

Statističke pogreške u rukopisima zaprimljenim u časopis *Biochemia Medica*

Statistical errors in manuscripts submitted to *Biochemia Medica* journal

Ana-Maria Šimundić, Nora Nikolac

Klinički zavod za kemiju, Klinička bolnica „Sestre milosrdnice“, Zagreb

University Department of Chemistry, Sestre milosrdnice University Hospital, Zagreb, Croatia

Sažetak

Uvod: Neispravna uporaba statističkih metoda može biti uzrokom iskrivljenja rezultata, netočnih zaključaka i znatnog gubitka financijskih i drugih sredstava i kao takva se smatra vrlo neetičnom. Nažalost, velik broj objavljenih radova u medicinskim časopisima sadrži neke statističke pogreške. Cilj ovog rada bio je utvrditi učestalost nekih najčešćih statističkih pogrešaka, u člancima zaprimljenim u uredništvo časopisa *Biochemia Medica*, tijekom četverogodišnjeg perioda, u razdoblju od 2006. do 2009. godine.

Materijali i metode: Obuhvaćeni su svi izvorni znanstveni i stručni radovi, koji su sadržavali statističku analizu podataka, zaprimljeni u uredništvo časopisa *Biochemia Medica* u razdoblju od 2006. do 2009. godine. Rukopisi su pregledani od strane dva recenzenta. Sljedeće pogreške su uključene u analizu: 1) pogrešna uporaba ili prikaz deskriptivne analize; 2) pogrešan izbor statističkog testa; 3) pogrešna uporaba statističkog testa za usporedbu tri ili više skupina; 4) pogrešan prikaz P vrijednosti; 5) pogrešno tumačenje P vrijednosti; 6) pogrešno tumačenje rezultata korelacije; 7) nije provedena analiza snage istraživanja.

Rezultati: Ukupno je obuhvaćeno 55 rukopisa. Niti jedan od rukopisa nije imao provedenu analizu snage istraživanja. Od ostalih šest pogrešaka, barem jedna pogreška utvrđena je u 48/55 (0,87) rukopisa. Najučestalije pogreške su bile neispravna uporaba statističkog testa za usporedbu tri ili više skupina (21/28 (0,75) članaka) i neispravan prikaz P vrijednosti (36/54 (0,66) članaka).

Zaključak: Statističke pogreške su vrlo učestale u rukopisima zaprimljenim u uredništvo časopisa *Biochemia Medica*. Sve potencijalne pogreške ustanovljene tijekom recenzije od strane statističkog recenzenta bile su ispravljene prije objavljivanja članaka, što je značajno unaprijedilo njihovu kvalitetu.

Ključne riječi: biostatistika; pogreške; analiza podataka; istraživačka etika

Abstract

Introduction: The inappropriate use of statistical methods may lead to distorted results, incorrect conclusions, and substantial waste of financial and other resources and is considered highly unethical. Unfortunately, a great number of published medical research contain some statistical errors. The aim of this study was to assess the frequency of several most common statistical errors, in manuscripts submitted to *Biochemia Medica* journal for possible publication, during the 4 year period, since 2006-2009.

Materials and methods: All original scientific and professional manuscripts submitted to *Biochemia Medica* during the 2006-2009 were eligible for the study, if they contained some kind of statistical analysis of the data. Manuscripts were reviewed manually by two reviewers. Following errors were included: 1) incorrect use or presentation of descriptive analysis; 2) incorrect choice of the statistical test; 3) incorrect use of statistical test for comparing three or more groups for differences; 4) incorrect presentation of P value; 5) incorrect interpretation of P value; 6) incorrect interpretation of correlation analysis; 7) power analysis not provided.

Results: A total of 55 eligible manuscripts were identified. None of these manuscripts reported power analysis. As of other 6 errors analyzed, at least one error was observed in 48/55 (0.87) manuscripts. Most common errors were incorrect use of statistical test for comparing three or more groups for differences (21/28 (0.75)) and incorrect presentation of P value (36/54 (0.66)).

Conclusion: The statistical errors are highly prevalent in manuscripts submitted to *Biochemia Medica*. All errors identified by statistical reviewer during the review process were successfully corrected prior to publication, what has greatly improved the overall quality of the manuscript.

Key words: biostatistics; errors; data analysis; research ethics

Pristiglo: 19. kolovoza 2009.

Prihvaćeno: 4. rujna 2009.

Received: August 19, 2009

Accepted: September 4, 2009

Uvod

Biostatistika je danas prihvaćena kao vrlo moćan alat za razumijevanje podataka i njihovu analizu i statističke metode se sve više koriste u biomedicinskoj znanstvenoj literaturi.

Introduction

Biostatistics is widely accepted as a powerful tool for understanding and analyzing data and there is a constant increase in the use of statistical methods in scientific biomedical literature. The inappropriate use of statistical

Neispravna uporaba statističkih metoda je ozbiljan problem koji može biti uzrokom iskrivljenja rezultata, netočnih zaključaka i znatnog gubitka financijskih i drugih sredstava (1). Nadalje, neispravna uporaba statističkih metoda može imati ozbiljne kliničke posljedice i kao takva se smatra vrlo neetičnom (2,3). Pogreške se događaju zbog manjka znanja ili slučajnog propusta, te čak zbog nemara i hotimične prijave (4). Nažalost, velik broj objavljenih radova u medicinskim časopisima sadrži neke statističke pogreške (5-8).

Kako bi spriječili nastanak pogrešaka ili barem smanjili njihov broj i unaprijedili kvalitetu članaka, mnogi su znanstveni časopisi uveli postupak statističke recenzije ili čak statističkog urednika (9,10). Nadalje, s istom su svrhom objavljene i mnoge smjernice koje se odnose na znanstveno-istraživačku metodologiju, ustroj istraživanja, analizu podataka i pisanje rada. No, napredak koji je uslijedio je bio vrlo skroman i neki su značajniji problemi u istraživačkoj metodologiji i statističkoj analizi još uvijek prisutni u biomedicinskoj literaturi.

Biochemia Medica je časopis s međunarodnom recenzijom, iz područja kliničke kemije, koji izlazi od 1991. godine. Godine 2006. značajno je promijenjena struktura uredničkog odbora i politika časopisa (11), ubrzo nakon čega je časopis uvršten u neke značajnije bibliografske baze s područja biomedicine (12,13), što govori o neprekidnom poboljšanju kvalitete sadržaja časopisa (14). Uredništvo časopisa od 2006. godine uvodi statističku recenziju za sve zaprimljene članke. S ciljem podučavanja i pomoći potencijalnim autorima u razumijevanju nekih osnovnih statističkih pojmova, časopis također 2006. godine uvodi rubriku *Odabrane teme iz biostatistike*, u kojoj su do sada objavljeni mnogi edukacijski članci (15-24).

Cilj ovog rada bio je utvrditi učestalost nekih najčešćih statističkih pogrešaka, u člancima zaprimljenih u uredništvo časopisa *Biochemia Medica*, tijekom četverogodišnjeg perioda, u razdoblju od 2006. do 2009. godine.

Treba istaknuti kako su svi autori bili obaviješteni o pogreškama ustanovljenim tijekom statističke recenzije, te da su svi rukopisi bili temeljito prepravljani i gotovo sve pogreške ispravljene prije objavljivanja rada.

Materijali i metode

U analizu smo uključili sve izvorne znanstvene i stručne radove, koji su sadržavali statističku analizu podataka, zaprimljene u uredništvo časopisa *Biochemia Medica* u razdoblju od 2006. do 2009. godine. Rukopisi su pregledani od strane dva recenzenta. U analizu su uključene sljedeće pogreške: 1) pogrešna uporaba ili prikaz deskriptivne analize; 2) pogrešan izbor statističkog testa; 3) pogrešna uporaba statističkog testa za usporedbu tri ili više skupina; 4) pogrešan prikaz P vrijednosti; 5) pogrešno tumačenje P vrijednosti; 6) pogrešno tumačenje rezultata korelacije; 7)

methods is a serious problem and may lead to distorted results, incorrect conclusions and a substantial waste of financial and other resources (1). Moreover, it can have serious clinical consequences and is therefore considered as highly unethical (2,3). Errors may occur due to the lack of competence, honest error or even negligence and deliberate deception (4). Unfortunately, a great number of published medical research contains some statistical errors (5-8).

To prevent or at least reduce the error rate and to improve the quality of their articles, many scientific journals have introduced a statistical peer-reviewing process, or even statistical editor (9,10). Furthermore, many guidelines concerning research methodology, study design, data analysis and reporting, have been published. However, the improvement has been only modest and some major problems with research methodology and statistical analysis still exist in the biomedical literature.

Biochemia Medica is a peer reviewed clinical chemistry journal published since 1991. Its editorial board and policy has gone through a major change in 2006 (11) and journal was soon accepted for indexing in some major biomedical bibliographic databases (12,13), pointing to the constant improvement of the quality of the journal content (14). Since 2006 journal has implemented a statistical review for all manuscripts submitted to the journal for possible publication. To educate and help its readers and potential authors in understanding some basic statistical concepts, journal has in 2006 launched a *Lessons in biostatistics* section, under which many educational articles have been published so far (15-24).

The aim of this study was to assess the frequency of several most common statistical errors, in manuscripts submitted to *Biochemia Medica* for possible publication, during the 4 year period, since 2006-2009.

It should be mentioned here that authors were informed about all errors identified by the statistical review, manuscripts were thoroughly revised and almost all errors were corrected prior to publication.

Materials and methods

All original scientific and professional manuscripts submitted to *Biochemia Medica* during the 2006-2009 were eligible for the study, if they contained some kind of statistical analysis of the data. Manuscripts were reviewed manually by two reviewers. Following errors were included: 1) incorrect use or presentation of descriptive analysis; 2) incorrect choice of the statistical test; 3) incorrect use of statistical test for comparing three or more groups for differences; 4) incorrect presentation of P value; 5) incorrect interpretation of P value; 6) incorrect interpretation of correlation analysis; 7) power analysis not provided. Errors under #2 did not include incorrect use of sta-

nije provedena analiza snage istraživanja. Pogreške pod točkom broj 2. nisu uključivale pogrešnu uporabu statističkog testa za usporedbu tri ili više skupina.

Statistička analiza

Podaci su prikazani opisno, kao broj i udio članaka u pojedinoj kategoriji. Statistička analiza je napravljena pomoću statističkog programa MedCalc[®] (MedCalc 9.3.0.0, Frank Schoonjans, Mariakerke, Belgium).

Rezultati

Ovim je istraživanjem ukupno obuhvaćeno 55 rukopisa zaprimljenih u uredništvo časopisa *Biochemia Medica*, od kojih je kasnije 18 odbijeno, a 37 prihvaćeno za objavljivanje. Niti jedan od rukopisa nije imao provedenu analizu snage istraživanja. Od ostalih šest pogrešaka, barem jedna pogreška utvrđena je u 48/55 (0,87) rukopisa. Od rukopisa koje smo obuhvatili ovom analizom, samo 7/55 (0,13) nije sadržavalo niti jednu pogrešku, osim analize snage istraživanja. Učestalost statističkih pogrešaka, u člancima zaprimljenim u uredništvo časopisa *Biochemia Medica*, tijekom perioda obuhvaćenog ovim istraživanjem, prikazana je u Tablici 1.

Najučestalije pogreške su bile neispravna uporaba statističkog testa za usporedbu tri ili više skupina i neispravan prikaz P vrijednosti.

Pogrešan izbor statističkog testa bio je prisutan u 34/55 analiziranih rukopisa. Najčešće su autori griješili: (i) koristeći parametrijski test (npr. t-test) umjesto neparametrijskog testa (npr. Mann-Whitney Rank Sum test), u slučaju kad su imali mali uzorak ili kad raspodjela nije bila normalna; (ii) koristeći neparni umjesto parnog testa za zavisne uzorke; (iii) koristeći χ^2 test umjesto Fisherovog egzaktnog testa kad su frekvencije u pojedinim razredima male.

TABLICA 1. Učestalost statističkih pogrešaka u člancima zaprimljenih u uredništvo časopisa *Biochemia Medica*, u periodu od 2006. do 2009. Pogreške su poredane prema učestalosti u padajući niz.

tistical test for comparing three or more groups for differences.

Statistical analysis

Descriptive data were presented as number and proportions of articles within various categories. Statistical analysis was done using MedCalc[®] statistical software (MedCalc 9.3.0.0, Frank Schoonjans, Mariakerke, Belgium).

Results

We identified a total of 55 eligible manuscripts submitted to *Biochemia Medica*, out of which 18 were subsequently rejected and 37 accepted for publication. None of the eligible manuscripts reported power analysis. As of other 6 errors analyzed, at least one error was observed in 48/55 (0.87) manuscripts. Only 7/55 (0.13) manuscripts had none of the errors analyzed in this study, with the exception of power analysis. The frequency of statistical errors in manuscripts submitted to *Biochemia Medica* during the studied period is presented in Table 1. Most common errors were incorrect use of statistical test for comparing three or more groups for differences and incorrect presentation of P value.

Incorrect choice of the statistical test was present in 34/55 studied manuscripts. Most commonly authors failed to: (i) use non-parametric test like Mann-Whitney Rank Sum test when sample is too small, or when data are not normally distributed instead of using parametric statistical methods (i.e. t-test); (ii) use paired tests when samples are dependent (paired); (iii) use Fisher exact test instead of testing data with chi-square test, due to the low cell frequencies.

As of the incorrect use or presentation of descriptive analysis, errors most often done by authors were: (i) usi-

TABLICA 1. The frequency of statistical errors in manuscripts submitted to *Biochemia Medica* during 2006-2009. Errors are sorted according to their frequency.

Error	Error rate N (proportion)
Power analysis not provided	55/55 (1.0)
Incorrect use of statistical test for comparing three or more groups for differences	21/28 (0.75)
Incorrect presentation of P value	36/54 (0.66)
Incorrect choice of the statistical test	34/55 (0.62)
Incorrect interpretation of correlation analysis	11/20 (0.55)
Incorrect use or presentation of descriptive analysis	19/55 (0.35)
Incorrect interpretation of P value	12/54 (0.22)

Što se tiče pogrešne uporabe ili prikaza opisne analize, autori su najčešće grijehili: (i) koristeći aritmetičku sredinu i standardnu devijaciju (umjesto medijana i interkvartilnog raspona) za opis skupina čiji podaci nisu normalno raspodijeljeni ili za male skupine; (ii) navodeći standardnu pogrešku aritmetičke sredine (engl. *standard error of the mean*, SEM) kao mjeru varijabilnosti podataka.

Rasprava

Neispravna uporaba ili čak zlouporaba biostatističkih metoda je dobro znana i obimno raspravljana u literaturi. Neispravan ustroj istraživanja i propusti u statističkoj analizi neizbježno vode k pogrešnim zaključcima i takvi nedostaci mogu dovesti do loše kliničke prakse i gubitka vrijednih financijskih sredstava u zdravstvu. Cilj ovog rada bio je utvrditi učestalost nekih najčešćih statističkih pogrešaka, u člancima zaprimljenim u uredništvo časopisa *Biochemia Medica*, u periodu od 2006. do 2009. godine. Naša je analiza pokazala kako znatan udio rukopisa zaprimljenih u uredništvo časopisa ima barem neku statističku pogrešku. Analiza snage istraživanja nije provedena niti u jednom radu, a najučestalije pogreške su bile uporaba pogrešnog statističkog testa za usporedbu tri ili više skupina i neispravan prikaz P vrijednosti.

Ispravno ustrojena istraživanja trebaju imati odgovarajuću snagu kako bi bila u stanju ustanoviti znatnije odstupanje od nulte hipoteze. Analiza snage istraživanja se koristi kako bi se odredila najmanja potrebna veličina uzorka za neku analizu. Općenito govoreći, što je veći uzorak, manja je pogreška uzorkovanja. Pogreška uzorkovanja je jedna od najučestalijih pogrešaka u znanstvenoj biomedicinskoj literaturi. Ispravno i primjereno uzorkovanje je ključni preduvjet za valjanost istraživanja. Ukoliko ispitivane skupine nisu reprezentativne za neku populaciju, autori ne bi smjeli koristiti inferencijsku statistiku, tj. donositi zaključke o populaciji (20,25). Budući je u svim člancima, obuhvaćenim ovom analizom, nedostajala analiza snage istraživanja, čini se da s tim u vezi ima prostora za poboljšanje. Autori bi trebali biti svjesni te činjenice i pokušati izbjeći takvu pogrešku. Prilikom planiranja istraživanja, autori ne bi smjeli zaboraviti izračunati odgovarajuću veličinu uzorka kako ne bi načinili pogrešku tipa II. Čitateljima valja dati do znanja je li veličina uzorka bila dovoljna kako bi se otkrio očekivani učinak od interesa. Ovdje valja naglasiti kako analiza snage istraživanja nije potrebna za neka jednostavnija preliminarina istraživanja.

Pogrešan izbor statističkog testa je također vrlo česta statistička pogreška (26,27). Autori često nisu svjesni uvjeta koje valja zadovoljiti pri izboru nekog statističkog testa. Ti se uvjeti odnose na značajke raspodjele, mjernu ljestvicu, veličinu uzorka, broj skupina i opažanja, te neke druge. Znatan udio članaka zaprimljenih u uredništvo časopisa *Biochemia Medica* ima upravo tu vrstu pogreške. Autori bi uvijek, prije izvođenja bilo kojeg statističkog testa, trebali

ng mean and standard deviation (instead of median and IQR, interquartile range) to describe data sets with non-Gaussian distribution or when samples are small; and (ii) using standard error of the mean (SEM) as a measure of data variability.

Discussion

The phenomenon of inappropriate use and misuse of biostatistical methods is well known and extensively discussed in the literature. Incorrect design of the study and statistical flaws inevitably lead to incorrect conclusions and such shortcomings may be misleading to the clinical practice and waste of the valuable health resources. The aim of this work was to assess the frequency of some most common statistical errors in manuscripts submitted to *Biochemia Medica* during 2006-2009. Our analysis revealed that a substantial proportion of manuscripts involved at least some statistical error. Power analysis was never provided and most common errors were incorrect use of statistical test and incorrect presentation of P value.

Properly designed experiments should ensure adequate power to detect reasonable departures from the null hypothesis. Power analysis is usually used to calculate the minimum sample size required to some specific analysis. In general, the larger the sample size, the smaller is the sampling error. Sampling error is one of the most commonly occurring mistakes in the scientific biomedical literature. Correct and appropriate sampling procedure is a key prerequisite to the validity of the study. If study groups are not representative for the population, authors should not use inferential statistical techniques, i.e. to make conclusions about the population (20,25). Since all eligible manuscripts in our study were lacking power analysis, there seems to be a lot of space for potential improvement, regarding this issue. Authors should be aware of that fact and try to avoid this mistake. When designing the study authors should always remember to calculate the adequate sample size in order not to do the type II error. Readers should be informed whether the size of the sample was appropriate to detect the effect of interest. It should be noted that power analysis is not necessary for simple exploratory studies reporting preliminary results. Failure to choose the proper statistical test is also a common statistical mistake (26,27). Authors are often unaware of the assumptions that need to be met when applying some statistical test for data analysis. These assumptions are concerning the data distribution characteristics, scale of measurement, sample size, number of groups and observations and some other. A substantial proportion of manuscripts submitted to *Biochemia Medica* suffer from this error. Authors are advised to always check for the assumptions of the test, prior to any analysis and to inform the readers why they chose the test, by stating for exam-

provjeriti jesu li zadovoljeni uvjeti za njegovo izvođenje i izvijestiti čitatelje zašto su odabrali upravo taj test, kao u sljedećem primjeru: „Za usporedbu više nezavisnih skupina primijenjen je ANOVA test, odnosno neparametrijski Kruskal-Wallis test za raspodjele koje nisu bile normalne.“ (28).

Kako proizlazi iz naših rezultata, P vrijednost nije bila ispravno prikazana u dvije trećine rukopisa, dok u više od 1/5 rukopisa P vrijednost nije bila ispravno tumačena. Takvo je opažanje u skladu s već objavljenim u znanstvenoj literaturi i ta je činjenica opsežno raspravljena u mnogim radovima (29-31).

Autori trebaju voditi računa o tome da uvijek valja navesti točnu P vrijednost za sve ispitane razlike i zaokružiti broj na tri decimalna mjesta (npr. treba napisati $P = 0,048$, umjesto $P < 0,05$). Valja izbjegavati izraze poput $P = NS$ (eng. *non significant*, nije značajno), $P > 0,05$ i $P < 0,05$. Također, P vrijednost ne treba navoditi s previše decimalnih mjesta, kao npr $P < 0,00001$, ukoliko ne postoje neki posebni razlozi za to. Nadalje, autori često izvode netočne zaključke temeljene isključivo na P vrijednosti, zanemarujući kliničku značajnost rezultata. Kako veličina opaženog učinka ne proizlazi iz saznanja o P vrijednosti, uvijek valja navoditi apsolutne razlike među skupinama i pripadajuće intervale pouzdanosti. Naime, male razlike mogu biti statistički značajne, a klinički potpuno besmislene, ako je uzorak veliki, dok velike razlike mogu biti klinički, no ne i statistički značajne, ukoliko je uzorak mali. Rezultate valja stoga uvijek tumačiti imajući na umu njihovu kliničku značajnost, kao što su npr. Pašalić i sur. u svom radu naveli: „Iako statistički značajne, razlike koncentracije kolesterola ($P = 0,001$), HDL kolesterola ($P < 0,001$), apolipoproteina A-I ($P < 0,001$) i triglicerida ($P = 0,002$) u plazmi između bolesnika s normalnom tjelesnom nasom, prekomjernom tjelesnom masom i onih pretilih bile su klinički nevažne.“ (32).

Još jedna pogreška, učestalo prisutna u rukopisima zaprimljenim u uredništvo časopisa *Biochemia Medica*, je pogrešan prikaz rezultata. Netočna uporaba ili prikaz opisne analize je bila zastupljena u više od jedne trećine rukopisa obuhvaćenih ovim istraživanjem. O učestalosti pogrešnog navođenja standardne pogreške aritmetičke sredine (SEM) pisali su i drugi autori. Nagele P. i suradnici su u svom radu ispitali učestalost pogrešne upotrebe SEM u četiri vodeća časopisa s područja anestezije tijekom 2001. godine. Autori tog rada su ustanovili kako je u svakom četvrtom članku (198/860, 23%) objavljenom u jednom od četiri analizirana časopisa tijekom 2001. godine, SEM pogrešno navedena kao mjera varijabilnosti podataka nekog uzorka (33).

Naposlijetku, pogreške vezane uz analizu korelacije su također vrlo učestale u znanstvenim člancima (7). Najčešće nalazimo sljedeće pogreške: (i) korištenje parametrijskog Pearsonovog testa korelacije iako nisu zadovoljeni uvjeti za njegovo izvođenje; (ii) pogrešno tumačenje rezultata analize korelacije (koeficijenta korelacije - r , i njegove statističke značajnosti - P); (iii) ucrtani korelacijski pravac

ple: „... ANOVA test was used for comparison of multiple independent groups and non-parametric Kruskal-Wallis test for distributions that were not normal.“ (28).

As of our results, P value was not presented correctly in two thirds of the submitted manuscripts and not interpreted correctly in more than 1/5 manuscripts. This observation is consistent with what has already been published in scientific literature and this issue has been extensively reviewed (29-31). Authors should have in mind that exact P values should be given for all tested differences and rounded on three decimal places (for example $P = 0.048$ should be written instead of $P < 0.05$). Expressions like $P = NS$ (non significant), $P > 0.05$, $P < 0.05$ should not be used. Also, P should not be reported with too many decimal places, like for example $P < 0.00001$, unless there are some exceptional reasons for that. Furthermore, authors often make incorrect conclusions based solely on the P value, disregarding the clinical significance of the results. Absolute differences between groups and their respective confidence intervals should be reported whenever possible, as the magnitude of an effect is not suggested by a P-value. Small differences can be statistically significant, but meaningless, if your sample is too large, whereas large differences can be clinically meaningful, but statistically insignificant, if your sample is too small. Results obtained should be always interpreted relevant to their clinical significance, as was for example, reported by Pasalic D. et al.: „... Though statistically significant, differences in plasma cholesterol ($P = 0.001$), HDL-cholesterol ($P < 0.001$), apolipoprotein A ($P < 0.001$) and triacylglycerol ($P = 0.002$) concentrations between normal, overweight and obese patients were clinically irrelevant.“ (32).

Another common mistake, quite often observed in manuscripts submitted to *Biochemia Medica*, is the erroneous data presentation. Incorrect use or presentation of descriptive analysis was present in more than one third of the manuscripts assessed in this study. The misuse of SEM was reported by some other as well. Nagele P. and his colleagues have aimed to evaluate the frequency of inappropriate use of the SEM in four leading anaesthesia journals in 2001 and have found that one in four articles (198/860, 23%) published in four anaesthesia journals in 2001 inappropriately used the SEM in descriptive statistics to describe the variability of the study sample (33).

Lastly, the errors in correlation analysis are also very common in the scientific articles (7). Most commonly, following mistakes are made: (i) parametric Pearson test is performed even though test assumptions are not met for the test; (ii) the results (correlation coefficient - r , and its statistical significance P) of the correlation analysis are misinterpreted; (iii) fitted line is extrapolated outside the data set; and (iv) conclusions on causality are made. Conclusions on causality of the observed relationship are not

je ekstrapoliran izvan skupa podataka; i (iv) donose se zaključci o uzročno-posljedičnom odnosu. Zaključivati o uzročno-posljedičnom odnosu nije dozvoljeno jer korelacija predstavlja samo povezanost dvaju obilježja, a ne i ovisnost jednog o drugom (16,34).

Više od polovice rukopisa analiziranih u našem radu je imalo neki problem vezan uz analizu korelacije. Autori trebaju znati da uvijek valja provjeriti jesu li zadovoljeni uvjeti za izvođenje Pearsonovog testa korelacije: tj. da su oba obilježja brojana, da barem jedno obilježje slijedi normalnu raspodjelu, da je uzorak dovoljno velik i da postoji dokaz o linearnom odnosu (prethodno provjeriti grafičkim prikazom i analizom tzv. *reziduala*) (35). Ukoliko navedeni uvjeti nisu zadovoljeni, valja upotrijebiti Spearmanov test korelacije. Kuo je sa suradnicima (7) proveo analizu učestalosti problema ekstrapolacije ucrtanog pravca u 4 časopisa iz područja opće medicine s visokim čimbenikom odjeka (*British Medical Journal, Lancet, JAMA, New England Journal of Medicine*). Radom je obuhvaćeno 37 članaka u kojima su autori ucrtali pravac na graf koji prikazuje odnos dvaju obilježja (engl. *scatter graf*). Ustanovljeno je da je 22/37 (59%) članaka imalo barem neki problem s ekstrapolacijom, 4/37 (11%) je imalo ucrtani pravac koji je zadirao u područje besmislenih vrijednosti, dok su u 3/37 (8%) članaka čak donošeni zaključci o vrijednostima izvan područja analiziranog skupa podataka.

Ovim radom nismo imali namjeru načiniti cjelovitu analizu svih statističkih pogrešaka koje se pojavljuju u radovima zaprimljenim u časopis *Biochemia Medica*. Izbor pogrešaka koje smo analizirali je bio proizvoljan i temeljen isključivo na našem iskustvu i spoznajama o pojedinim osobitostima ispitivanih rukopisa. Na taj smo način možda propustili obuhvatiti neke pogreške, kao što su problem višestrukog testiranja, propusti u grafičkom prikazu podataka, pitanje podataka koji nedostaju, problemi vezani uz linearnu regresiju te problem vrijednosti koje su znatno izvan skupa podataka (engl. *outliers*) i neke druge. Glavni cilj ovog rada je u prvom redu bio upozoriti čitatelje te sve potencijalne autore Časopisa na neke osnovne pogreške i propuste, ukazujući tako na prostor za poboljšanje.

Zaključno, pogreške u uporabi biostatističkih metoda su vrlo učestale u rukopisima zaprimljenim u uredništvo časopisa *Biochemia Medica*. Sve potencijalne pogreške ustanovljene tijekom recenzije od strane statističkog recenzenta bile su ispravljene prije objavljivanja članka, što je značajno unaprijedilo njegovu kvalitetu. I autori i urednici bi trebali biti svjesni važnosti pravilne i nepravilne primjene statističkih metoda. Medicinski biokemičari i znanstvenici drugih grana biomedicine uključeni u bilo koju vrstu statističke analize podataka, morali bi provoditi svoja istraživanja profesionalno, kompetentno i etično (2). Svaki korak naprijed u smislu edukacije budućih potencijalnih autora je na dobrobit etičnosti, profesionalnosti i kompetentnosti postupaka istraživanja te kvalitete objavljenih radova.

allowed because correlation only represents the association between variables, but not causation (16,34).

More than half of the studied articles in our analysis were having some problem with correlation analysis. Authors should be sure to check the assumptions for the Pearson correlation analysis: i.e. that both variables are numeric, at least one variable is normally distributed, sample is large and there is evidence for linear correlation (as observed from a scatter plot, or by plotting residuals) (35). If those assumptions are not met, Spearman correlation analysis should be performed. Kuo (7) has studied the extrapolation problem in 4 general medical journals with high impact factor (*British Medical Journal, Lancet, JAMA, New England Journal of Medicine*). He reviewed a total of 37 articles with scatter plot and has observed that 22/37 (59%) of the published articles had some extrapolation problem, 4/37 (11%) had fitted line reaching meaningless value and 3/37 (8%) stated conclusions about the values outside the range of observed data.

This study was not aimed to be a comprehensive analysis of all potential statistical errors occurring in the manuscripts submitted to *Biochemia Medica*. The choice of errors to be analyzed was arbitrary and solely based on our experience and knowledge about the specificities in the studied manuscripts. As such, we may have missed some errors not included in this study, such as multiple hypothesis testing, erroneous graphical data presentation, missing data, issues in linear regression analysis, outliers data issues and some other. The main goal of this article was primarily to point out some basic errors and pitfalls to the readers of the Journal and all potential authors, thus identifying potential for improvement.

Herein we conclude that the misuse of biostatistics is highly prevalent in manuscripts submitted to *Biochemia Medica*. All potential errors identified by statistical reviewer during the review process were successfully corrected prior to publication, what has greatly improved the overall quality of the manuscript. Both, authors and editors should be aware of the importance of uses and misuses of statistical analysis. Clinical chemists and all other scientists in biomedicine involved in any kind of statistical analysis of data should perform their work in a professional, competent, and ethical manner (2). Each step forward in educating potential future authors is beneficial to the level of the ethical, professional and competent research practice as well as to the quality of the papers to be submitted.

Zahvala

Autori zahvaljuju Marijani Miler za korisne sugestije na završnu verziju rukopisa.

Adresa za dopisivanje:

Ana-Maria Šimundić
Klinički zavod za kemiju
Klinička bolnica „Sestre milosrdnice“
Vinogradska 29
10000 Zagreb
e-pošta: am.simundic@gmail.com

Acknowledgments

The authors wish to thank to Marijana Miler for her valuable comments to the final version of the manuscript.

Corresponding author:

Ana-Maria Simundić
University Department of Chemistry
Sestre milosrdnice University Hospital
Vinogradska 29
10000 Zagreb
Croatia
e-mail: am.simundic@gmail.com

Literatura/References:

1. Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research - a review of common pitfalls. *Swiss Med Wkly* 2007;137:44-9.
2. Committee on Professional Ethics. Ethical guidelines for statistical practice. Available at: <http://www.amstat.org/about/ethicalguidelines.cfm>. Accessed June 2nd, 2009.
3. Bilic-Zulle L. Scientific integrity - the basis of existence and development of science. *Biochem Med* 2007;17:143-50.
4. Gardenier JS, Resnik DB. The misuse of statistics: concepts, tools, and a research agenda. *Account Res* 2002;9:65-74.
5. McKinney WP, Young MJ, Hartz A, Lee MB. The inexact use of Fisher's Exact Test in six major medical journals. *JAMA* 1989;261:3430-3.
6. Kanter MH, Taylor JR. Accuracy of statistical methods in Transfusion: a review of articles from July/August 1992 through June 1993. *Transfusion* 1994;34:697-701.
7. Kuo YH. Extrapolation of correlation between 2 variables in 4 general medical journals. *JAMA* 2002;287:2815-7.
8. Simundić AM, Nikolac N, Topic E. Methodological issues in genetic association studies of inherited thrombophilia. *Clin Appl Thromb Hemost.* 2009;15:327-33.
9. Lukic IK, Marusic M. Appointment of statistical editor and quality of statistics in a small medical journal. *Croat Med J* 2001;42:500-3.
10. Petrovecki M. The role of statistical reviewer in biomedical scientific journal. *Biochem Med* 2009;19:223-30.
11. Topic E, Cvoriscec D. *Biochemia Medica* in the new guise. *Biochem Med* 2006;16:3-4.
12. Simundić AM. *Biochemia Medica* indexed. *Biochem Med* 2006;16:91-2.
13. Simundić AM, Topic E, Cvoriscec D. *Biochemia Medica* indexed in Science Citation Index Expanded and Journal Citation Reports/Science Edition citation databases. *Biochem Med* 2008;18:141-2.
14. Gasparac P. The role and relevance of bibliographic citation databases. *Biochem Med* 2006;16:93-102.
15. Bossuyt PMM. Clinical evaluation of medical tests: still a long road to go. *Biochem Med* 2006;16:103-6.
16. Udovicic M, Bazdaric K, Bilic-Zulle L, Petrovecki M. What we need to know when calculating the coefficient of correlation? *Biochem Med* 2007;17:10-5.
17. Raslich MA, Markert RJ, Stutes SA. Selecting and interpreting diagnostic tests. *Biochem Med* 2007;17:151-61.
18. McHugh ML. Standard error: meaning and interpretation. *Biochem Med* 2008;18:7-13.
19. Simundić AM. Confidence interval. *Biochem Med* 2008;18:154-61.
20. McHugh ML. Power analysis in research. *Biochem Med* 2008;18:263-74.
21. Ilakovac V. Statistical hypothesis testing and some pitfalls. *Biochem Med* 2009;19:10-6.
22. McHugh ML. The odds ratio: calculation, usage, and interpretation. *Biochem Med* 2009;19:120-6.
23. Bartolucci AA. Describing and interpreting the methodological and statistical techniques in meta-analyses. *Biochem Med* 2009;19:127-36.
24. McHugh ML. Risk reduction statistics. *Biochem Med* 2009;19:231-5.
25. Simundić AM. Types of variables and distributions. *Acta Med Croatica* 2006;60(Suppl 1):17-35.
26. Neville JA, Lang W, Fleischer AB Jr. Errors in the Archives of Dermatology and the Journal of the American Academy of Dermatology from January through December 2003. *Arch Dermatol* 2006;142:737-40.
27. Kotur PF. Statistics in biomedical journals. *Indian J Anaesth* 2006;50:166-8.
28. Dodig S, Galez D, Zoricic-Letoja I, Kristic-Kirin B, Kovac K, Nogalo B, et al. C-reactive protein and complement components' C3 and C4 in children with latent tuberculosis infection. *Biochem Med* 2008;18:52-8.
29. Ludwig DA. Use and misuse of p-values in designed and observational studies: guide for researchers and reviewers. *Aviat Space Environ Med* 2005;76:675-80.
30. Berle D, Starcevic V. Inconsistencies between reported test statistics and p-values in two psychiatry journals. *Int J Methods Psychiatr Res* 2007;16:202-7.
31. Chinn S. Statistics for the European Respiratory Journal. *Eur Respir J* 2001;18:393-401.
32. Pasalic D, Ferencak G, Grskovic B, Stavljenic-Rukavina A. Body mass index in patients with positive or suspected coronary artery disease: a large Croatian cohort. *Biochem Med* 2008;18:321-30.
33. Nagele P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br J Anaesth* 2003;90:514-6.
34. McClure P. Correlation statistics: review of the basics and some common pitfalls. *J Hand Ther* 2005;18:378-80.
35. Dawson B, Trapp RG. *Basic and Clinical Biostatistics*. 4th Ed. New York: Lange Medical Books/McGraw-Hill; 2004.