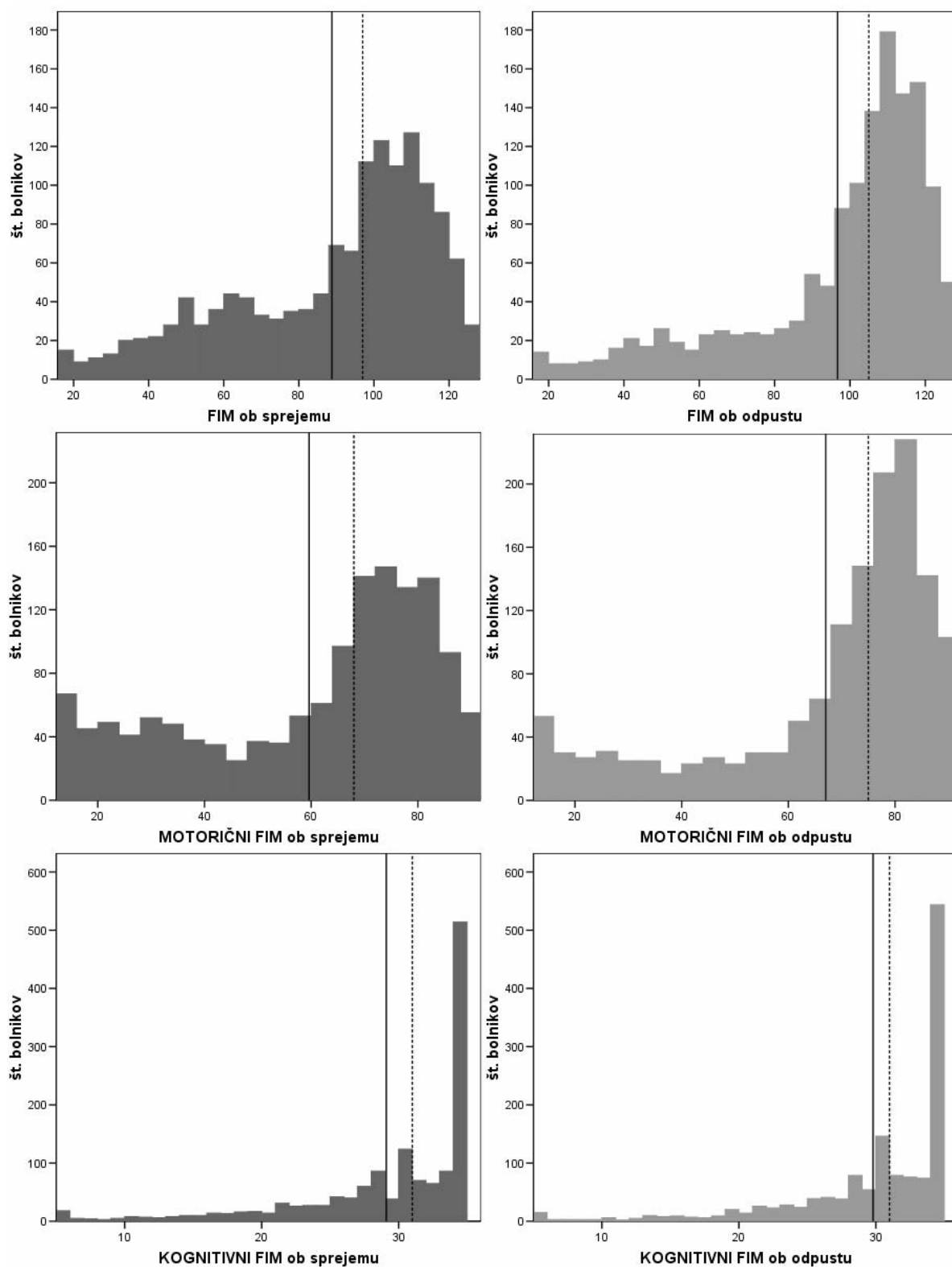


**Slika 1** Porazdelitev starosti bolnikov na IRSR v letih 2004-2006.

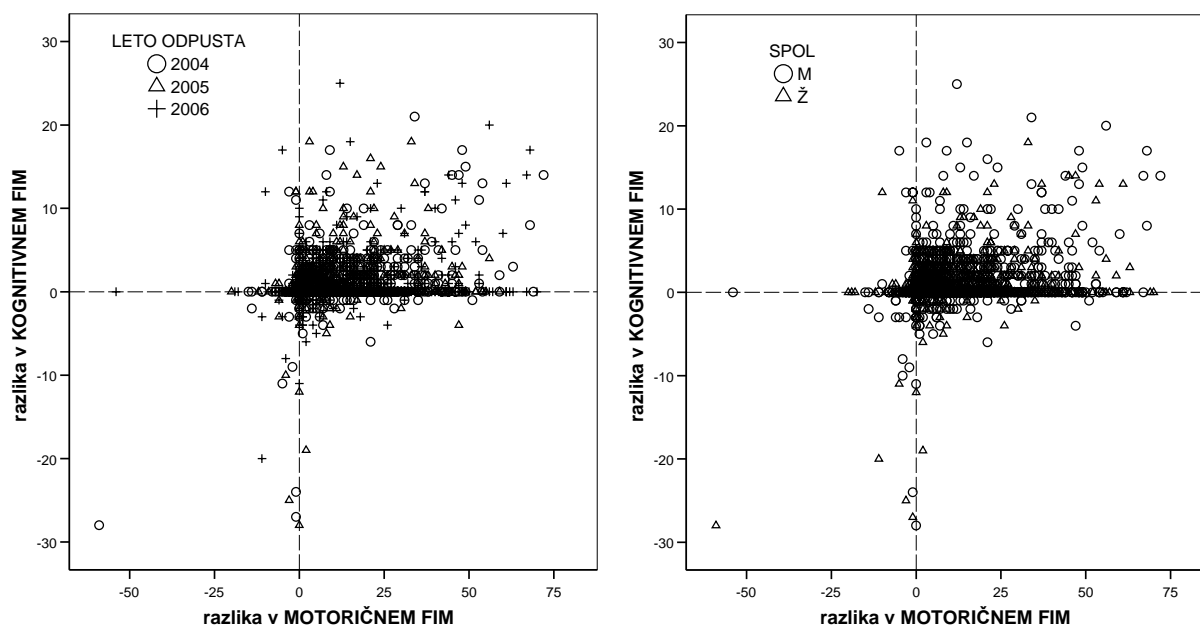
**Tabela 3** Struktura sprejetih bolnikov glede na skupine FIC v letih 2004-2006.

Skupina FIC	Leto odpusta			Skupaj (delež)	Grafični prikaz
	2004	2005	2006		
Amputacije udov	386 (29,7%)	384 (29,4%)	406 (29,1%)	1176 <b>(29,4%)</b>	
Možganska kap	258 (19,9%)	271 (20,8%)	276 (19,8%)	805 <b>(20,1%)</b>	
Nevrološka stanj	209 (16,1%)	250 (19,2%)	235 (16,9%)	694 <b>(17,4%)</b>	
Ortopedska stanja	142 (10,9%)	153 (11,7%)	158 (11,3%)	453 <b>(11,3%)</b>	
Okvara hrbtenjače	134 (10,3%)	143 (11,0%)	129 (9,3%)	406 <b>(10,2%)</b>	
Okvara možganov	61 (4,7%)	50 (3,8%)	122 (8,8%)	233 <b>(5,8%)</b>	
Bolečinski sindromi	28 (2,2%)	19 (1,5%)	30 (2,2%)	77 <b>(1,9%)</b>	
Artritis	12 (0,9%)	14 (1,1%)	14 (1,0%)	40 <b>(1,0%)</b>	
Hude multiple poškodbe	19 (1,5%)	7 (0,5%)	6 (0,4%)	32 <b>(0,8%)</b>	
Opekline	1 (0,1%)	5 (0,4%)	2 (0,1%)	8 <b>(0,2%)</b>	
Okvare srca in ožilja	2 (0,2%)	0 (0,0%)	1 (0,1%)	3 <b>(0,1%)</b>	
Prirojene nepravilnost	1 (0,1%)	0 (0,0%)	1 (0,1%)	2 <b>(0,1%)</b>	
Oslabelost	1 (0,1%)	1 (0,1%)	0 (0,0%)	2 <b>(0,1%)</b>	
Druge okvare, ki vpl. na zmožnosti	45 (3,5%)	8 (0,6%)	14 (1,0%)	67 <b>(1,7%)</b>	

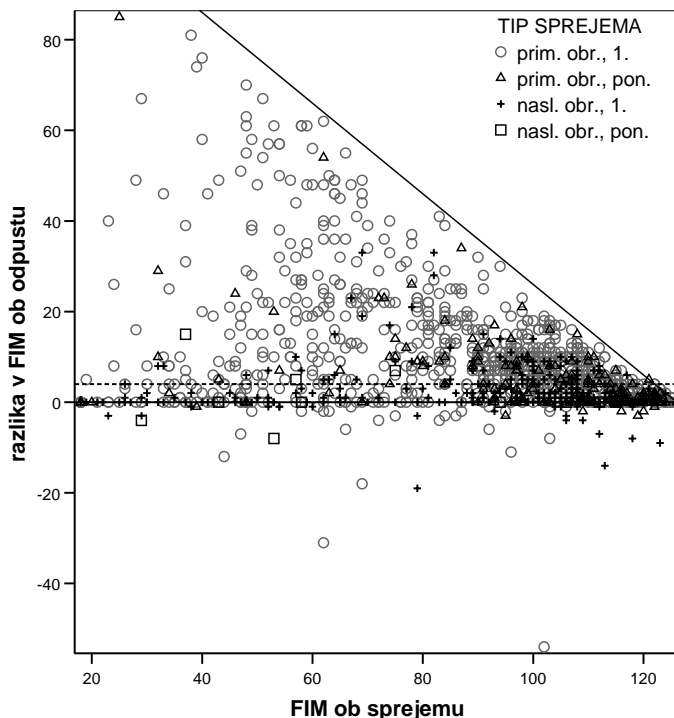
Osenčene so odstopajoče celice s standardiziranim ostankom pri  $\chi^2$  testu povezanosti pod -2 (svetlo) oziroma nad 2 (temno). Med vodoravnimi stolpci, ki prikazujejo robno porazdelitev FIC, so črno obarvani tisti z odstopajočimi celicami, torej izrazitimi razlikami deleža kategorije med leti.



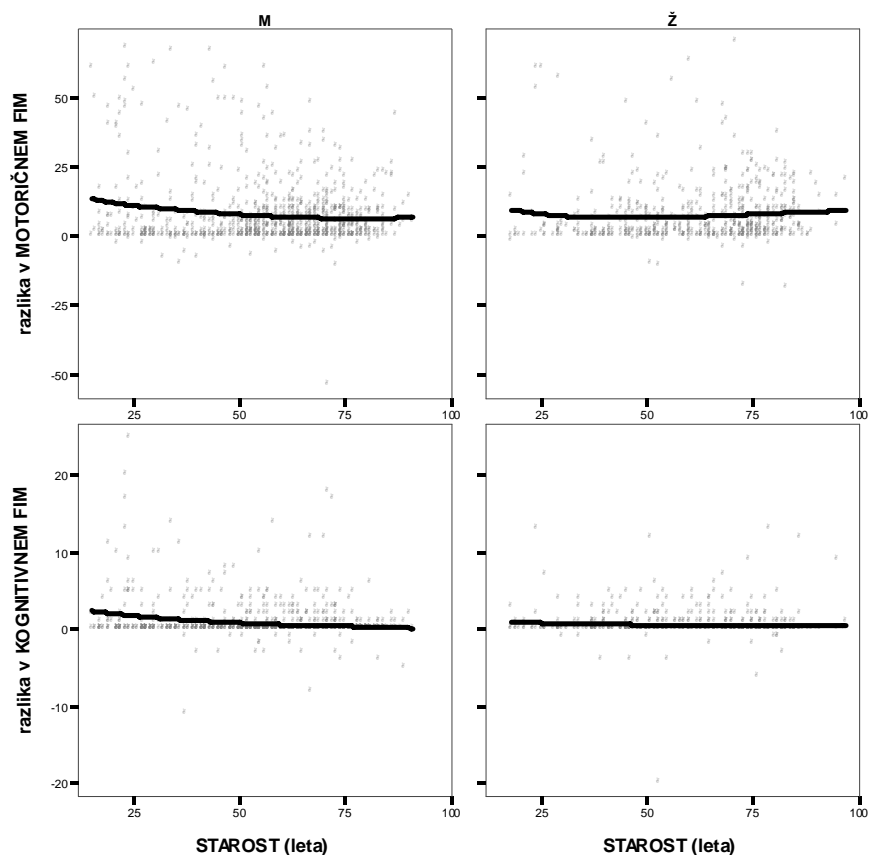
**Slika 2** Ocene na skupni lestvici in obeh podlestvicih FIM ob sprejemu (levo, temneje) in odpustu (desno, svetleje) leta 2006. Polna črta označuje aritmetično sredino, črtkana mediano; vsaka dvojica histogramov ima enako razdeljeni osi. Javno opazimo pomik porazdelitve v okviru razpona lestvice ter povečanje obeh mer srednje vrednosti ob odpustu.



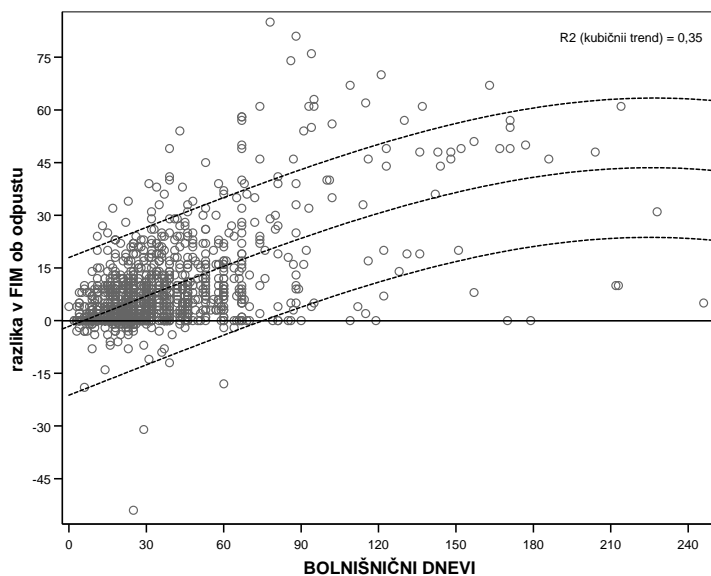
**Slika 3** Povezanost med razliko v motoričnem in kognitivnem FIM po rehabilitaciji glede na leto (levo) in spol (za vsa opazovana leta skupaj, desno). Sistematičnih razlik med leti oz. spoloma ni; opazna je korelacija razlik med podlestvicama.



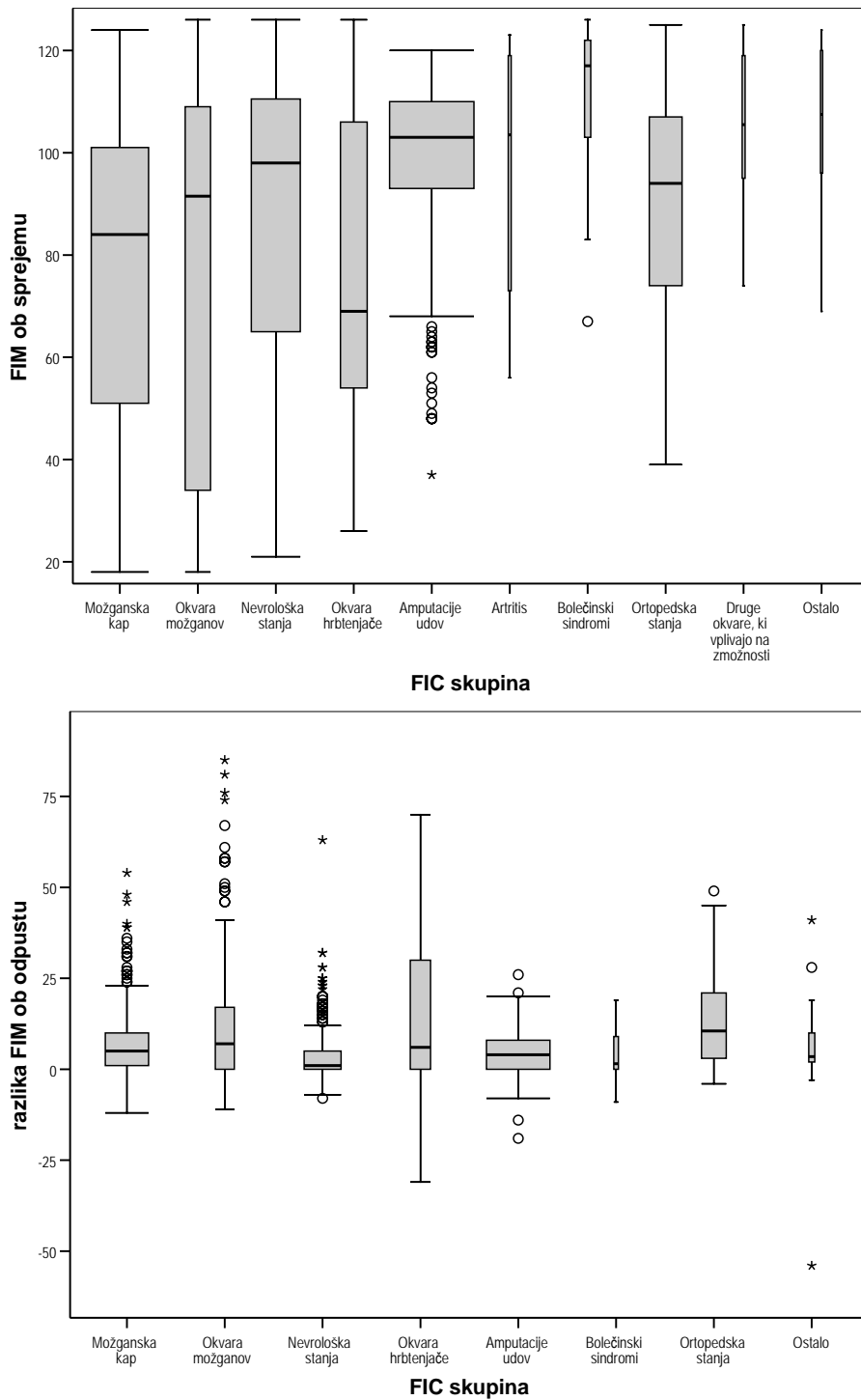
**Slika 4** Sprememba FIM po rehabilitaciji v odvisnosti od sprejemnega FIM glede na tip obravnave (primarna ali naslednja) in sprejema (prvi ali ponovni) v letu 2006. Poševna črta označuje zgornjo mejo možnega povečanja; črtna vodoravna črta označuje povprečno spremembo.



**Slika 5** Odvisnost spremembe motoričnega in kognitivnega FIM od starosti glede na spol za leto 2006. Razsevnim diagramom je dodana robustna neparametrična regresija z glajenjem (LOESS).



**Slika 6** Odvisnost spremembe FIM od trajanja hospitalizacije za leto 2006. Vrisan je polinomski trend tretje stopnje (z navedenim popravljeni determinacijskim koeficientom) z ocenjenim 95% intervalom zaupanja za napoved; izpuščena sta bolnika z ležalno dobo nad 8 mesecev. Vidi se, da rehabilitacija, daljša od pol leta, praviloma ni smiselna.



**Slika 7** Porazdelitev FIM ob sprejemu ter razlik v FIM ob odpustu po skupinah FIC v letu 2006. Širina škatlastega diagrama kvantilov je sorazmerna s številom primerov v skupini FIC. Bolniki z artritisom in kategorije Druge okvare, ki vplivajo na zmožnosti, so zaradi heterogenosti in majhnega števila na spodnjem diagramu vključeni v kategorijo Ostalo.

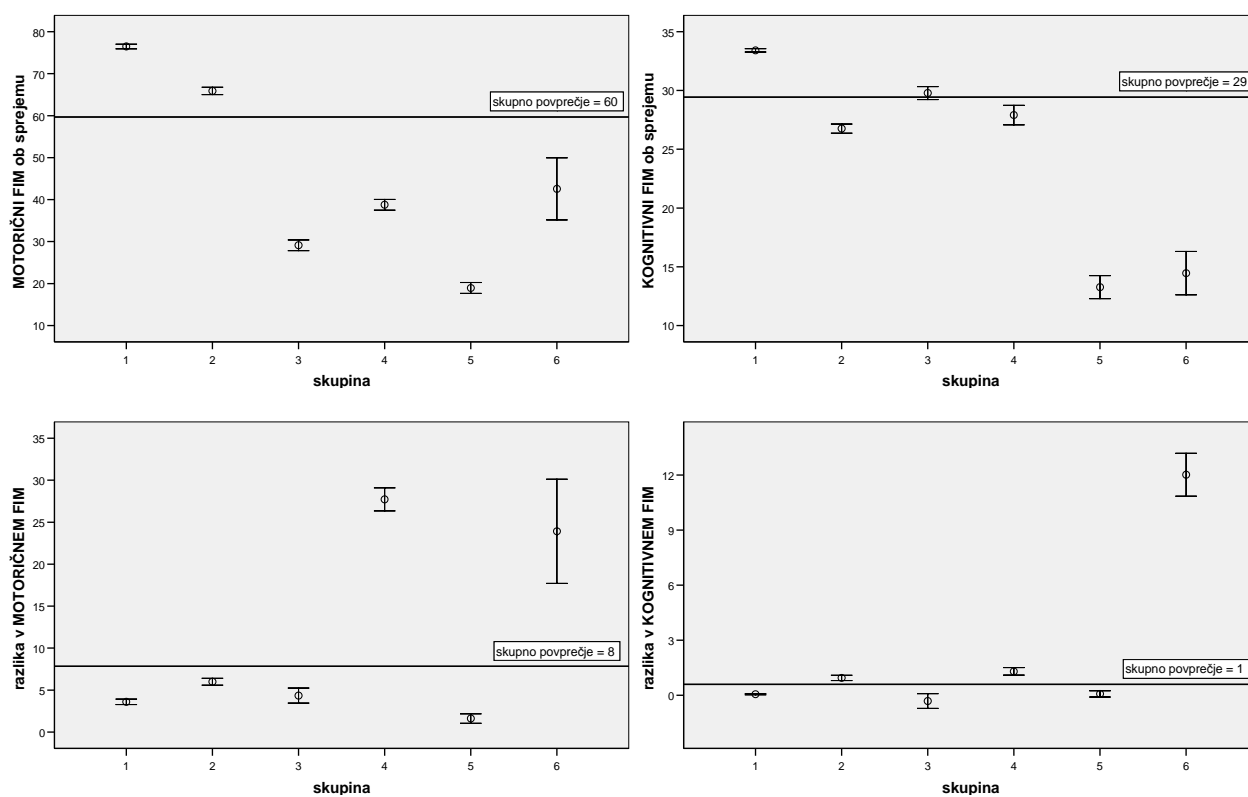
**Združevanje bolnikov v skupine glede na sprejemno vrednost in spremembo FIM**

Pri statističnem združevanju bolnikov v skupine je bil uporabljen dvostopenjski postopek, ki daje zanesljive rezultate. Najprej smo s hierarhičnim združevanjem v skupine določili najprimernejše število skupin (ob upoštevanju več kriterijev), nato pa smo izvedli razdeljevanje v vnaprej določeno število skupin (postopek *k*-means). Pri obeh korakih je bila kot mera različnosti med primeri uporabljena evklidska razdalja. Spremenljivke smo zaradi različnih lestvic pred postopkom združevanja v skupine standardizirali. Za hierarhično združevanje v skupine smo uporabili Wardov postopek. Kot najprimernejša se je pokazala rešitev s šestimi razredi, ki ustreza tako Bayesovemu informacijskemu kriteriju kot kriteriju deleža pojasnjene variance. Za podrobnosti postopkov združevanja in izbora najprimernejšega števila skupin tu ni prostora, vsekakor pa sliki 8 in 9 v povezavi z oceno deleža sprejemov v posamezni skupini ter analizami FIC, ARPF in sprejemnih diagnoz glede na skupine (kontingenčne tabele s  $\chi^2$  testi povezanosti in prepoznavanjem odstopajočih celice – prav tako izpuščene zaradi preobsežnosti) potrjujeta, da je smiselno razlikovati šest skupin sprejemov. Njihove značilnosti povzema tabela 4.

Izdelali smo tudi dvoskupinski model porazdelitve sprejemnega in odpustnega FIM. Oblika porazdelitve FIM ob sprejemu in odpustu (slika 2) nakazuje, da je primeren model opaženih vrednosti FIM zmes dveh unimodalnih porazdelitev. Lahko predpostavimo, da sta ti porazdelitvi normalni (kot pri preprosti obliki analize latentnih razredov), zaradi omejenega razpona lestvice FIM pa je še boljši model vsakega od razredov omejena normalna slučajna spremenljivka (pri kateri se vse latentne vrednosti pod oziroma nad določeno mejo izrazijo kot minimalna oziroma maksimalna opažena vrednost). Najprej smo z algoritmom EM določili parametre zmesi dveh normalnih porazdelitev ob sprejemu, nato pa ocenjeni delež obeh komponent zmesi pustili konstanten ob sprejemu in odpustu, ocenjeni povprečji in standardna odklona uporabili kot začetne približke ter ob sprejemu in ob odpustu ocenili parametre zmesi omejenih normalnih porazdelitev po metodi največjega verjetja. Na tem mestu naj omenimo le še, da je bilo ujemanje modela s podatki odlično, podrobnejša predstavitev in rehabilitacijska interpretacija modela pa zahteva poseben prispevek, ki ga nameravamo objaviti v mednarodnem znanstvenem časopisu.

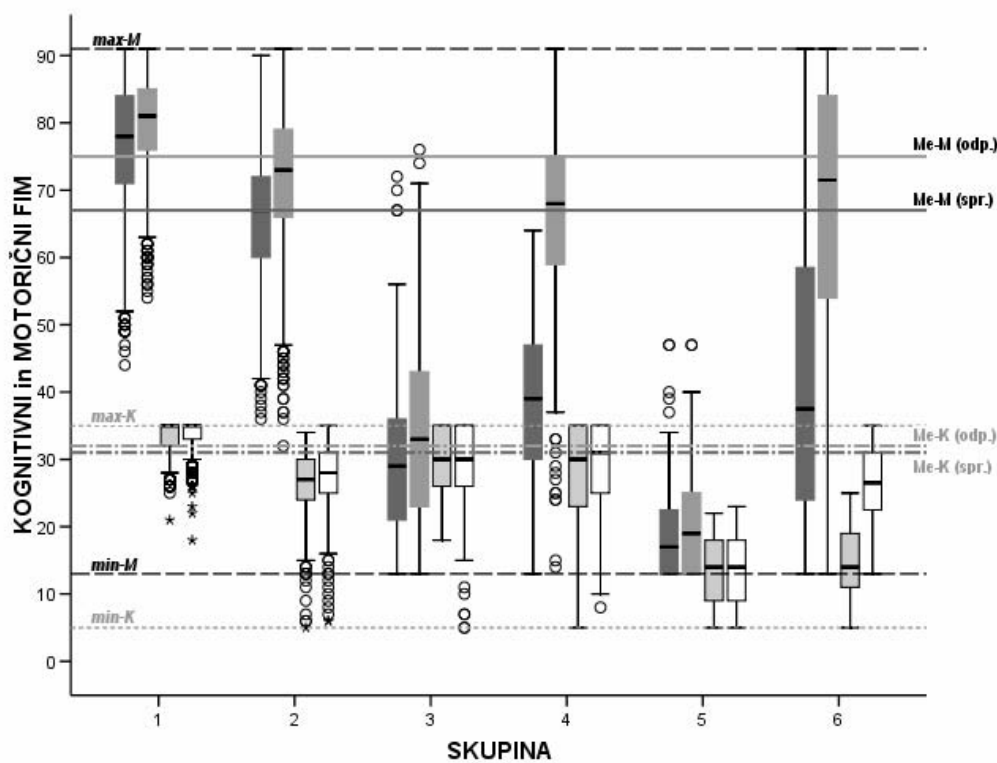
**Tabela 4** Značilnosti skupin, dobljenih z združevanjem sprejemov v 6 skupin.

Skupina	Delež sprejemov	FIM ob sprejem		Napredek na FIM		Prevladujoče diagnoze
		motorični	kognitivni	motoričnem	kognitivnem	
1	45% (~1/2)	blizu maksimuma	blizu maksimuma	majhen (nekaj točk)	ne	nevrološka stanja, amputacije
2	22% (~1/5)	nekoliko nadpovprečen	podpovprečen	5 točk	2 točki	možganska kap, amputacije
3	12% (~1/7)	nizek (povprečje 30)	povprečen (okoli 30)	okoli 5 točk	ne	različne (okvare hrbtenjače, možg. kap, ortopedska stanja idr.)
4	14% (~1/7)	okoli 40	okoli 30	zelo velik (30 točk)	majhen (2 točki)	~1/3 možganska kap, ~1/5 okvare hrbtenjače, ~1/5 ortopedska stanja
5	5% (1/20)	zelo nizek (povprečje 20)	zelo nizek (povprečje < 15)	ne	ne	glede na splošni delež: možganska kap in okvare možganov
6	2% (~1/50)	povprečje okoli 45	povprečje okoli 15	zelo velik (tipično 25t)	zelo velik (tipično 12t)	okvare možganov oz. poškodbe glave



**Slika 8** Profili skupin (ocenjeni sočasni 95% intervali zaupanja za povprečja) na vseh štirih spremenljivkah, vključenih v postopek razvrščanja, za razpoznanih 6 skupin sprejemov.





**Slika 9** Porazdelitev sprejemnega (temnejše) in odpustnega (svetlejše) motoričnega (temnejše sivi zaboji brez obrobe) in kognitivnega FIM (zaboji z obrobo, svetlo sivi in beli) za sprejeme, razvrščene v šest skupin. Za lažjo interpretacijo sta za obe podlestvici FIM z vodoravnimi črtami prikazana možni razpon ocen (črtkano za motorični, pikčasto za kognitivni FIM) in medianska ocena (polni črti za motorični, črtkano-pikčasti črti za kognitivni FIM).

## Razprava

S strokovnega vidika so ocenjevalne lestvice nepogrešljiv pripomoček v procesih načrtovanja obravnave in spremljanja bolnikov. Enako pomembne so v kliničnem raziskovalnem delu. Zato vodstvo IRSR vsestransko podpira ocenjevanje z lestvico FIM in izdelavo analiz, ki pomagajo odgovarjati na aktualna strokovna vprašanja in olajšujejo uvid v rehabilitacijsko dejavnost v obravnavanem obdobju. Seveda pa se pri uporabi tovrstnih instrumentov srečujemo z vrsto vprašanj in dilem, zlasti če se opredelimo za en sam merski instrument, ki naj bi bil oziroma postal univerzalno orodje.

Skozi opravljene analize, ki jih je tu predstavljen le manjši del, smo temeljito preverili različne merske

lastnosti FIM. Nekatere so se pokazale kot odlične (npr. zanesljivost z vidika notranje skladnosti, saj so vrednosti Cronbachovega  $\alpha$  okoli 0,95), nekatere kot primerne pri veliki večini bolnikov (npr. občutljivost, saj je le okoli 1% bolnikov sprejetih z največjo možno vrednostjo skupnega in motoričnega FIM), nekatere pa so težko preverljive (zlasti kriterijska veljavnost) ali praktično neugotovljive (iz bioloških in etičnih razlogov to velja za zanesljivost z vidika stabilnosti v času). Pričakovano smer in velikost sprememb skupne ocene FIM ter ocen na obeh podlestvicah se da oceniti na podlagi združevanja bolnikov v skupine, s takimi modeli pa se da tudi predvideti ocene uspešnosti rehabilitacije ob morebitni spremembi strukture sprejetih bolnikov.

Analize so potrdile, da bi se morebitno financiranje rehabilitacijske dejavnosti na podlagi vrednosti FIM moralo opirati zgolj na značilnosti

sprejetih bolnikov (v smislu razvrstitve ARPF), a tudi napredek na lestvici FIM je lahko pomemben in uporaben z vidika zdravstvene politike.

Ocenjevanje s FIM je za rehabilitacijsko ustanovo tudi sicer pomembno z vidikov vodenja, upravljanja s kakovostjo in informatike. IRSR je zato vključil spremljanje FIM v novi bolnišnični informacijski sistem, kar bo omogočalo redne analize v obliki poročil (reporting) in sprotnih poizvedb (OLAP). Za strokovne in raziskovalne potrebe bomo morda izdelali tudi intranetni portal z rezultati zahtevnejših statističnih analiz.

Poleg predstavitve našega dela in spoznanj je namen prispevka tudi zavzemanje za dvig nivoja statistične prakse v bolnišnicah in drugih ustanovah na področju zdravstva. Poudarek je na kakovostnem prikazu podatkov<sup>8</sup>, kjer je stanje v Sloveniji posebej daleč od zaželenega. Z informatizacijo življenja in napredkom informacijske tehnologije sta se zaželenost in dostopnost izdelave statističnih poročil in grafičnih prikazov podatkov povečala neprimerno bolj, kot se je dvignila stopnja statistične pismenosti. V tako pomembni dejavnosti, kot je zdravstvo, si državni organi, javni zavodi, ponudniki programja ter zdravniki, informatiki in vsi drugi vpleteni strokovnjaki delijo pomembno odgovornost, da ne izvajajo neustreznih statističnih analiz, ki vodijo do neustreznih sklepov, ter ne izdelujejo risbosmetja (chartjunk<sup>9</sup>), ki zavaja uporabnike informacij. Tudi tistim, ki se jim zdi paradigma medicine, podprte z dokazi, premalo oprijemljiva, nudijo pretresljivi primeri iz prakse<sup>10,11</sup> dovolj dokazov, da bi več znanja na področju analize podatkov na vseh nivojih zdravstvenega sistema lahko veliko prispevalo h kakovosti v zdravstvu.

## Literatura

1. Cohen ME, Marino RJ: The tools of disability outcomes research functional status measures. *Arch Phys Med Rehab* 2000; 81(12 Suppl. 2): S21-S29.
2. Granger CV, Gresham GE: *Functional assessment in rehabilitation medicine*. Baltimore 1984: Williams & Wilkins.
3. Pajntar M, Leskošek B: Rezultati projekta: Kakovost v zdravstvu Slovenije. *Zdrav vestn* 2002; 71: 765-771.
4. Burger H, Goljar N (ur.): *Ocenjevanje izida v medicinski rehabilitaciji – zbornik predavanj, 14. dnevi rehabilitacijske medicine, 4. in 5. april 2003*. Ljubljana 2003: Inštitut Republike Slovenije za rehabilitacijo.
5. Burger H: Poročilo o ocenjevanju z Lestvico funkcijske neodvisnosti FIM na Inštitutu RS za rehabilitacijo v letu 2004. Ljubljana 2005: Inštitut Republike Slovenije za rehabilitacijo.
6. Eagar K: The Australian National Sub-Acute and Non-Acute Patient Classification (AN-SNAP): Report of the National Sub-Acute and Non-Acute Casemix Classification Study. Wollongong 1997: Centre for Health Service Development, University of Wollongong.
7. Brock KA, Vale SJ, Cotton SM: The effect of the introduction of a case-mix-based funding model of rehabilitation for severe stroke: an Australian experience. *Archives of Physical Medicine and Rehabilitation* 2007; 88(7): 827-832.
8. Wainer H, Thissen D: Graphical data analysis. *Ann Rev Psychol* 1981; 32: 191-241.
9. Tufte ER: *The visual display of quantitative information*. Cheshire 1998: Graphics Press.
10. Mohammed MA, Cheng KK, Rouse A, et al: Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001; 357(9254): 463-467.
11. Marshall T, Mohammed MA, Rouse A: A randomized controlled trial of league tables and control charts as aids to health service decision-making. *Int J Qual Health Care* 2004; 16(4): 309-315.

Strokovni članek ■

## Farmakogenomika in informacijski izzivi

## Pharmacogenomics and informatical challenges

**Polonca Ferk, Branimir Leskošek**

**Izvleček.** Farmakogenomika je interdisciplinarno področje, ki združuje znanja farmakologije in genomike. Njen osnovni namen je opredeliti, kako in v kolikšni meri posameznikov edinstven genetski zapis določa farmakološki odgovor na specifično zdravilo in terapijo. Z namenom odkriti klinično uporabne povezave med genetskim zapisom in farmakološkim odgovorom je potrebno iz različnih virov pridobiti in analizirati velike količine podatkov, pri čemer je ključna uporaba ustreznih informacijskih orodij. V prispevku je opisan farmakogenomski primer iz klinične prakse ter v nadaljevanju sodobni izzivi informatike v farmakogenomiki.

**Abstract.** Pharmacogenomics is an interdisciplinary field which merges knowledge from pharmacology and genomics. Its basic aim is to find out how the unique genotype of an individual determines specific pharmacological responses. To be able to discover clinically relevant associations between the specific genotype and pharmacological response, huge amounts of data should be collected from different sources and further analyzed which is possible only by using suitable and effective informatics tools. In the paper a clinically relevant pharmacogenomic example is first presented and then contemporary informatics challenges are described.

■ **Infor Med Slov:** 2008; 13(1): 33-40

---

Instituciji avtorjev: Katedra za farmakologijo, Medicinska fakulteta, Univerza v Mariboru (PF), Inštitut za biomedicinsko informatiko, Medicinska fakulteta, Univerza v Ljubljani (BL).

Kontaktna oseba: Polonca Ferk, Medicinska fakulteta, Univerza v Mariboru, Slomškov trg 15, 2000 Maribor. email: polonca.ferk@guest.arnes.si.

## Uvod

### Farmakogenomika in njen pomen

Farmakogenomika je interdisciplinarno področje, ki združuje znanja farmakologije in genomike. Njen osnovni namen je opredeliti, kako in v kolikšni meri posameznikov edinstven genetski zapis določa farmakološki odgovor na specifično zdravilo in terapijo. Končni cilj je optimizacija in individualna prilagoditev že prve terapije (izbira zdravila, odmerka, režima odmerjanja; *angl. gene-dependent dosing*), vse v smislu:

- večje učinkovitosti,
- večje varnosti,
- zmanjšanja neželenih stranskih učinkov terapije (kar je še posebej pomembno pri zdravilih, ki imajo ozko terapevtsko območje),
- posledično pa tudi v smislu boljšega sodelovanja bolnika v procesu zdravljenja,
- ter nenazadnje, v smislu racionalizacije terapije s farmakoeconomskega vidika (zmanjšanje stroškov zdravljenja).

### Farmakogenomika in genetska raznolikost med posamezniki

Osnovo razvoja farmakogenomike predstavlja genetska raznolikost med posamezniki, v glavnem polimorfizmi posameznih nukleotidov (*angl. single-nucleotide polymorphisms –SNPs*) ter mikrosatelitski polimorfizmi (kratke tandemske ponovitve DNA). Med genetskimi polimorfizmi je potrebno prepoznati tiste, ki so informativni za napoved izida zdravljenja z izbranim zdravilom.

Informativni genetski polimorfizmi se intenzivno iščejo na različnih farmakodinamičnih in farmakokinetičnih nivojih: v genih za presnovne encime, za tarčne molekule za zdravila (receptorje, encime, ionske kanale, prenašalne molekule), ...

### Farmakogenomske raziskave

Dva osnovna pristopa v farmakogenomskih raziskavah sta:

- pristop “od genotipa k fenotipu”, pri katerem najprej identificiramo tarčne gene (preiskovanje genomskih podatkovnih baz, uporaba metod primerjalne genomike), poiščemo genetske polimorfizme v teh genih (z uporabo *high-throughput* eksperimentalnih metod), nato pa zberemo in s specifično genetsko različico povežemo fenotipske (molekularne, celične, klinične) podatke.
- pristop “od fenotipa h genotipu”, kjer najprej poiščemo značilne, merljive in klinično uporabne razlike v fenotipskih značilnostih med posamezniki (npr. časovni potek koncentracij zdravila v plazmi po aplikaciji določenega odmerka), v drugem koraku pa poiščemo gene in genotipe za pojasnitev fenotipske raznolikosti (uporaba živalskih modelov, metod primerjalne genomike, mikromrež cDNA, literaturnih in DNA-podatkovnih baz).

V obeh primerih je potrebno potrditi klinični pomen ugotovljene povezave med specifičnima genotipom in fenotipom.

Cilj farmakogenomskih raziskav je uvedba molekularno-genetskega testiranja informativnih in za specifično terapevtsko indikacijo validiranih genetskih polimorfizmov v klinično prakso.

### Farmakogenomika v klinični praksi – primer

Farmakogenomska odkritja se postopoma in v vedno večji meri vključujejo v klinično prakso, in sicer v smislu:

- izbire primernega zdravljenja
- napovedi poteka bolezni (prognoze),

- (zgodnje) diagnostike,
- napovedovanja tveganja oz. nagnjenosti za razvoj bolezni.

Kot primer navajamo povezavo med genetskimi različicami visoko polimorfne gena CYP2D6, ki nosi zapis za enega izmed najpomembnejših jetrnih oksidativnih encimov iz družine citokromov P450, in uporabo tamoksifena v preventivi oz. zdravljenju podtipa raka dojke s prekomerno izraženim estrogenskim receptorjem (ER+). Genotip CYP2D9 se je namreč izkazal za neodvisen napovedni dejavnik kliničnega izida zdravljenja zgodnjega raka dojke postmenopavzalnih bolnic s tamoksifenom.<sup>1</sup>

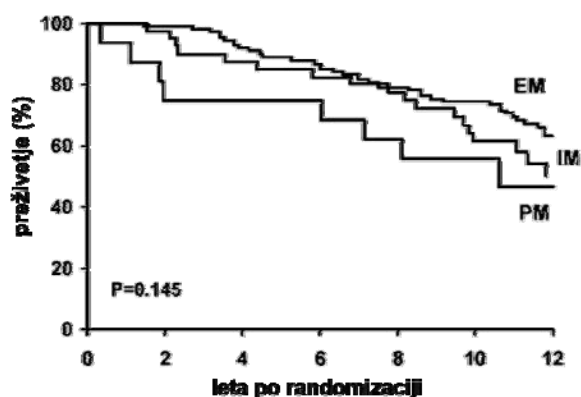
Za gen CYP2D6 je doslej znanih več kot 63 alelnih različic, ki kodirajo funkcionalno različni encim, z večjo oz. manjšo presnovno aktivnostjo. Stopnja aktivnosti encima CYP2D6 ima velik vpliv na plazemske koncentracije in s tem na učinke zdravil, ki se presnavljajo preko tega encima. Približno četrtnina zdravil, ki je danes na tržišču, se presnavlja s pomočjo encima CYP2D6, npr. številni blokatorji beta, antipsihotiki, antidepresivi, antiaritmiki, varfarin.<sup>2</sup> Pri tamoksifenu, ki je farmakološko neaktiven (t.i. predzdravilo), pa izkoriščamo njegovo presnovo pod vplivom encima CYP2D6 za pretvorbo v farmakološko aktivne presnovke (zlasti endoksifen). Slednji z vezavo na estrogenske receptorje delujejo kot anti-estrogeni in zavirajo celično proliferacijo. Hitrost presnavljanja tamoksifena pri posamezni bolnici je odvisna od stopnje presnovne aktivnosti izooblike encima CYP2D6, na slednjo pa vplivajo zlasti genetski polimorfizmi v genu CYP2D6 ter sočasno jemanje drugih zdravil, ki so induktorji oz. inhibitorji encima CYP2D6. Najpogostejši med aleli CYP2D6, ki vodijo v odsotnost encimske aktivnosti, je alel CYP2D6\*4, prisoten pri približno eni petini belcev. Posledica različne aktivnosti encima CYP2D6 in s tem hitrosti presnove tamoksifena so različni fenotipi, ki zelo dobro korelirajo s kliničnim izidom in prognozo pri terapiji raka dojke s tamoksifenom:<sup>1</sup>

- Počasni metabolizatorji (PM – *angl. poor metabolizers*) tamoksifen presnavljajo počasi in imajo posledično nizke plazemske koncentracije endoksifena, na terapijo s tamoksifenom pa se ne odzovejo zadostno in imajo slabšo prognozo (čas do ponovnega pojava bolezni je krajši, preživetje slabše), zato je pri njih potrebno razmisliti o alternativni terapiji.
- Srednje hitri metabolizatorji (IM – *angl. intermediate metabolizers*)
- Hitri metabolizatorji (EM – *angl. extensive metabolizers*) presnavljajo tamoksifen hitro in tako dosegajo visoke koncentracije endoksifena v plazmi. Posledično je njihov odziv na terapijo s tamoksifenom hiter in učinkovit, prognoza pa boljša.

Leta 2006 je ameriška agencija za hrano in zdravila (FDA – *angl. food and drug administration*) odobrila enostaven molekularno genetski test, s katerim še pred začetkom zdravljenja bolnic z ER+ rakom dojke določimo genotip CYP2D6 in na osnovi tega predvidimo odziv posamezne bolnice na terapijo s tamoksifenom.

**Tabela 1** Farmakogenomika tamoksifena v zvezi z encimom CYP2D6.

fenotip	genetska različica CYP2D6	plazemske koncentracije endoksifena
PM	prisotna 2 alela CYP2D6*4	nizke
IM	prisoten 1 alel CYP2D6*4	srednje
EM	odsotnost CYP2D6*4	visoke



**Slika 1** Preživetje (v %) po zdravljenju s tamoksifenom glede na metabolni status bolnic (PM, IM, EM), določen z genotipom CYP2D6.<sup>1</sup>

## Povezovanje farmakogenomike in informatike

S farmakogenomskimi raziskavami pridobivamo ogromno količino genetskih, kliničnih in farmakoloških podatkov. Učinkovito in urejeno shranjevanje ter prikaz podatkov v podatkovnih zbirkah, iskanje informacij, priprava podatkov za nadaljnjo obdelavo in analizo ter odkrivanje novih povezav in znanj so možni le z razvojem in uporabo učinkovitih informacijskih in programskih orodij. Potreba po urejenih podatkovnih bazah se je povečala po razkritju zaporedja nukleotidov, odkritju genov in začetku opredeljevanju funkcionalnih vidikov posameznih genetskih elementov v človeškem genomu v okviru projekta "Človeški genom" (angl. *Human Genome Project*). Lažji dostop do teh informacij in njihovo uporabo pri raziskovalnem delu je omogočil zlasti NCBI (angl. *National Center for Biotechnology Information*) z razvojem podatkovnih baz z zbranimi kratkimi opisnimi informacijami (ter povezavami do drugih podatkovnih virov), dostopnimi na svetovnem spletu, npr. *GenBank DNA sequence database*, dbSNP, GEO, OMIM (angl. *Online Mendelian Inheritance in Man*).

Specifična genetska različica lahko ima prevladujoč vpliv na posameznikov odziv na terapijo, največkrat pa je farmakološki odgovor določen poligeno (s kombinacijo vseh posameznikovih informativnih genetskih polimorfizmov), ob součinkovanju drugih dejavnikov (starost, spol, funkcionalnost notranjih organov, sočasna terapija in interakcije z drugimi zdravili, narava bolezni, splošno zdravstveno stanje, način prehranjevanja, telesna aktivnost).

V primerih multifaktorsko določenega farmakološkega odgovora se posebej izrazito pokaže potreba po informacijskem sistemu, v katerem bi bili relevantni genetski podatki zbrani in korelirani s kliničnimi podatki. Tak sistem bi na osnovi enostavnega predhodnega genetskega testiranja nudil pomoč pri izbiri za posameznika ustrezne in optimalne terapije, vključno s predlogi za morebiti ustrežnejšo alternativno terapijo. Za nekatere konkretne klinične primere so že postavljeni matematični algoritmi, ki z zelo veliko verjetnostjo napovejo odgovor bolnika na specifično terapijo.

## Informacijski izzivi v farmakogenomiki

Urejeno zbiranje in vrednotenje podatkov je pomembno tudi za spremljanje dejanske uspešnosti (na osnovi bioloških označevalcev) individualno prilagojene terapije. Primer v Sloveniji je spremljanje uspešnosti zdravljenja s transtuzumabom (Herceptin®, Roche), ki je monoklonsko protitelo za usmerjeno zdravljenje bolnic z metastatsko obliko raka dojke v primerih, ko tumorji čezmerno izražajo humani receptor za endotelijski rastni dejavnik (HER2) oz. imajo povečano število kopij gena HER2.<sup>3</sup>

Farmakogenomika je relativno nova veda, s še veliko raziskovalnimi priložnostmi in bioinformacijskimi izzivi. Slednje lahko znotraj farmakogenomike razdelimo na več področij:<sup>4</sup>

1. prikaz raznolikosti farmakogenomskih podatkov
2. razvoj standardov za izmenjavo podatkov
3. združevanje podatkov iz različnih virov
4. odkrivanje znanja iz literature
5. povezava med prostorskimi strukturami in genetsko raznolikostjo
6. uporaba primerjalne genomike
7. upravljanje z laboratorijskimi podatki
8. zaščita zaupnosti in zasebnosti osebnih kliničnih podatkov

### **Prikaz raznolikosti farmakogenomskih podatkov**

Začetni izziv farmakogenomike je izdelava ustreznega informacijskega podatkovnega modela, ki bi omogočal enostavno shranjevanje in prikaz podatkov. Glede na različna področja, s katerimi se farmakogenomika povezuje, so podatki zelo raznovrstni:

- a) Genomski podatki (sekvence DNA). Obstaja večje število standardov za zapis genomskih podatkov kot je npr. BIOML (*angl. Biopolymer Markup Language*) ali pa standardi, ki jih predlagata zbirki *Genbank* in *Human Genome Database*. Obstaja več brskalnikov po človeškem genomu, ki omogočajo različen pogled na gensko strukturo: <http://genome.ucsc.edu/>, <http://www.ensembl.org/>, <http://vega.sanger.ac.uk/>, <http://www.ncbi.nlm.nih.gov/genome/guide/human/>, <http://pipeline.lbl.gov>
- b) Molekularni in celični podatki (encimska kinetika, mikromreže cDNA). Težko jih je opisati enoumno in hkrati primerno za

računalniško obdelavo, zato so standardi za opis takšnih podatkov še posebej potrebni.

- c) Klinični podatki. S stališča farmakogenomike je najpomembnejša povezava genomskih in molekularno-celičnih podatkov s kliničnimi. Obstaja več različnih standardov za opis kliničnih podatkov. V evropskem in tudi svetovnem merilu prevladujeta standard za klasifikacijo bolezni Mednarodna klasifikacija bolezni (MKB) in terminološki standard SNOMED (*angl. Systematized Nomenclature of Medicine*), pri katerem je najbolj razširjen SNOMED-CT (*angl. Clinical Terms*). MKB obstaja v več revizijah in je preveden tudi v slovenščino.<sup>5</sup> SNOMED-CT<sup>6</sup> vsebuje preko 300.000 izrazov (med drugim je možna enoumna povezava z MKB).

Vse tri predstavljene vrste podatkov poskuša povezati farmakogenomska zbirka znanja PharmGKB,<sup>7,8</sup>

### **Razvoj standardov za izmenjavo podatkov**

Standarde za izmenjavo biomedicinskih podatkov lahko razdelimo na:

- slovarje izrazov, ki omogočajo strukturirani opis prostega besedila bibliografskih enot ali kliničnih podatkov. Veliko slovarjev je bilo vzpostavljenih v okviru projekta UMLS (*angl. Unified Medical Language System*)<sup>9</sup> v ameriški državni medicinski knjižnici (NLM – National Library of Medicine). Kot primer opisa bibliografskih enot navajamo podatkovno zbirko Medline<sup>10</sup> kjer so vsakemu povzetku dodane ključne besede iz slovarja MeSH (*angl. Medical Subject Headings*),<sup>11</sup> ki je preveden tudi v slovenščino.<sup>12</sup> Za opis kliničnih podatkov pa poleg že omenjenega terminološkega slovarja SNOMED-CT obstaja še cela množica posebnih slovarjev, ki jih lahko uporabljamo pri farmakogenomiki kot so npr: imena genov pri človeku, imena zdravil in zdravilnih učinkovin, oznake za stranske učinke zdravil.<sup>11</sup>

- Druga veja so tehnični standardi, kjer je zaradi razumljivosti in enostavnosti najbolj priljubljen standard XML (*angl. eXtensible Markup Language* – razširjeni označevalni jezik), ki omogoča izdelavo lastnih označevalnih jezikov.<sup>13</sup> Zaželeno je, da podatki zapisani v XML obliki vključujejo tudi semantično razlago vsebine. V ta namen obstajajo različni načini preverjanja veljavnosti s shemami. Najbolj znana shema je shema XML, ki jo razvija konzorcij W3C, obstaja pa še več drugih.<sup>13</sup>

### **Združevanje podatkov iz različnih virov**

Iskanje po različnih podatkovnih zbirkah je sestavni del raziskovanja tudi v farmakogenomiki. Npr. za iskanje tridimenzionalne strukture proteina moramo združiti podatke iz vsaj treh različnih genetskih podatkovnih zbirk kot so GenBank, dbSNP, in OMIM ter iz literature. Za pridobivanje informacij o delovanju in medsebojnem učinkovanju zdravil moramo raziskovati po različnih, večinoma za javnost zaprtih zbirkah. Najlažje si pri organizaciji informacij pomagamo tako, da izdelamo svojo podatkovno zbirko, ki vsebuje zelene podatke iz različnih virov. Pri tem se moramo zavedati, da takšno združevanje v začetni fazi in tudi dolgoročno zahteva veliko človeških in finančnih virov, saj je potrebno izdelati in stalno vzdrževati vmesnike, ki podatke iz originalnih virov pretvarjajo v našo obliko. Morda je prihodnost v izdelavi javno dostopnih spletnih vmesnikov, ki za posamezno iskalno zahtevo preiščejo več virov in vrnejo rezultat v poenoteni obliki.

### **Odkrivanje znanja iz literature**

Najbolj zanimiva bibliografska podatkovna zbirka za odkrivanje znanja je Medline/PubMed,<sup>10</sup> v kateri obstaja skoraj 20 milijonov zapisov o biomedicinskih publikacijah. Kot smo že omenjali je pridobivanje informacij iz pisanih člankov enostavno za človeškega bralca, veliko težje pa za računalnike. Po drugi strani pa lahko računalniki veliko hitreje pregledajo veliko količino besedila

kot ljudje. Avtomatizacija iskanja informacij s tehnikami procesiranja naravnega jezika (*NLP – Natural Language Processing*) iz literature je bistveni del sistemov za odkrivanja znanja, ki se najbolj uporabljajo predvsem na področju genetike. Sistemi za odkrivanje znanja se lahko uporabljajo za odkrivanje novih povezav, npr. če poznamo dve različni povezavi s skupnim imenovalcem (*če je koncept X povezan s konceptom Y in hkrati koncept Z povezan s konceptom Y, potem lahko sledi, da je koncept X povezan s konceptom Z*) ali za izbiro ožje skupine podatkov iz velike množice (npr. podatki o izražanju mRNA pri mikromrežah). Sistem lahko uporabljamo tudi za zožanje ali odkrivanje novih potencialno zanimivih raziskovalnih področij. Tudi v Sloveniji smo z javno dostopnim sistemom Bitola<sup>14</sup> razvitim na Inštitutu za biomedicinsko informatiko v samem vrhu omenjenih raziskav.

### **Upravljanje z laboratorijskimi podatki**

Kot je pronicljivemu bralcu verjetno znano, so dobri eksperimentalni podatki, prav tako kot teoretični, ključni za uspešno raziskovalno delo in jih ne smemo zanemarjati. Še vedno velja pregovor, da slabi vhodni podatki ne morejo dati dobrih rezultatov (*angl. garbage in garbage out*). Podobno velja tudi v farmakogenomiki, kjer je samo dober oz. odličen laboratorijski informacijski sistem lahko osnova za dobre teoretične rezultate. Dober informacijski sistem lahko definiramo kot zanesljiv, varen (tako s stališča zaščite podatkov kot s stališča nepooblaščenih vpogledov v podatke), povezljiv in vedno dostopen sistem, ki ima vse dele in postopke tudi ustrezno dokumentirane.

### **Zaščita zaupnosti in zasebnosti osebnih kliničnih podatkov**

Uporaba informacijske tehnologije v farmakogenomiki omogoča lažjo in hitrejšo obdelavo podatkov, hkrati pa je njihova morebitna zloraba lažja. Zelo pomembno je, da se pred vsako objavo npr. na spletu zakrijejo vsi osebni podatki oz. se javno objavlja samo razosebljene podatke.



Seveda je istočasno potrebno ustrezno zaščititi informacijski sistem, v katerem so shranjeni vsi podatki, in sicer tako z informacijskimi kot tudi s fizičnimi ukrepi. Za varovanje podatkov priporočamo uporabo pravil (certificiranje je zelo drago), skladnih s priporočili za upravljanje z informacijsko varnostjo glede na družino standardov ISO27000.<sup>15</sup>

## Razprava

Terapevtski pristop prihodnosti je nedvomno v optimalni obravnavi vsakega posameznega bolnika na osnovah njegovih individualnih specifičnih lastnosti, med njimi tudi na osnovi specifičnega genetskega zapisa. Tako uporaba farmakogenomskih znanj za prilagajanje terapije obeta spremembe v smislu prehoda iz izdelave in uporabe zdravil za množično in splošno uporabo v zdravila za individualno terapijo. Kljub nedvomnim koristim takega pristopa za zdravje bolnika ter konkretnim primerom, kjer je učinkovitost farmakogenomskega pristopa jasno dokazana, pa nekateri dejavniki zavirajo oz. omejujejo hitrejši in obsežnejši prehod farmakogenomskih znanj v klinično prakso:

- obstoječe doktrine na principu "trial-and-error" ter zadržki zdravnikov do sprejemanja novih tehnologij in smernic zdravljenja
- farmacevtska industrija (k posamezniku usmerjeno zdravljenje pomeni manjši trg za posamezno zdravilo)
- farmakogenomski podatki so problematični s stališča zavarovalnic in zdravstvenih sistemov, ki bi jih utegnilo zanimati posameznikovo tveganje za razvoj posameznih bolezni in farmakološkega odziva na specifično terapijo. Kako torej zagotoviti informacijsko varnost in opredeliti etična merila?

## Izzivi za prihodnost

Hitro uvajanje izsledkov farmakogenomskih raziskav v klinično prakso lahko učinkovito doprinese k dvigu vseh vidikov kakovosti zdravljenja in preventive (strokovni, ekonomski in družbeni vidik). Navajamo nekaj možnih izzivov za prihodnost:

- Prenos farmakogenomskih znanj v klinično prakso hitreje ter v čim večji meri tudi v slovenskem prostoru. Razvoj smernic za razvoj, uporabo in posredovanje farmakogenomskih podatkov.
- Za "reorganiziranje" mišljenja zlasti izkušenih zdravnikov, ki že vrsto let zdravijo po obstoječih doktrinah "trial-and-error", bodo potrebni zelo prepričljivi dokazi o smiselnosti prilagajanja terapije na osnovi posameznikovega genetskega profila s terapevtskega in farmakoekonomskega vidika.
- Izboljšati izobraženost (skupaj z uvedbo v izobraževalni proces na fakultetah) zdravnikov in drugih zdravstvenih delavcev s področja farmakogenomike, farmakokinetike in farmakodinamike, zlasti s klinično aplikativnega vidika, da se novih znanj ne bomo bali in zaradi nevednosti zavirali napredka, ki bi ga ta znanja prinesla v klinično obravnavo bolnikov.
- Genetski zapis posameznika bo del zdravstvene dokumentacije oz. elektronskega zdravstvenega zapisa. Pojavlja se vprašanje, v kakšni obliki naj bo shranjen, na kakšen način naj bo dostopen (s certifikatom na osebni izkaznici oz. kartici zdravstvenega zavarovanja), kdo in kdaj naj ima dostop do njega?
- Prenos farmakogenomskih podatkov v povzetke temeljnih značilnosti zdravila (SmPC - angl. *Summary of Product Characteristics*) ter v navodila za uporabo zdravil.
- Za uspeh farmakogenomike tudi v Sloveniji je potrebno čim prejšnje sprejetje ustreznih

informacijskih standardov na državnem nivoju (uskklajenih z EU), na področju molekularne biologije, še bolj pa na področju zbiranja kliničnih podatkov. Nekatere aktivnosti v smislu podpore strategiji eZdravje2010 se na Ministrstvu za zdravje že dogajajo, bi jih pa bilo potrebno pospešiti in to najprej na področjih, kjer vlada veliko soglasje, kot je npr. pri uporabi standarda SNOMED-CT za klinične izraze v kliničnih informacijskih sistemih.

### Literatura

1. Goetz MP, Knox SK, Suman VJ, et al.: The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res Treat* 2007; 101(1): 113-121.
2. Goetz MP, Kamal A, Ames MM: Tamoxifen pharmacogenomics: the role of CYP2D6 as a predictor of drug response. *Clin Pharmacol Ther* 2008; 83(1): 160-166.
3. Leskošek B: A clinical research information system: An example of prospective observational study in oncology. *Proceedings of the American medical informatics association AMIA*, sprejeto v objavo 2008.
4. Altman RB, Klein TE: Challenges for biomedical informatics and pharmacogenomics. *Annu Rev Pharmacol Toxicol* 2002; 42:113-133.
5. Mednarodna klasifikacija bolezni, 10. revizija. Ljubljana 2005: IVZ.
6. SNOMED-CT - Wikipedia Web Page. [http://en.wikipedia.org/wiki/SNOMED\\_CT](http://en.wikipedia.org/wiki/SNOMED_CT), 16.06.2008.
7. Altman RB: PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. *Nat Genet* 2007; 39(4): 426.
8. PharmGKB Web Page. <http://pharmgkb.org>, 16.06.2008.
9. Unified Medical Language System (UMLS) Web Page. <http://www.nlm.nih.gov/research/umls/umlsmain.html>, 16.06.2008.
10. Medline/PubMed Web Page. <http://www.ncbi.nlm.nih.gov/pubmed/>, 16.06.2008.
11. Medical Subject Headings (MeSH) Web Page. <http://www.nlm.nih.gov/mesh/>, 16.06.2008.
12. MeSH v slovenščini, interno gradivo. Inštitut za biomedicinsko informatiko Ljubljana, 2007.
13. Specifikacija za XML - Wikipedia Web Page. <http://en.wikipedia.org/wiki/XML>, 16.06.2008
14. Hristovski D, Peterlin B: BITOLA - Biomedical Discovery Support System. <http://www.mf.uni-lj.si/bitola>, 16.06.2008
15. Družina standardov ISO27000 - Wikipedia Web Page. [http://en.wikipedia.org/wiki/ISO/IEC\\_27000-series](http://en.wikipedia.org/wiki/ISO/IEC_27000-series), 16.06.2008.

*Bilten SDMI* ■

## Vabilo na kongres MI 2008 z mednarodno udeležbo: Od e-zdravja k zdravju

Slovensko društvo za medicinsko informatiko (SDMI) prireja ob 20. obletnici svojega delovanja kongres MI 2008 z mednarodno udeležbo: Od e-zdravja k zdravju. Kongres bo v Termah Zreče od 5. do 7. oktobra 2008.

SDMI organizira vsako leto strokovno srečanje in vsaki dve leti kongres. Tokratni kongres MI 2008 je povezan z jubilejno 20. obletnico uspešnega delovanja društva. Poleg pregleda uspešne poti bo kongres s sodelovanjem tujih priznanih strokovnjakov obravnaval aktualne razvojne teme in omogočil izmenjavo informacij, pogledov in izkušenj s področja zdravstvene informatike v Sloveniji in v tujini.

Vabimo vas, da na kongresu aktivno sodelujete in v obliki referata ali posterja predstavite svoje delo, poglede in dognanja.

### Predvidena vsebina kongresa

- e-Zdravje kot ključno orodje za uresničevanje razvojnih ciljev v zdravstvu,
- ureditveni okvirji zdravstvenega informacijskega sistema Slovenije (eZIS):
  - pravna ureditev,
  - tehnična ureditev (standardi, arhitektura, modeli, infrastruktura, web, ...),
- razvojna dejavnost na področju informatike v zdravstvu (vsebina, rešitve, viri),

- dobra praksa pri uporabi informacijskih in telekomunikacijskih orodij v zdravstvu,
- IT in upravljanje procesov (znanje, kakovost, varnost, spremembe, organizacija),
- novi modeli, tehnologije in storitve v zdravstvu.

V okviru kongresa bo organiziranih več delavnic in okrogla miza o informatiziranju zdravstva gledano iz različnih perspektiv: upravljavcev / menedžmenta, zavarovancev, zdravstvenih strokovnjakov, administrativnih delavcev...

### Navodilo avtorjem

- Povzetki naj vsebujejo seznam avtorjev z navedbo organizacije oz. podjetja in elektronskim naslovom, naslov prispevka in povzete vsebine v obsegu 200 besed.
- Povzetke pošljite v elektronski obliki po elektronski pošti na naslov kongres@sdmi.si.
- Programski odbor bo avtorja prispevka obvestil o uvrstitvi v program kongresa in hkrati podal podrobnejša navodila za izdelavo in oblikovanje prispevkov – referatov oziroma posterjev.

### Pomembni datumi

- 30. junij 2008 – rok za oddajo povzetkov prispevkov,

- 10. julij 2008 – obvestilo avtorjem o uvrstitvi v program kongresa,
- 1. september 2008 – rok za oddajo končnih prispevkov.

Dodatne informacije o izvedbi in vsebini kongresa bodo objavljene v naslednjem obvestilu in na spletni strani Slovenskega društva za medicinsko informatiko <http://www.sdmi.si>, [info@sdmi.si](mailto:info@sdmi.si).

Uradna jezika kongresa bosta slovenščina in angleščina.

#### **Predsednik SDMI**

- Ivan Eržen

#### **Programski odbor**

- Leo Ciglencčki,
- Jože Gašperšič, predsednik,
- Rade Ilijaž,
- Janko Kersnik
- Brane Leskošek,
- Tomaž Marčun,
- Dorjan Marušič,

- Vesna Prijatelj,
- Drago Rudel,
- Igor Rus,
- Smiljana Slavec,
- Marjan Sušelj,
- Gordana Živčec Kalan.

#### **Organizacijski odbor**

- Nina Dolenc,
- Ivan Eržen,
- Andrej Orel,
- Mojca Paulin, predsednica,
- Danila Perhavec,
- Marija Zevnik.

Slovensko društvo za medicinsko informatiko  
Katedra za javno zdravje  
Zaloška 4, 1000 Ljubljana  
<http://www.sdmi.si>

■ **Infor Med Slov:** 2008; 13(1): 41-42