

# PRIMER UPORABE NAJPREPROSTEJŠEGA STATISTIČNEGA TESTA: ALI SE ZAHTEVNOST REHABILITACIJE BOLNIŠNIČNIH PACIENTOV POVEČUJE? AN APPLICATION OF THE SIMPLEST STATISTICAL TEST: IS THE DEMANDINGNESS OF INPATIENT REHABILITATION INCREASING?

doc. dr. Gaj Vidmar, univ. dipl. psih.  
Inštitut Republike Slovenije za Rehabilitacijo, Ljubljana

## Izvleček

### Izhodišča:

Na Inštitutu Republike Slovenije za rehabilitacijo (IRSR) smo zasnovali in izvedli različne analize zahtevnosti obravnave pacientov skozi čas za obdobje od leta 2000 dalje, med katere sodi tudi analiza pogostnosti zapletov in dodatnih diagnoz.

### Metode:

Analizo smo izvedli na podlagi medicinske dokumentacije – zdravstvenih kartonov pacientov. Za leta 2001, 2004 in 2007 smo naključno izbrali po 20 kartonov iz vsakega od šestih obravnavanih oddelkov IRSR. Ocenjeno število diagnoz in zapletov na pacienta smo med leti primerjali z asimptotičnim testom razlike med dvema Poissonovima razmerjema.

### Rezultati:

Število diagnoz na pacienta je bilo tako v letu 2004 kot v letu 2007 statistično značilno večje kot v letu 2001. Povečanje med letoma 2004 in 2007 ni statistično značilno. Tudi zabeleženo število zapletov na pacienta je bilo v letih 2004 in 2007 večje kot v letu 2001, a razlika ni statistično značilna.

### Zaključki:

Uporaba statističnega testa, ki si zasluži naziv najpreprostejši, nam je dala uporaben občutek za podatke, ki kažejo na povečevanje zahtevnosti bolnišnične rehabilitacije.

### Ključne besede:

Bolnišnična rehabilitacija, sprejemi, knjiga dežurnega zdravnika, medicinska dokumentacija, primerjava Poissonovih razmerij

## Abstract

### Background:

*The Institute for Rehabilitation, Republic of Slovenia (IRSR), has recently conducted several analyses of changes in demandingness of inpatient care since 2000, which include the study of frequency of complications and concomitant diagnoses.*

### Methods:

*The study was based on medical documentation – patient records. For 2001, 2004 and 2007, we randomly selected 20 records from each of the six studied wards of the IRSR. The estimated number of diagnoses and complications per patient was compared between the years using the asymptotic test for difference between two Poisson ratios.*

### Results:

*The number of diagnoses per patient was statistically significantly larger in 2004 as well as in 2007 compared to 2001. The increase from 2004 to 2007 was not statistically significant. The recorded number of complications per patient was also larger in 2004 and 2007 than in 2001, but the difference is not statistically significant.*

### Conclusions:

*The application of the »simplest statistical test« gave us a useful »feel of the data«, showing increased demandingness of inpatient rehabilitation.*

### Key words:

*Inpatient rehabilitation, admissions, on-duty physician log, medical records, comparison of Poisson rates*

## UVOD

Daljšanje življenjske dobe in napredek zdravstvenega varstva s seboj prinašata vse večjo zahtevnost bolnišnične obravnave tudi na področju rehabilitacije (1). Hkrati se pacienti tudi bolje zavedajo svojih pravic (2) in pričakujejo vse višji standard zdravstvenih storitev. Na Inštitutu Republike Slovenije za rehabilitacijo (IRSR) smo zato na pobudo vodstva zasnovali in izvedli različne analize zahtevnosti obravnave pacientov skozi čas za obdobje od leta 2000 dalje. Del teh analiz predstavlja primer, o katerem poročamo.

Prispevek pa ima tudi izobraževalen namen, saj predstavlja uporabo malo znanega, a izredno preprostega in široko uporabnega statističnega testa, ki lahko pride prav vsakemu zdravstvenemu strokovnjaku. Zahtevnost sodobnih statističnih metod za analizo kliničnih podatkov večini praktikov namreč močno otežuje razumevanje poročil o kliničnih poskusih v znanstveni literaturi, a izsledke številnih kliničnih poskusov je moč predstaviti in razumeti tudi z razmeroma preprosto statistiko. Zato želimo spodbuditi bralke in bralce k pridobivanju »občutka za podatke«, ki naj zamenja neproduktivno izgubljanje v podrobnostih zahtevnejših statističnih analiz in poskuse njihove izvedbe brez potrebnega predznanja.

Glavni izid številnih kliničnih poskusov je izid bolezni. Praviloma gre za to, ali se je izbrani dogodek (npr. ozdravitev) zgodil ali ne. Ozdravitev je enostaven dogodek, lahko pa gre za sestavljen dogodek, npr. bodisi smrt bodisi preživet srčni infarkt pri študiji ogroženih za srčnožilne bolezni (odsotnost dogodka v tem primeru predstavlja življenje brez infarkta ali smrt iz razloga, ki ni povezan s srčnožilnimi boleznimi). S sestavljenimi dogodki imamo opravka tudi na področju rehabilitacije (npr. kakršenkoli zaplet s protezo). Statistične metode analize preživetja (npr. test logrank ali Coxov regresijski model sorazmernih tveganj) pri primerjavi izida med skupinami v tovrstnih študijah upoštevajo različnost časa spremljanja bolnikov, a so za številne uporabnike medicinskih raziskav prezahtevne za ustrezno izvedbo oziroma pravilno razumevanje rezultatov. Tudi matematični temelji testa hi-kvadrat in Fisherjevega eksaktnega testa, ki se pogosto uporabljata za primerjavo deležev glavnega izida, če je čas spremljanja konstanten, niso splošno poznani oziroma razumljivi širši (strokovni) javnosti. Zato tu predstavljamo preprostejšo metodo, s katero lahko hitro ocenimo moč dokazov o razliki v izidu (npr. med dvema načinoma zdravljenja oziroma rehabilitacije).

Predstavljajmo si randomiziran klinični poskus z dvema približno enako velikima skupinama udeležencev, v katerem je izid, ki nas zanima, določen klinični dogodek. Bistveni podatek je število udeležencev, ki jih je v posamezni skupini doletel dogodek – v eni skupini naj jih bo  $a$ , v drugi pa  $b$ . Tedaj lahko testno statistiko za primerjavo med skupinama (3) izračunamo po preprostem obrazcu

$$z = \frac{a - b}{\sqrt{a + b}}$$

Razliko med številoma dogodkov torej delimo s kvadratnim korenem vsote števila dogodkov. Dobljena vrednost  $z$  se pod ničelno hipotezo, da imata obe zdravljenji enak učinek na tveganje za nastop dogodka, porazdeljuje približno po standardni normalni porazdelitvi, torej splošno znani Gaussovi porazdelitvi s povprečjem 0 in varianco 1, ki je tabelirana na koncu večine statističnih učbenikov. S tabelami, z elektronsko preglednico ali pa s pomočjo katerega od številnih spletnih virov s področja verjetnosti in statistike vrednost  $z$  preprosto pretvorimo v  $p$ , seveda pa marsikdo že na pamet ve, da  $z > 1,96$  pomeni  $p < 0,05$  ( $z > 2,58$  pa  $p < 0,01$ ). Pri tem smo seveda predpostavili, da je  $a$  večje izmed obeh števil dogodkov, tako da je vrednost  $z$  pozitivna in se nam ni treba ubadati z negativnimi vrednostmi (če sta  $a$  in  $b$  enaka, seveda brez računanja sklenemo, da je  $p = 1$ , in torej ničelno hipotezo obdržimo), ter da smo opravili dvosmerno testiranje (torej z alternativno hipotezo, da sta učinka različna v katerikoli smeri).

Test je, kot rečeno, približen, a navadno daje dovolj zanesljive rezultate. Pri randomizaciji je število udeležencev v obeh skupinah namreč praviloma (skoraj) enako, to pa velja tudi za čas spremljanja. Pogostnost kliničnega dogodka je praviloma majha, npr. manjša od 20%, često pa še mnogo manjša, torej lahko predpostavimo, da se število udeležencev v izbrani skupini, ki jih doletel dogodek, porazdeljuje po Poissonovi porazdelitvi (4). Če skupno število dogodkov ( $a + b$ ) ni premajhno, npr. če ni manjše od 20, potem približek z normalno porazdelitvijo za primerjavo dveh Poissonovih slučajnih spremenljivk\* vodi do predstavljenega obrazca.

Za izvedbo testa zadošča že kalkulator, torej tudi mobilni telefon. Če želimo izvedbo združiti z zbiranjem podatkov in s prikazom rezultatov, potrebujemo le elektronsko preglednico, priljubljeni Microsoftov Excel ali kakšen brezplačen nadomestek zanj pa ima dandanes nameščen praktično vsak osebni računalnik. Uporaba statističnega programskega paketa torej ni potrebna.

Za boljše razumevanje in pravilno uporabo testa je potrebno razumeti še nekaj njegovih statističnih lastnosti. Informacija znotraj vsake skupine leži le v števcu razmerja, torej v številu udeležencev z dogodkom – imenovalca ni pomemben, torej je vseeno, ali smo isto število dogodkov opazili v poskusu z večjima skupinama. Po drugi strani je od števila dogodkov odvisna moč testa, ki je seveda večja pri večjem številu dogodkov. Poleg tega ima metoda dve kritični predpostavki oziroma omejitvi: če se imenovalca razmerij (t.j. velikosti skupin, ki ju primerjamo) nezamisljivo razlikujeta, je test seveda pristranski; in če je pogostnost dogodkov velika, postane test konzervativen (t.j. daje vrednosti  $p$ , ki so večje, kot bi morale biti). Za

\* Poissonova porazdelitev, znana tudi kot porazdelitev redkih dogodkov, je ena od osnovnih teoretičnih porazdelitev diskretnih slučajnih spremenljivk. Ker ima zanimivo zgodovino in je izredno široko uporabna, a za razumevanje ne zahteva izjemnega matematičnega znanja (5, 6, 7), jo priporočamo kot temo za dodatno branje.

večino študij ti omejitvi sicer nista bistveni, saj sta primerjani skupini praviloma enako veliki, klinični dogodki pa so praviloma redki, torej je test široko uporaben. Podrobno analizo matematičnih lastnosti različnih statističnih testov za primerjavo dveh Poissonovih razmerij najdemo v statistični literaturi (8).

Za konec uvoda je potrebno opozoriti, da je pri objavljanju rezultatov kliničnih poskusov v mednarodnih znanstvenih časopisih namesto predstavljenega »priročnega« postopka praviloma potrebno navesti rezultate »običajnih« statističnih testov (logrank, hi-kvadrat ipd.). Podobno je glede velikosti učinka zdravljenja: praviloma ga ocenimo z razmerjem tveganj ali razmerjem obetov, a dobro »oceno za občutek« nam nudi že preprosto razmerje udeležencev, ki jih je doletel dogodek, med opazovanima skupinama.

## METODE

Analizo podatkov smo izvedli na podlagi medicinske dokumentacije – zdravstvenih kartonov pacientov. Za leta 2001, 2004 in 2007 smo naključno izbrali po 20 kartonov iz vsakega od šestih obravnavanih oddelkov IRSR iz časovnega obdobja september-november. Pregledali smo sprejemne, odpustne in temperaturne liste.

Če sta bili zapisani etiološka in funkcijska diagnoza, smo to šteli za eno diagnozo. Taka primera sta paraplegija in zlom vretenca ter možganska kap in hemiplegija. Po drugi strani gre pri amputaciji spodnjega uda zaradi sladkorne bolezni za dve diagnozi, pri amputaciji spodnjega uda zaradi ateroskleroze pa za eno diagnozo, če ni bilo podatka o okvarah drugih organov, in dve diagnozi, če je bil npr. narejen žilni obvod na drugem spodnjem udju.

Analizirali smo tudi zaplete (reanimacija; aspiracija; vstavljen stalni urinski kateter; katetrizacija ali menjava katetra; novo nastala rana ali preležanina; nastavitev infuzije; transfuzija; nadzor bilance tekočin; motnje zavesti brez reanimacije; poslan na IPP zaradi zapleta; telesna temperatura povišana nad 37,9°C več kot 1 dan; globoka venska tromboza; padec brez poškodbe; padec s poškodbo; avtonomna hiperrefleksija; hipertenzivna reakcija; urejanje krvnega sladkorja več kot 1 dan; angina pectoris; akutni abdomen; izpad in vstavljanje nazogastrične sonde; akutna krvavitev iz notranjih organov; alergična reakcija; menjava kanile; bolečine z intramuskularno ali intravenozno terapijo; bolečine več kot 1 dan z dodatno peroralno terapijo; nemir z intramuskularno terapijo; nemir več kot 1 dan s peroralno terapijo; epileptični napad; težave z dihanjem; inhalacije; kisik; uroinfekt; respiratorni infekt; inhalacije; kisik; bruhanje; težave s spanjem več kot 1 dan z dodatno terapijo; drugo relevantno). Tu poročamo le o kvantitativni analizi skupnega števila zapletov, posamezne zaplete pa smo analizirali tudi kvalitativno oziroma so bili obravnavani na kliničnih sestankih.

Za zbiranje in analizo podatkov smo uporabili elektronsko preglednico Microsoft® Excel (vezija 2003 SP3; Microsoft Corporation, Redmond, WA).

## REZULTATI

Tabela 1 povzema zbrane podatke in ugotovitve glede pogostnosti diagnoz in zapletov v opazovanih letih. Vidimo, da je bilo število diagnoz na pacienta tako v letu 2004 kot v letu 2007 statistično značilno večje kot v letu 2001. Povečanje med letoma 2004 in 2007 ni statistično značilno. Tudi zabeleženo število zapletov na pacienta je bilo v letih 2004 in 2007 večje kot v letu 2001, a razlika ni statistično značilna.

**Tabela 1:** Primerjava pogostnosti diagnoz in zapletov na IRSR, ocenjene s pregledom medicinske dokumentacije, med leti 2001, 2004 in 2007

	2001		2004		2007	
	diagnoze	komplikacije	diagnoze	komplikacije	diagnoze	komplikacije
število na karton	2,5	1,0	3,1	1,3	3,4	1,2
primerjava			$z=2,704$	$z=1,793$	$z=4,197$	$z=1,040$
z letom 2001			$p=0,007$	$p=0,073$	$p<0,001$	$p=0,298$
primerjava					$z=1,502$	$z=-0,754$
z letom 2004					$p=0,133$	$p=0,451$

## RAZPRAVA

Za analizo medicinskih podatkov se upravičeno uporablja mnogo zahtevnih in zapletenih statističnih metod. Toda tudi brez njih se da izluščiti ključne informacije, na katerih temeljijo sklepi raziskav. V randomiziranih poskusih, kjer je izid dogodek, je ključno število udeležencev v vsaki opazovani skupini, ki so izkusili dogodek. Statistični test, ki smo ga predstavili, uporablja zgolj ta podatek in daje hiter, a zanesljiv vpogled v temeljno sporočilo poskusa. Test predpostavlja, da je število dogodkov v posamezni skupini slučajna spremenljivka, porazdeljena po Poissonovi porazdelitvi. Pod ničelno hipotezo se tedaj število dogodkov, ki od danega skupnega števila dogodkov odpade na vsako od skupin, porazdeljuje kot binomska slučajna spremenljivka z verjetnostjo posameznega dogodka 0,5. Na tem in na normalnem približku k binomski porazdelitvi temelji McNemarjev test razlike med odvisnima deležema (9), ki je opisan v večini osnovnih statističnih učbenikov. Bistveno manj znano pa je, da lahko iz iste osnove izpeljemo test, ki smo ga predstavili.

Test je uporaben tudi, če je imamo več možnih dogodkov oziroma alternativnih izidov, kot npr. več različnih neželenih dogodkov pri preizkušanju zdravil. Seveda se z večanjem števila testov, ki jih naredimo, povečuje verjetnost »lažno pozitivnega« sklepa, toda če imamo opravka s tremi ali

štirimi izidi, torej s tremi ali štirimi sočasnimi testi, in se odločamo na podlagi približnega pravila, da razlika po vsej verjetnosti ni slučajna, če smo dobili  $z$  večji od 2 (oziroma manjši od -2), lahko tak test nudi dragoceno opozorilo izvajalcem, za katere lahko potem statistiki izvedejo zahtevnejše analize. V našem primeru gre za istovrsten problem, saj smo leti 2004 in 2007 primerjali z letom 2001, nato pa še med seboj. Ker so primerjave le tri, ker so bile vnaprej načrtovane (ne pa izvedene *post-hoc* glede na rezultate drugih statističnih analiz), ker so bili testi zaradi pogostnosti dogodka konzervativni (kot je pojasnjeno v nadaljevanju) in ker je namen prej praktično klinično in organizacijsko odločanju kot znanstveno spoznavanje, se prav tako nismo odločili za Bonferronijev popravek oziroma se sploh nismo spuščali v zapleteno problematiko večkratnih testiranj, o kateri so tudi mnenja statistikov pogosto deljena (10).

Naši podatki od predpostavk uporabljenega testa odstopajo tudi glede pogostnosti dogodka. Tako pri diagnozah kot pri zapletih je razmerje na pacienta namreč večje od 1 (kot smo omenili, pa naj bi bilo po možnosti manjše od 0,2). Uporabljeni test je zato izrazito konzervativen (to pomeni, da je verjetnost napake 1. vrste – t.j. neupravičene zavrnitve ničelne hipoteze – manjša, kot kaže dobljena vrednost  $p$ ), a v našem primeru je to (več kot zadostna) protiutež povečani verjetnosti napake 1. vrste kot posledici večkratnih testiranj. Zato smo lahko toliko bolj prepričani v veljavnost sklepa o povečevanju zahtevnosti bolnišnične rehabilitacije.

Izognili smo se tudi problematiki statističnega sklepanja v nerandomiziranih študijah. V predstavitvi testa in v vsej s tem povezani statistični teoriji je govora le o randomiziranih poskusih, naši podatki pa so seveda opazovalne narave. A podobno kot zgoraj, bomo tudi tu zahtevno statistično (in tudi filozofsko oziroma epistemološko) problematiko, ki ostaja pereča od samih začetkov uporabne statistike do danes (11), pustili ob strani ter sprejeli uporabljeni test kot robustnega in priročnega za hiter vpogled v stanje.

## ZAKLJUČKI

Uporaba statističnega testa, ki ga je upravičeno imenovati »najpreprostejši«, nam je dala uporaben »občutek za podatke«. Ti kažejo na povečevanje zahtevnosti bolnišnične rehabilitacije na IRSR skozi zadnje desetletje.

## ZAHVALA

Zahvaljujem se sodelavki prof.dr. Heleni Burger, dr.med., ki je zasnovala in vodila pregled medicinske dokumentacije ter skrbno pregledala rokopis prispevka, ter sodelavcem asist. Metki Moharič, dr.med., dipl.fiziot., asist. Nataši

Puzić, dr.med., Meliti Marolt, dr.med., in Martinu Zorku, dr.med., ki so pregledali zdravstvene kartone in opravili vnos podatkov. Hvaležen sem tudi anonimnemu recenzentu za tehtne pripombe.

## Literatura:

1. Vidmar G, Burger H, Marinček Č. Time-trend in ability level of stroke and multiple sclerosis patients undergoing complex rehabilitation in Slovenia. V: Marinček Č, Burger H (ur.). 7th Mediterranean Congress of Physical and Rehabilitation Medicine, Porotorž, Slovenia, September 18-21, 2008: Book of Abstracts. Ljubljana: Institute for Rehabilitation, Republic of Slovenia, 2008: 110.
2. Kalčina L, Močnik-Drnovšek V. Pacientove pravice. Ljubljana: Informacijsko dokumentacijski center Sveta Evrope, 2005.
3. Pocock SJ. The simplest statistical test: how to check for a difference between treatments. *BMJ* 2006; 332: 1256-8.
4. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. (4. izd.). Oxford: Blackwell, 2002.
5. Bullmer, MG. Principles of statistics (2. izd.). New York: Dover, 1979.
6. Avtorji Wikipedije. Poisson distribution. Wikipedia, The Free Encyclopedia. Dostopno na: [http://en.wikipedia.org/w/index.php?title=Poisson\\_distribution&oldid=249579128](http://en.wikipedia.org/w/index.php?title=Poisson_distribution&oldid=249579128). Dostop 5. novembra 2008.
7. Cedilnik A. Uvod v verjetnostni račun. Ljubljana: FDV, 2003.
8. Ng HKT, Gua K, Tang ML. A comparative study of tests for the difference of two Poisson means. *Comp Stat Data Anal* 2007; 51: 3086-99.
9. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; 12: 153-7.
10. Howell D. Statistical methods for psychology (6. izd.). Belmont: Duxbury, 2007.
11. Reeves BC. Principles of research: limitations of non-randomized studies. *Surgery* 2006; 24: 263-7.