Introduction to Statistics

Anna-Bettina Haidich

Associate Professor in Medical Statistics-Epidemiology

Department of Medicine AUTH

haidich@auth.gr

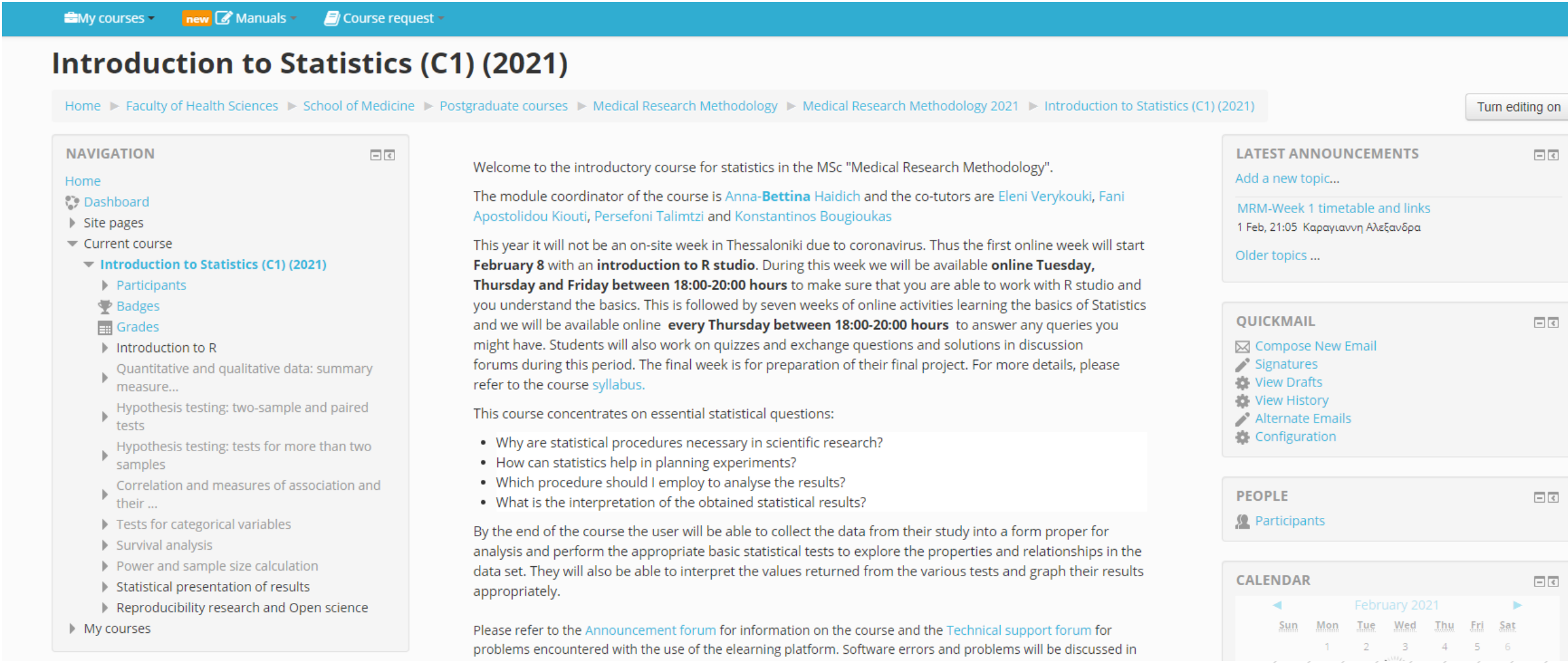
MSc in Medical Research Methodology

Medical School
Aristotle University of Thessaloniki

Materials and announcements

- elearning.auth.gr
- School of Medicine
- Postgraduate Courses
- Medical Research Methodology
- Course “Introduction to Statistics (C1) (2021)”

elearning platform (moodle)



On-line outline

Starting date		Topics	Tutors
8 Feb 2021	Online 1	Introduction to R	All
15 Feb 2021	Online 2	Summary measures for quantitative and qualitative data - Graphs - Normal distribution	Bettina+Eleni+Persefoni
22 Feb 2021	Online 3	Hypothesis testing- paired and two-sample t-tests: Mann-Whitney U test and Wilcoxon Signed Ranks test	Bettina+ Konstantinos
1 Mar 2021	Online 4	Hypothesis testing - tests for more than two samples: ANOVA and Kruskal-Wallis test	Eleni + Konstantinos
8 Mar 2021	Online 5	Correlation and Measures of association and their confidence intervals	Konstantinos+ Bettina + Persefoni
15 Mar 2021	Online 6	Tests for categorical variables : $\chi 2$, Fisher's exact test, Mc Nemar's test	Bettina + Persefoni
22 Mar 2021	Online 7	Survival analysis: Log-rank test and Kaplan-Meier plots	Eleni+ Persefoni
29 Mar 2021	Online 8	Power and sample size calculation	Bettina + Persefoni
5 April 2021	Free		
12 April 2021	Assignment	Final Project	All

On-line support

- Three days this week **8-12/2/2021** we will be online between **18:00-20:00** using Zoom
  - **https://authgr.zoom.us/j/94316381406**

- Thereafter every **Thursday 18:00-20:00** till **1/4/2021**

- All problems with R will be solved!

Grading system

Component	% of grade	When
Lecture participation through online lectures and forum discussion	15%	Daily
Quizzes	15%	Weekly
Final Project	70%	**April 12, 2021**
Total	**100%**	

For the Final Project you will be analyzing either your own dataset or one provided by us depending on your choice

Communication

- Anna-**Bettina** Haidich haidich@auth.gr
- Eleni Verykouki everykouki@auth.gr
- Konstantinos Bougioukas mpougioukas@auth.gr
- Persefoni Talimtzi talimtzi@auth.gr
- Fani Apostolidou-Kiouti fania@auth.gr

Why do we need Statistics?

Statistics!

Google's prediction: What will be the "sexy" job in the next ten years?

Here's a strange prediction from Google's Chief Economist: "I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding."

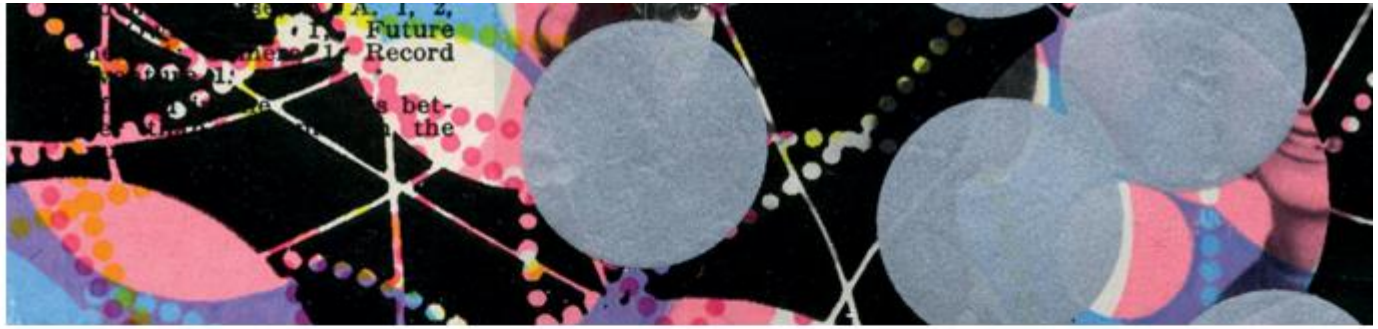
That quote came from a New York Times article about the rapidly increasing demand for statisticians.

- Hal Varian, Google Chief Economist, August 2009

Statistics!

Headline from Harvard Business Review:

Harvard Business Review



DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Statistics!

The New York Times

TECHNOLOGY

For Today's Graduate, Just One Word: Statistics

By STEVE LOHR AUG. 5, 2009

Statistical analysis section is a MUST in medical research papers

- Statistical analysis section

We summarized continuous variables with mean and standard deviation or median and interquartile range, whenever they were not normally distributed and presented categorical variables with frequencies and the corresponding percentages. We created a dichotomous variable using a cutoff point of 30 in the score of IAT to distinguish between students with normal Internet use (below 31) and students with Internet addiction (from 31 to 100), including all students with mild, moderate, and severe Internet addiction. We used a *t* test or Mann-Whitney *U* test for not normally distributed variables to compare continuous variables between the two groups. We assessed the association of the categorical variables between the two groups with the chi-square test or Fisher's exact test, whenever the expected cell frequency was <5. Logistic regression with Internet addiction as a dependent variable was used adjusting for sex and age. A multivariable model was built with backward elimination based on likelihood ratio criteria, where univariable predictors with $p<0.20$ were considered for inclusion. We conducted statistical analyses in IBM/SPSS Statistics version 21.0. *p* values are two-tailed.

Internet Addiction in Greek Medical Students: an Online Survey

Zoi Tsimtsiou • Anna-Bettina Haidich • Dimitris Spachos • Stamatia Kokkali • Panagiotis Bamidis • Theodoros Dardavesis • Malamatenia Arvanitidou

Internet Addiction in Greek Medical Students: an Online Survey

Zoi Tsimtsiou · Anna-Bettina Haidich · Dimitris Spachos · Stamatia Kokkali · Panagiotis Bamidis · Theodoros Dardavesis · Malamatenia Arvanitidou

Table 1 Sociodemographic characteristics of students with Internet addiction and normal Internet users

	Internet addiction	Normal Internet use	*p* value
Mean age in years (SD)	21.2 (2.7)	21.4 (3.1)	0.570
	n (%)	*n* (%)	
Gender			0.274
Males	76 (32.6)	157 (67.4)	
Females	85 (28.2)	216 (71.8)	
Nationality			0.288
Greek	148 (29.6)	352 (70.4)	
Other[a]	13 (38.2)	21 (61.8)	
Academic year			0.764
1st	31 (31.3)	68 (68.7)	
2nd	29 (32.6)	60 (67.4)	
3rd	40 (30.8)	90 (69.2)	
4th	19 (22.4)	66 (77.6)	
5th	25 (32.5)	52 (67.5)	
6th	6 (27.3)	16 (72.7)	
Degree	11 (34.4)	21 (65.6)	

Internet Addiction in Greek Medical Students: an Online Survey

Zoi Tsimtsiou · Anna-Bettina Haidich · Dimitris Spachos · Stamatia Kokkali · Panagiotis Bamidis · Theodoros Dardavesis · Malamatenia Arvanitidou

Abstract

Objective The authors investigated the prevalence of Internet addiction (IA) in undergraduate medical students to identify possible associations with sociodemographics and Internet habits.

Methods All students at the Aristotle University of Thessaloniki School of Medicine, Greece, were invited to complete the online Internet Addiction Test (IAT) along with sociodemographics and preferences on Internet activities.

Results The authors received 585 responses after three reminders (23.5 % response rate). Mild IA was found in 24.5 %, moderate in 5.4 %, and severe in 0.2 %. In multivariable analysis, the odds to develop IA were increased with visits in Internet cafes (Odds Ratio [OR] 3.49, 95 % Confidence Interval [CI]: 1.45, 8.46), the use of Facebook (OR 2.43, 95 % CI: 1.35, 4.38), Twitter (OR 2.45, 95 % CI: 1.37, 4.39), and online games (OR 1.95, 95 % CI: 1.29, 2.94). Using e-mails seemed to be protective against IA (OR 0.59, 95 % CI: 0.37, 0.94).

Conclusion This is the first IA prevalence study in a European medical school. Early-detection systems and other ways to help students with pathological behaviors should be developed.

Data are everywhere!

Original Investigation

Postoperative Glaucoma Following Infantile Cataract Surgery
An Individual Patient Data Meta-analysis

Asimina Mataftsi, MD, PhD, MRCOphth; Anna-Bettina Haidich, PhD; Stamatia Kokkali, MD, PhD;

RESULTS Seven centers contributed individual patient data on 470 infants with a median age at surgery of 3.0 months and median follow-up of 6.0 years. Eighty patients (17.0%) developed glaucoma at a median follow-up of 4.3 years. Only 2 of these patients had a pseudophakic eye. The risk for postoperative glaucoma appeared to be lower after primary implantation (hazard ratio [HR], 0.10 [95% CI, 0.01-0.70]; $P = .02$; $I^2 = 34\%$), higher after surgery at 4 weeks or younger (HR, 2.10 [95% CI, 1.14-3.84]; $P = .02$; $I^2 = 0\%$), and higher after additional procedures (HR, 2.52 [95% CI, 1.11-5.72]; $P = .03$; $I^2 = 32\%$). In multivariable analysis, additional procedures independently increased the risk for glaucoma (HR, 2.25 [95% CI, 1.20-4.21]; $P = .01$), and primary implantation independently reduced it (HR, 0.10 [95% CI, 0.01-0.76]; $P = .03$). Results were similar in the aggregate data meta-analysis that included data from 10 published articles.

CONCLUSIONS AND RELEVANCE Although confounding factors such as size of the eye and surgeon experience are not accounted for in this meta-analysis, the risk for postoperative glaucoma after infantile cataract surgery appears to be influenced by the timing of surgery, primary implantation, and additional intraocular surgery.

JAMA Ophthalmol. 2014;132(9):1059-1067. doi:10.1001/jamaophthalmol.2014.1042
Published online June 12, 2014.

Efficacy and safety of tigecycline for the treatment of infectious diseases: a meta-analysis

Efthimia Tasina, Anna-Bettina Haidich, Stamatia Kokkali, Malamatenia Arvanitidou

Lancet Infect Dis 2011; 11: 834–44

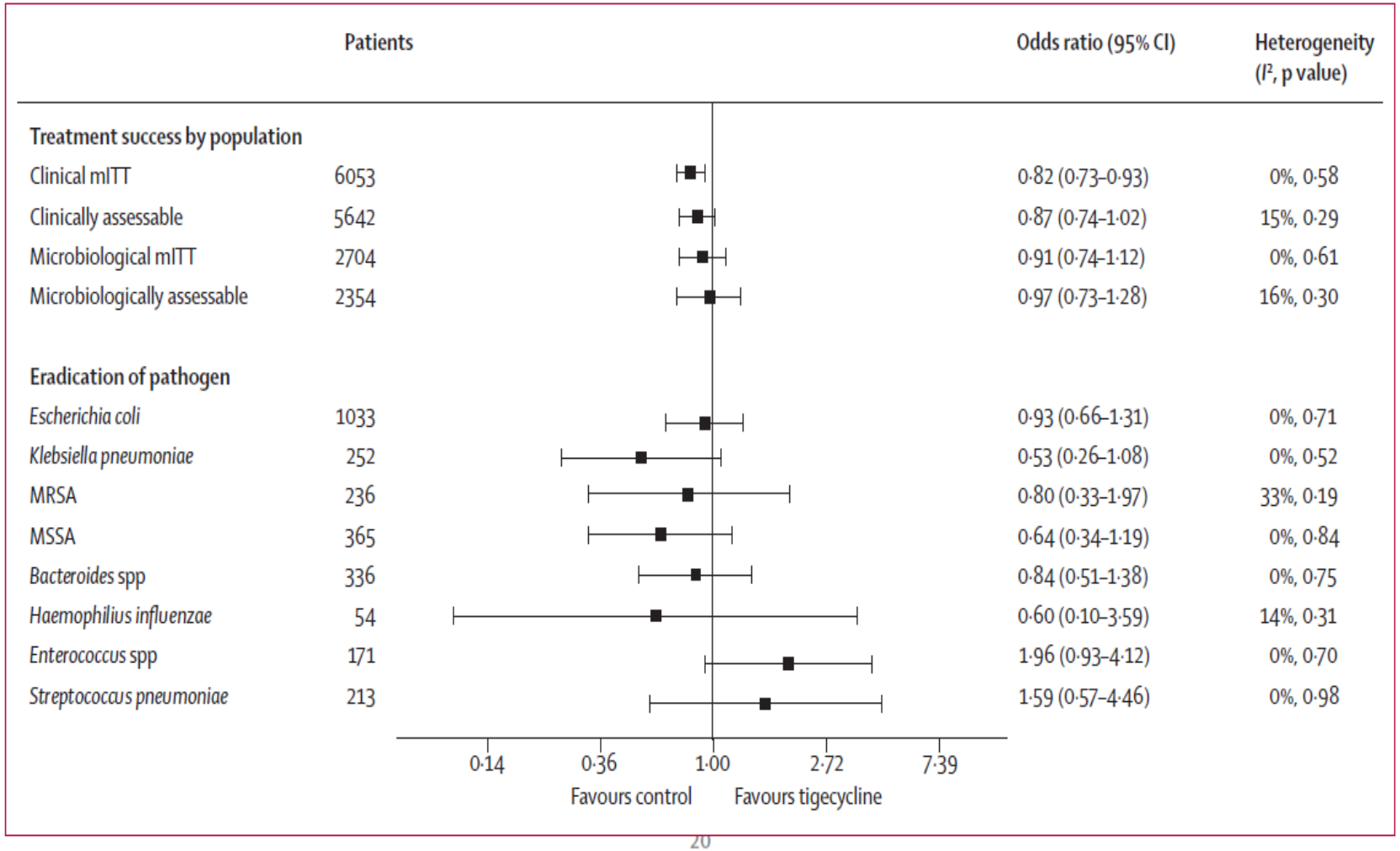


Figure 4: Comparative effectiveness of tigecycline versus comparator antibiotics

ELSEVIER

Journal of Clinical Epidemiology 106 (2019) 70–79

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Reporting guidelines on how to write a complete and transparent abstract for overviews of systematic reviews of health care interventions

Konstantinos I. Bougioukas, Emmanouil Bouras, Fani Apostolidou-Kiouti, Stamatia Kokkali, Malamatenia Arvanitidou, Anna-Bettina Haidich*

Department of Hygiene, Social-Preventive Medicine & Medical Statistics, Faculty of Health Sciences, School of Medicine, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece

Abstract

Objective: An overview of systematic reviews (OoSRs) is a study designed to offer a broad view of evidence from existing systematic reviews (SRs). The abstract is an important part of an OoSRs as it can determine whether reading the full text is of interest. The aim of this article is to offer guidelines to promote transparent and sufficient reporting in abstracts of OoSRs of health care interventions.

Study Design and Setting: The items were developed by combining key features from abstracts of OoSRs, PRISMA for abstracts, and our published reporting guidelines for OoSRs. The initial version was distributed to experts to give feedback; pilot testing by a group of researchers followed. The refined checklist was applied by two reviewers independently in a sample of 40 abstracts.

Results: The developed instrument "Preferred Reporting Items for OoSRs abstracts" (PRIO for abstracts) consists of six sections with 15 topics including 20 items in total. The mean inter-rater reliability was 0.87 (95% confidence interval: 0.82, 0.92). An explanation and at least one published example of good reporting per item are provided.

Conclusion: This instrument will assist authors in writing transparent and informative abstracts for OoSRs and can be adopted by journals that publish OoSRs.

Keywords: Abstract; Checklist; Overview of systematic reviews; PRIO for abstracts; PRIO-Harms; Reporting guidelines

Journal of Clinical Epidemiology 106 (2019) 70–79

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Reporting guidelines on how to write a complete and transparent abstract for overviews of systematic reviews of health care interventions

Konstantinos I. Bougioukas, Emmanouil Bouras, Fani Apostolidou-Kiouti, Stamatia Kokkali, Malamatenia Arvanitidou, Anna-Bettina Haidich*

Department of Hygiene, Social-Preventive Medicine & Medical Statistics, Faculty of Health Sciences, School of Medicine, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece

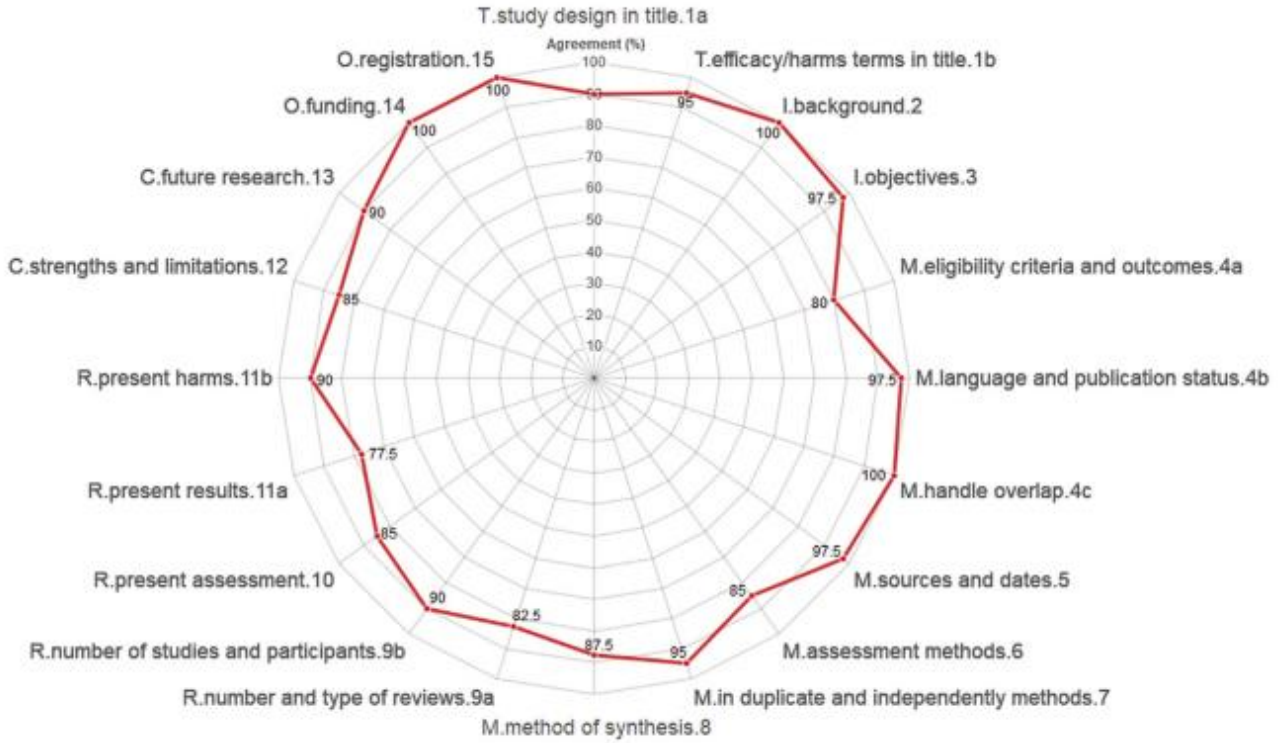
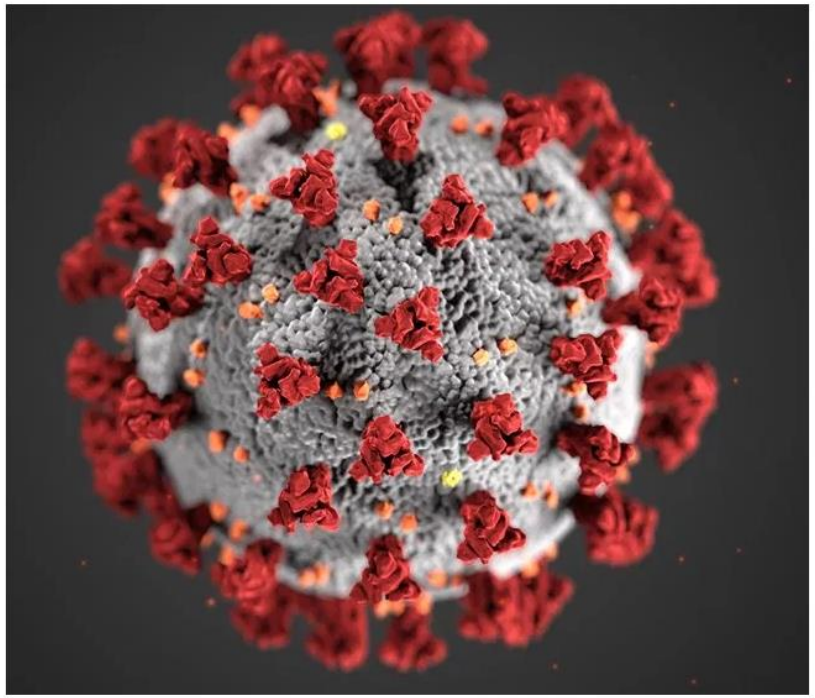


Fig. 2. Radar plot showing the percentage of agreement between the two reviewers for each abstract item. T, Title; I, Introduction; M, Methods; R, Results; C, Conclusions; O, Other.

https://retractionwatch.com/retracted-coronavirus-covid-19-papers/

Retracted coronavirus (COVID-19) papers

N=69

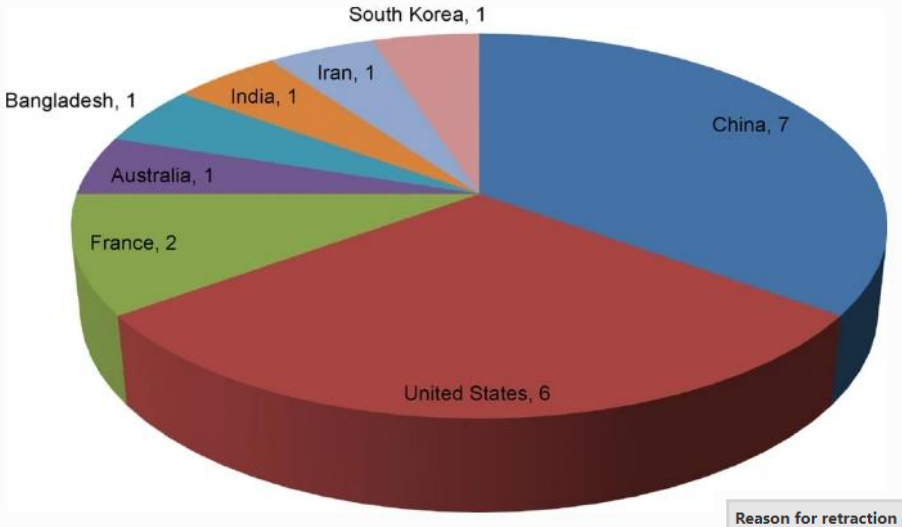


Published: 05 August 2020

Retracted COVID-19 articles: a side-effect of the hot race to publication

Parisa Soltani & Romeo Patini

Fig. 1



Number of retracted/withdrawn COVID-19 articles from different countries

Reason for retraction or withdrawal	Number of cases
Concerns, issues, or errors in results and/or conclusions	8
Concerns, issues, or errors in data	7
Duplication due to error of the journal/publisher	2
Duplication due to report of the case by another team	1
Error in analysis	1
Copyright claims	1
Ethical violation by the author	1
Lack of approval from a third party	1
Concerns/issues about third party involvement	1
Investigation by a third party	1
Concerns/issues about authorship (dishonest presentation)	1
Not mentioned	2

Data Provides Information

- Good data can be analysed and summarized to provide useful information

  (Good reporting for good health)

- Bad data can be analyzed and summarized to provide incorrect/harmful/non-informative result

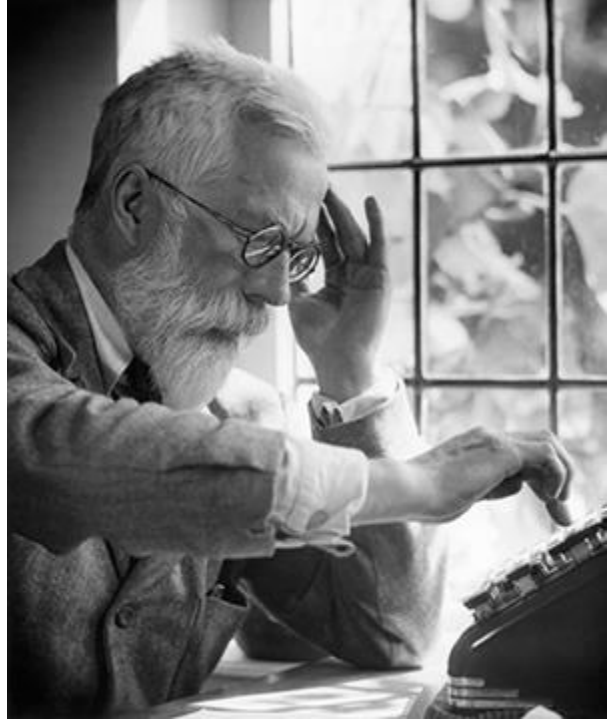
Steps in Research Project

- Planning/Design of Study
- Data collection
- Data analysis
- Presentation
- Interpretation

Statistics CAN play a role in each of these steps! (but mostly is only called upon the data analysis part)

Ronald Fisher

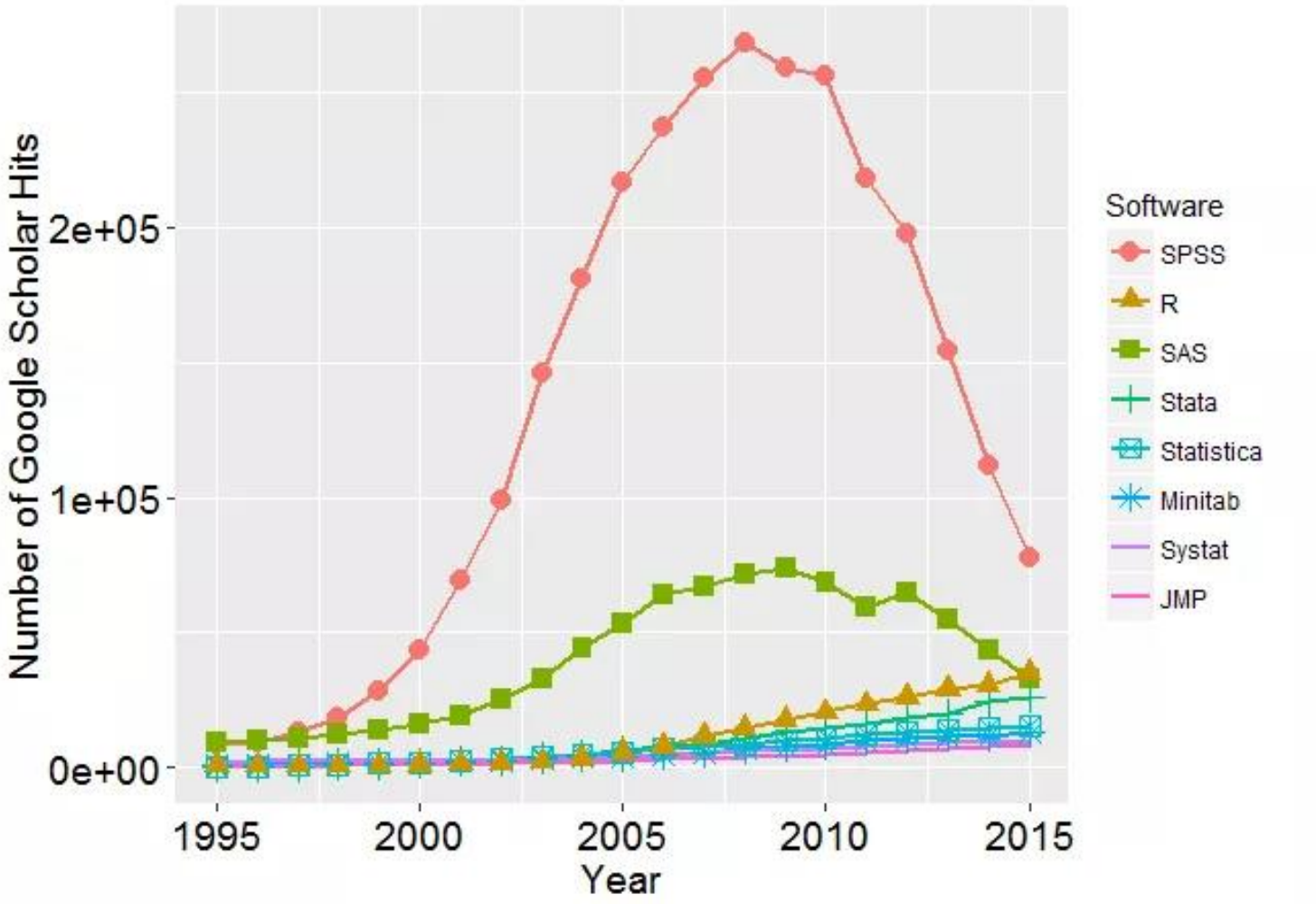
- As the influential twentieth-century statistician Ronald Fisher said: “To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.”

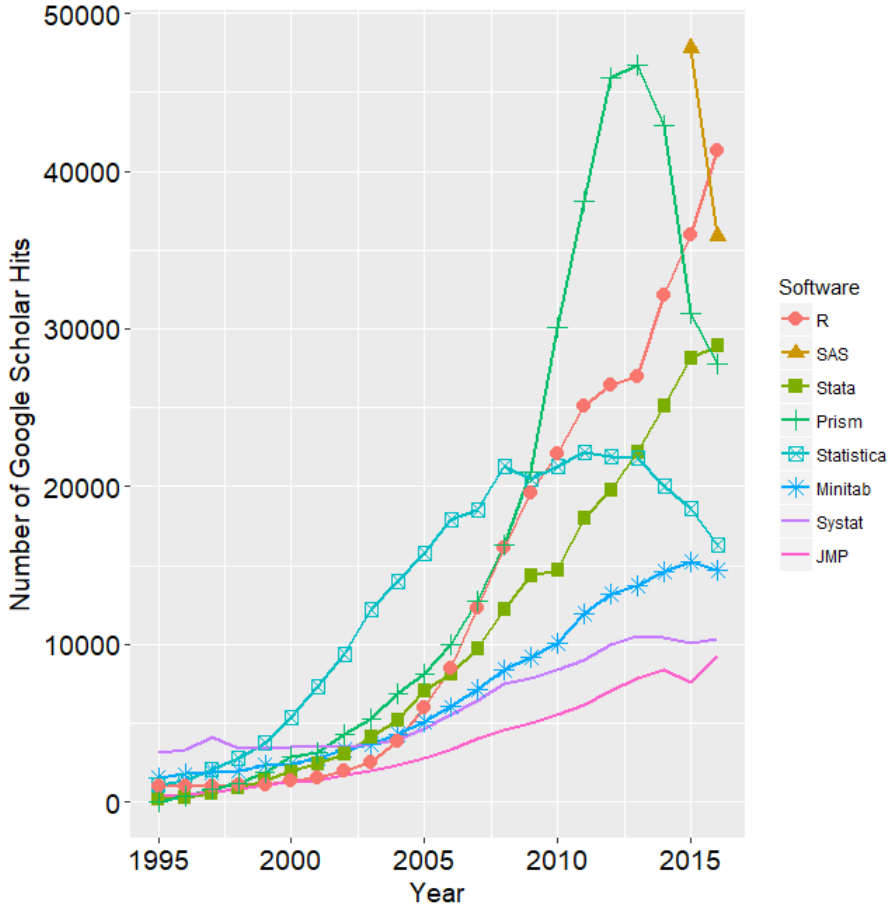


Aim of the course

- Distinguish between qualitative and quantitative data and know how to summarize them
- Hypothesis testing
- Which statistical procedure should I employ to analyse the results?
- What do the statistical results actually mean when I have got them and how to present them ?
- What sample size is needed for my study?
- You will also be trained to calculate and generate these statistics yourself using freely available statistical software R

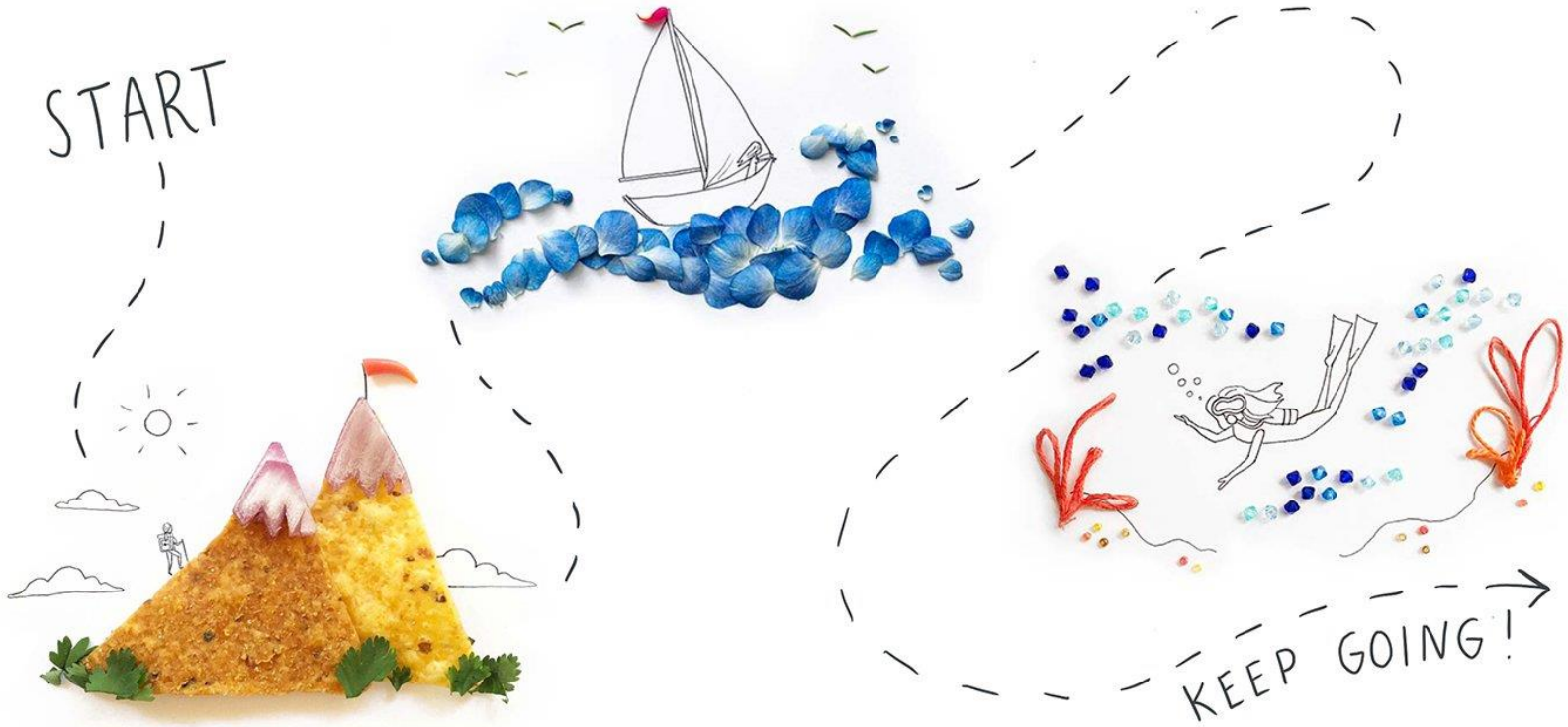
Why R?





http://r4stats.com/articles/popularity/

Learning R



Let's get started with R !