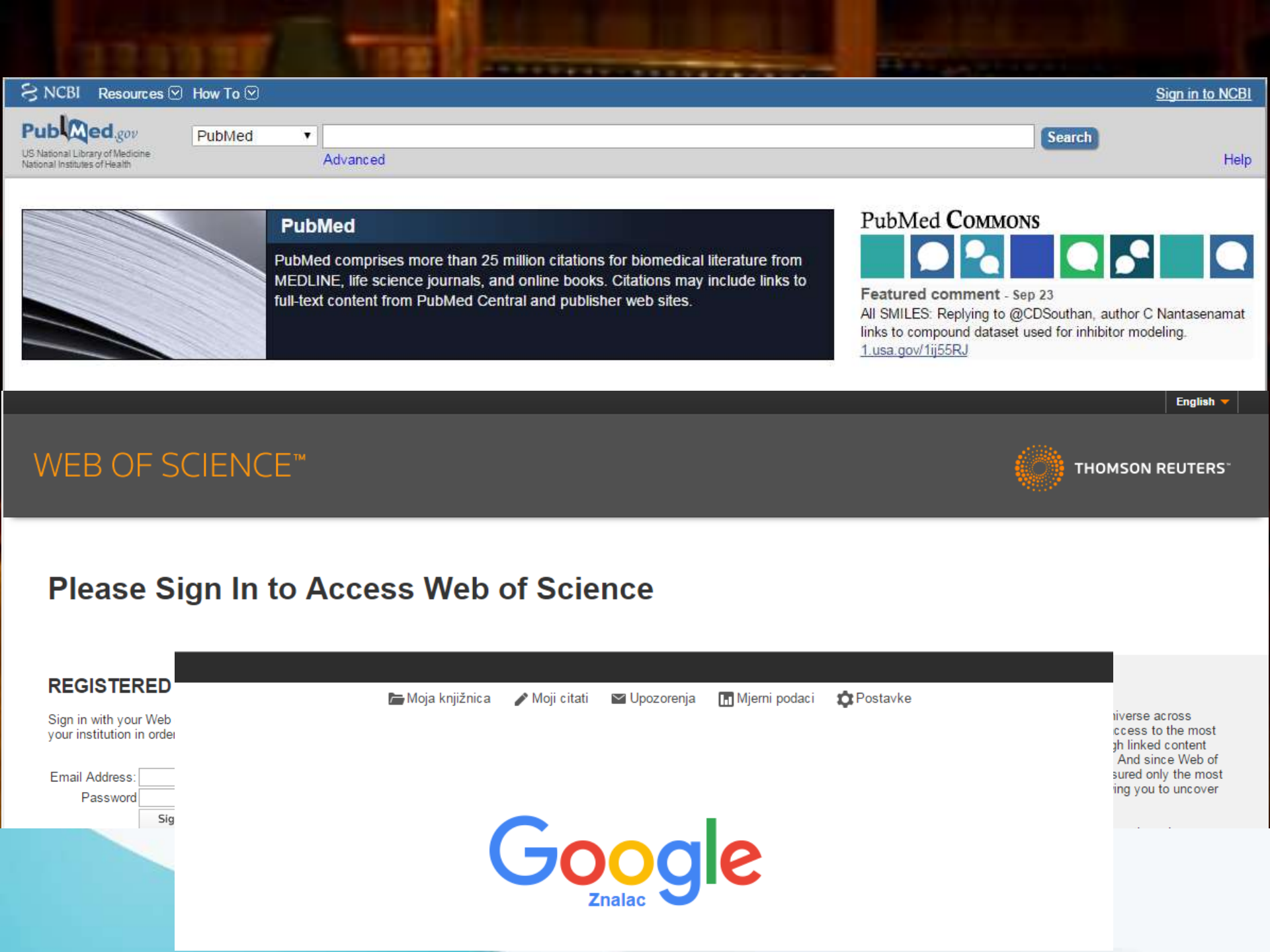
Najčešće pogrješke u statističkoj obradbi podataka (i kako ih izbjeći)

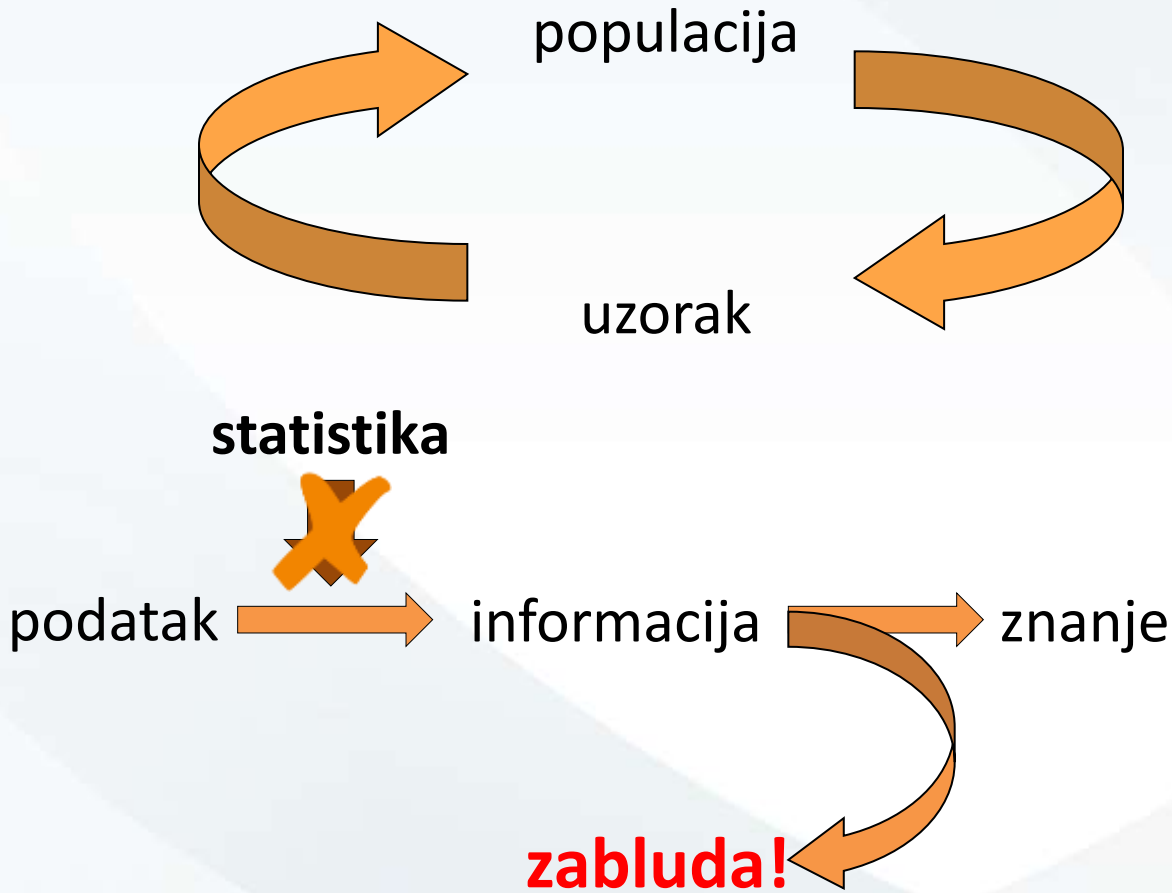
Lidija Bilić-Zulle

Klinički zavod za laboratorijsku dijagnostiku Kliničkog bolničkog centra Rijeka

Katedra za medicnsku informatiku Medicnskog fakulteta Sveučilišta u Rijeci


NCBI Resources How To
Sign in to NCBI
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Search
Advanced
Help
PubMed
PubMed comprises more than 25 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.
PubMed Commons
Featured comment - Sep 23
All SMILES: Replying to @CDSouthan, author C Nantasenamat links to compound dataset used for inhibitor modeling.
1.usa.gov/1ij55RJ
English
WEB OF SCIENCE™
THOMSON REUTERS
Please Sign In to Access Web of Science
REGISTERED
Sign in with your Web
your institution in orde
Email Address:
Password
Sig
Moja knjižnica
Moji citati
Upozorenja
Mjerni podaci
Postavke
iverse across
ccess to the most
gh linked content
And since Web of
sured only the most
ing you to uncover
Google
Znalac

Zašto je važno NE griješiti



Najčešće statističke pogrješke...

- ... u oblikovanju uzorka
- ... u odabiru mjernih varijabli
- ... u unosu i razvrstavanju podatka
- ... u prikazu podataka
- ... u odabiru statističkih testova
- ... u prikazu rezultata statističkih testova
- ... u tumačenju rezultata statističkih testova

MedCrave
Step into the World of Research
Biometrics & Biostatistics International Journal

Statistical Errors in Medical Publication

Abstract

While errors can be application-related, they can also occur at different stages of the study such as planning, implementation, analysis, interpretation, and presentation, all of which are related to the statistical topics. The statistical errors in the study process can be categorized as the ones that occur

(i) In the research process (before reporting)

(ii) In presentations or publications.

Some statistical errors in a publication can be assessed, while some others cannot be. In publications, not all of the statistical quantities can be checked; it is only possible to check the accuracy of some of the statistics via reported descriptive values. In addition, some terminology, demonstration and interpretation errors can be identified. As such, discussions related to the statistical errors in publications are limited to errors that can possibly be determined.

Review Article

Volume 2 Issue 1 - 2015

Ilker Ercan[1]* and Hakan Demirtas[2]

[1]*Department of Biostatistics, Uludag University, Turkey*

[2]*Division of Epidemiology and Biostatistics, University of Illinois, USA*

***Corresponding author:** Ilker Ercan, Department of Biostatistics, Uludag University, Medical Faculty, Bursa, Turkey, Tel: +90 224 2953881; Email: ercan@uludag.edu.tr

Received: January 28, 2014 | **Published:** February 19, 2015

A. Errors related to p-values
   a. p-values given in closed form,
   b. p-values lacking after statistical tests,
   c. Incorrect p-values,
   d. Incorrect demonstration of p-values
B. Errors related to tests
   I. Statistical technique used but not defined,
   II. Insufficient data presented for the statistical test,
   III. Incorrect name of the statistical test,
   IV. Statistical technique defined but not used,
   V. Use of incorrect test,
   VI. Statistical analysis required but not performed
C. Errors in the summarization of data
D. Mathematical demonstration errors
E. Statistical symbol errors
F. Incomprehensible statistical terms
G. Inappropriate interpretation
H. Errors in (statistical) terminology
I. Incorrect and insufficient demonstration of descriptive statistics
J. Presentation of statistical method-analysis and results in the incorrect section of the manuscript

Review article SWISS MED WKLY 2007;137:44–49 · www.smw.ch 44
Peer reviewed article

Statistical errors in medical research – a review of common pitfalls

Alexander M. Strasak, Qamruz Zaman, Karl P. Pfeiffer, Georg Göbel, Hanno Ulmer

Department of Medical Statistics, Informatics and Health Economics, Medical University Innsbruck

Table 1

Statistical errors and deficiencies related to the design of a study.

Study aims and primary outcome measures not clearly stated or unclear
Failure to report number of participants or observations (sample size)
Failure to report withdrawals from the study
No a priori sample size calculation/effect-size estimation (power calculation)
No clear a priori statement or description of the Null-Hypothesis under investigation
Failure to use and report randomisation
Method of randomisation not clearly stated
Failure to use and report blinding if possible
Failure to report initial equality of baseline characteristics and comparability of study groups
Use of an inappropriate control group
Inappropriate testing for equality of baseline characteristics

Review article
SWISS MED WKLY 2007;137:44–49 · www.smw.ch 44
Peer reviewed article

Statistical errors in medical research – a review of common pitfalls

Alexander M. Strasak, Qamruz Zaman, Karl P. Pfeiffer, Georg Göbel, Hanno Ulmer

Department of Medical Statistics, Informatics and Health Economics, Medical University Innsbruck

Table 2

Statistical errors and deficiencies related to data analysis.

Use of wrong statistical tests
Incompatibility of statistical test with type of data examined
Unpaired tests for paired data or vice versa
Inappropriate use of parametric methods
Use of an inappropriate test for the hypothesis under investigation
Inflation of Type I error
Failure to include a multiple-comparison correction
Inappropriate post-hoc Subgroup analysis
Typical errors with Student's t-test
Failure to prove test assumptions
Unequal sample sizes for paired t-test
Improper multiple pair-wise comparisons of more than two groups
Use of an unpaired t-test for paired data or vice versa
Typical errors with χ^2-tests
No Yates-continuity correction reported if small numbers
Use of chi-square when expected numbers in a cell are <5
No explicit statement of the tested Null-Hypotheses
Failure to use multivariate techniques to adjust for confounding factors

Peer reviewed article

Statistical errors in medical research – a review of common pitfalls

Alexander M. Strasak, Qamruz Zaman, Karl P. Pfeiffer, Georg Göbel, Hanno Ulmer

Department of Medical Statistics, Informatics and Health Economics, Medical University Innsbruck

Table 3

Errors related to the documentation of statistical methods applied.

Failure to specify/define all tests used clear and correctly
Failure to state number of tails
Failure to state if test was paired or unpaired
Wrong names for statistical tests
Referring to unusual or obscure methods without explanation or reference
Failure to specify which test was applied on a given set of data if more than one test was done
"Where appropriate" statement

Review article SWISS MED WKLY 2007;137:44–49 · www.smw.ch 44
Peer reviewed article

Statistical errors in medical research – a review of common pitfalls

Alexander M. Strasak, Qamruz Zaman, Karl P. Pfeiffer, Georg Göbel, Hanno Ulmer

Department of Medical Statistics, Informatics and Health Economics, Medical University Innsbruck

Table 4

Statistical errors and deficiencies related to the presentation of study data.

Inadequate graphical or numerical description of basic data
Mean but no indication of variability of the data
Giving SE instead of SD to describe data
Use of mean (SD) to describe non-normal data
Failure to define ± notion for describing variability or use of unlabeled error bars
Inappropriate and poor reporting of results
Results given only as p-values, no confidence intervals given
Confidence intervals given for each group rather than for contrasts
"p = NS", "p <0.05" or other arbitrary thresholds instead of reporting exact p-values
Numerical information given to an unrealistic level of precision

Peer reviewed article

Statistical errors in medical research – a review of common pitfalls

Alexander M. Strasak, Qamruz Zaman, Karl P. Pfeiffer, Georg Göbel, Hanno Ulmer

Department of Medical Statistics, Informatics and Health Economics, Medical University Innsbruck

Table 5

Statistical errors and deficiencies related to the interpretation of study findings.

Wrong interpretation of results
"non significant" interpreted as "no effect", or "no difference"
Drawing conclusions not supported by the study data
Significance claimed without data analysis or statistical test mentioned
Poor interpretation of results
Disregard for Type II error when reporting non-significant results
Missing discussion of the problem of multiple significance testing if done
Failure to discuss sources of potential bias and confounding factors

Biochemia Medica 2007;17(1):10-5

Odabrane teme iz biostatistike **Lessons in biostatistics**

Što treba znati kada izračunavamo koeficijent korelacije?

What we need to know when calculating the coefficient of correlation?

Martina Udovičić[1], Ksenija Baždarić[1], Lidija Bilić-Zulle[1,2], Mladen Petrovečki[1,3]

[1]Katedra za medicinsku informatiku, Medicinski fakultet Sveučilišta u Rijeci, Rijeka
[1]Department of Medical Informatics, School of Medicine, University of Rijeka, Rijeka, Croatia
[2]Zavod za laboratorijsku dijagnostiku, Klinički bolnički centar „Rijeka", Rijeka

- uvjeti za izračunavanje povezanosti (veličina uzorka i raspodjela podataka)
- tumačenje značajnosti koeficijenta korelacije
- zaključivanje o uzročno posljedičnoj vezi
- usporedba dvaju koefcijenata korelacije

Biochemia Medica 2009:19(3):294-300

Izvorni stručni članak **Original professional article**

Statističke pogreške u rukopisima zaprimljenim u časopis *Biochemia Medica*

Statistical errors in manuscripts submitted to *Biochemia Medica* journal

Ana-Maria Šimundić, Nora Nikolac

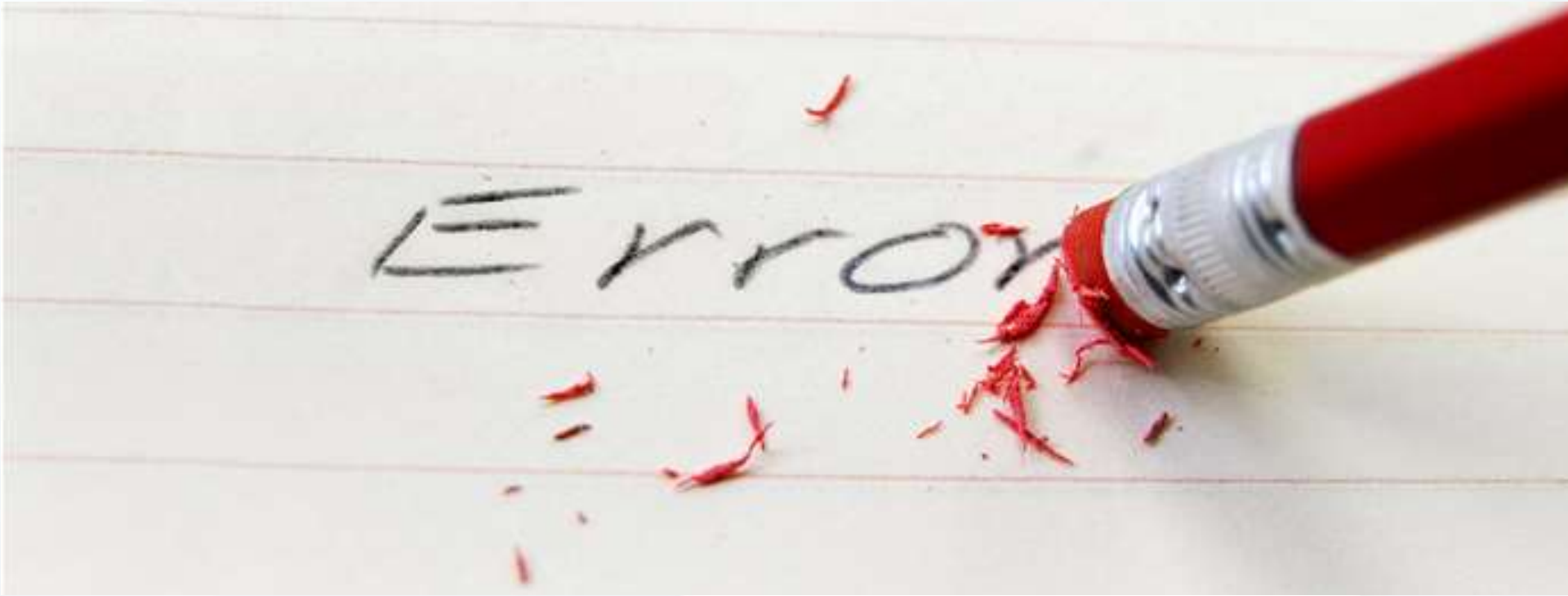
Klinički zavod za kemiju, Klinička bolnica „Sestre milosrdnice", Zagreb

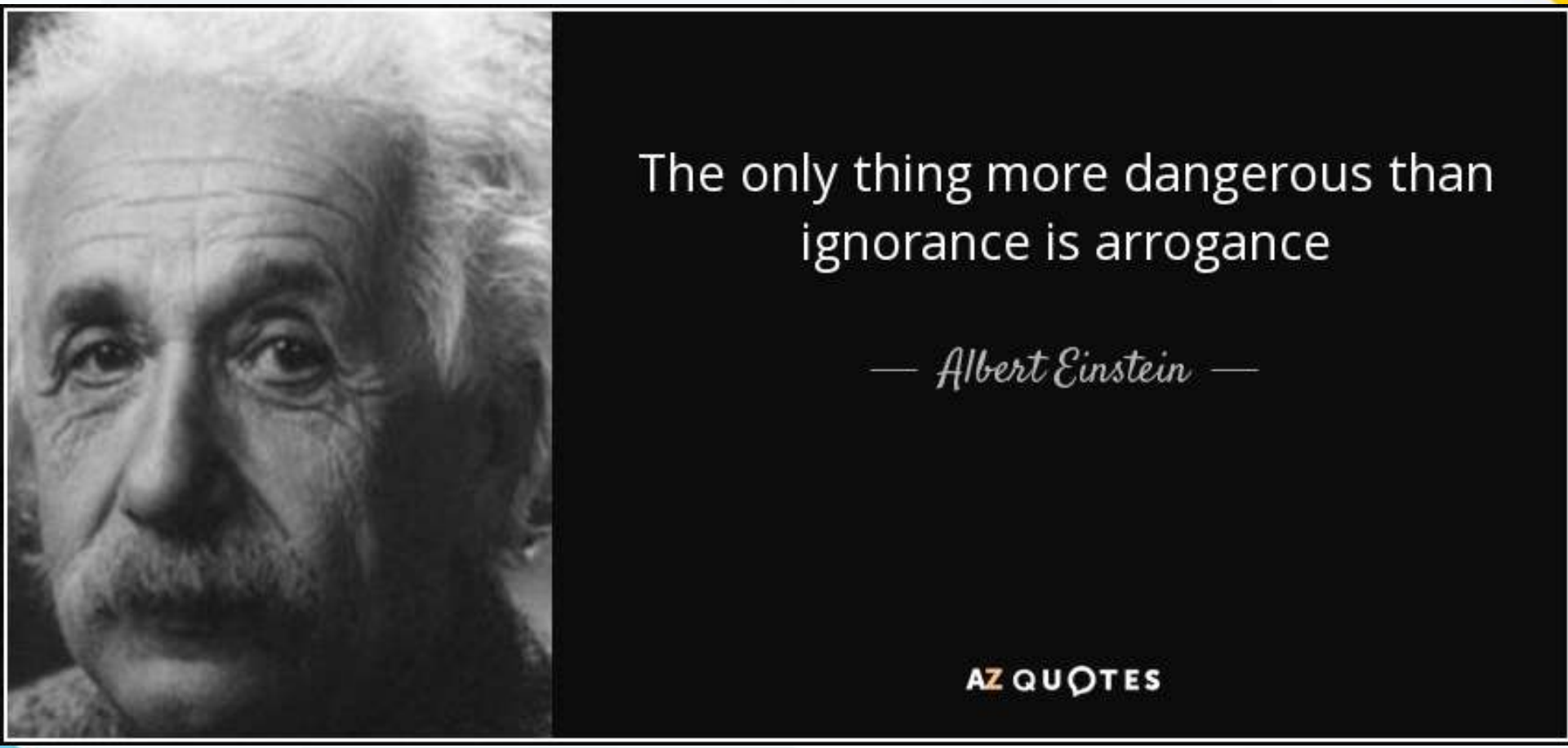
University Department of Chemistry, Sestre milosrdnice University Hospital, Zagreb, Croatia

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.

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

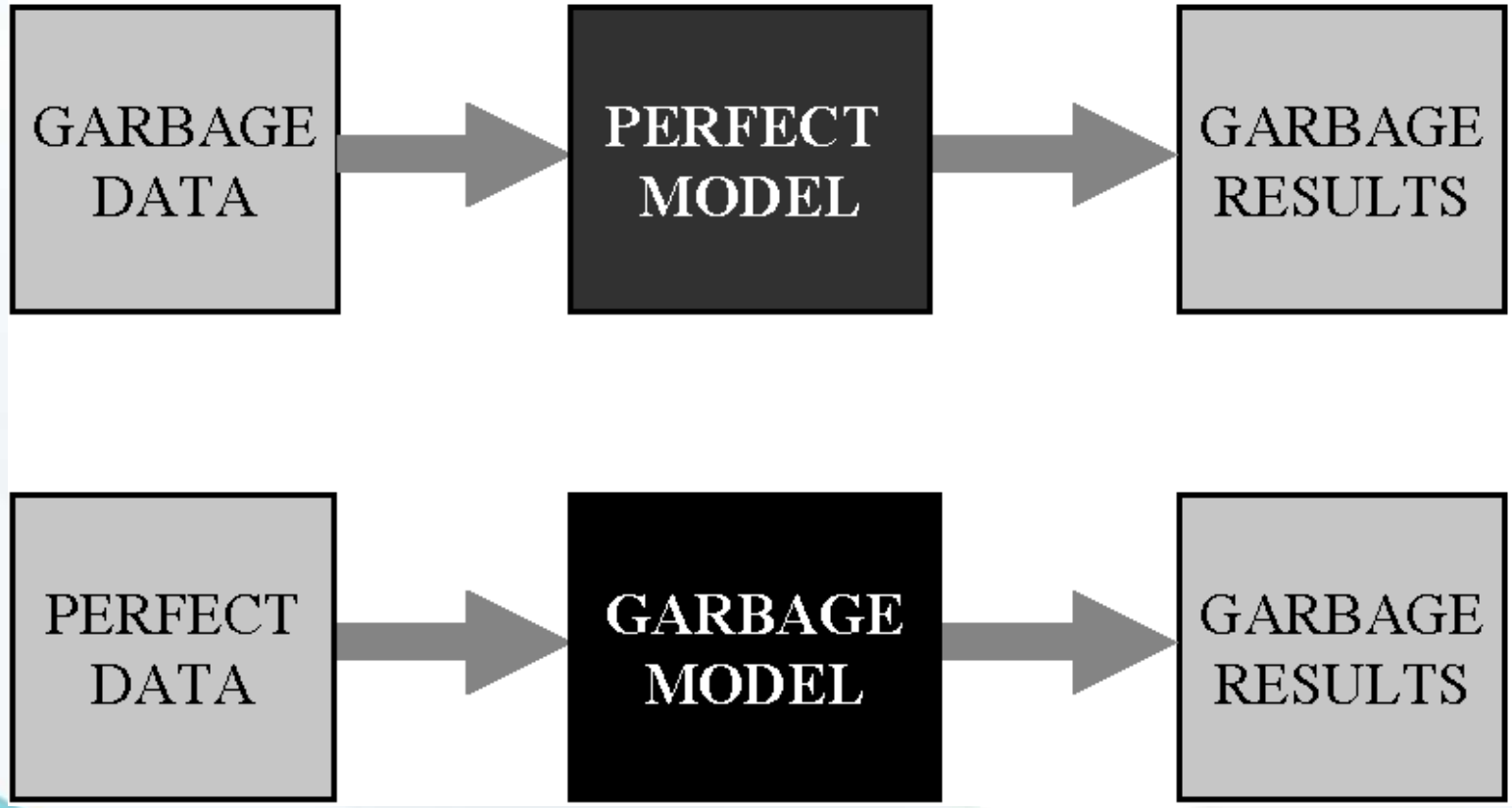
Error	Error rate N (proportion)
Power analysis not provided	55/55 (1.0)
Incorrect use of statistical test for comparing three or more groups for differences	21/28 (0.75)
Incorrect presentation of P value	36/54 (0.66)
Incorrect choice of the statistical test	34/55 (0.62)
Incorrect interpretation of correlation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


ERROR


The only thing more dangerous than ignorance is arrogance
— Albert Einstein —
AZ QUOTES

MODEL CALCULATIONS

"Garbage In-garbage Out" Paradigm



http://101proofsforgod.blogspot.hr/2014/02/52-garbage-in-garbage-out-gigo.html

"If you torture the data long enough, it will confess."

Ronald Coase

Professor Emeritus of Economics
University of Chicago Law School

[WhatYouCanDo:]

- spoznati važnost pravilne obrade podataka
- naučiti osnove biostatistike
- uključiti statističara u tim prije početka istraživanja
- dati rad na statističku predrecenziju
- ne kopirati tuđe statističke obrade
- oprez s lako dostupnom programskom potporom za obradu podataka

Što časopisi mogu učiniti...

- objaviti opširne upute za autore
- uvoditi statističke urednike i recenzente
- za svaki rad s obradom podataka učiniti statističku recenziju
- educirati recenzente, autore i urednike

THE LANCET

Zahvaljujem na pozornosti ☺

lidija.bilic.zulle@uniri.hr