EVALUACIJA PRIMENJENIH STATISTIČKIH METODA U OPSERVACIONIM STUDIJAMA IZ OBLASTI NEUROLOGIJE

ORIGINALNI RAD

ORIGINAL ARTICLE

EVALUATION OF STATISTICAL METHODS APPLIED IN OBSERVATIONAL STUDIES IN NEUROLOGY

Jelena Tomašević[1], Mirjana Pajčin[1], Dejana Stanisavljević[2], Nenad Milošević[1], Smiljana Kostić[3], Milena Stanić[4]

[1] Univerzitet u Prištini sa privremenim sedištem u Kosovskoj Mitrovici, Medicinski fakultet, Kosovska Mitrovica, Srbija

[2] Univerzitet u Beogradu, Medicinski fakultet, Beograd, Srbija

[3] Univerzitet odbrane, Medicinski fakultet Vojnomedicinske akademije, Klinika za neurologiju, Beograd, Srbija

[4] Dom zdravlja Užice, Užice, Srbija

[1] *University of Pristina in Kosovska Mitrovica, Faculty of Medicine, Kosovska Mitrovica, Serbia*

[2] *University of Belgrade, Faculty of Medicine, Belgrade, Serbia*

[3] *University of Defense, Medical Faculty of the Military Medical Academy, Clinic of Neurology, Belgrade, Serbia*

[4] *Health Care Center Užice, Užice, Serbia*

SAŽETAK

Uvod/Cilj: Statističke metode imaju ključnu ulogu u planiranju, analizi i interpretaciji rezultata medicinskih istraživanja. Zbog kompleksnosti neuroloških oboljenja i heterogenosti ispitivanih populacija, pravilna primena statističkih metoda u neurologiji je od posebnog značaja za donošenje pouzdanih kliničkih zaključaka. Cilj ovog rada je da se izvrši evaluacija primene statističkih metoda u opservacionim studijama iz oblasti neurologije, sa posebnim osvrtom na učestalost korišćenja različitih statističkih tehnika, adekvatnost njihove primene i način tumačenja rezultata.

Materijal i metode: Sprovedena je pretraga literature u bibliografskoj bazi *MEDLINE/PubMed* za period do 1. decembra 2024. godine korišćenjem termina *„neurology"* uz limiter *„observational study"*. Nakon primene kriterijuma za uključivanje i isključivanje iz studije, metodom slučajnog izbora analizirano je 95 opservacionih studija publikovanih između 2013. i 2024. godine. Dva istraživača su nezavisno ekstrahovala podatke korišćenjem unapred pripremljenog obrasca.

Rezultati: Retrospektivni dizajn je imalo 53,7% studija, dok je 46,3% bilo prospektivno. Etičko odobrenje je navedeno u 78,9% radova. Procena normalnosti raspodele podataka je prikazana u 29,5% studija. Najčešće korišćeni statistički testovi su bili hi-kvadratni test (46,3%), *t*-test (36,8%), Fišerov test (36,8%) i Man-Vitnijev test (33,7%). Regresiona analiza je primenjena u 33,7% studija. Planiranje veličine uzorka je bilo jasno navedeno u samo 10,5% radova. Primena *t*-testa je bila statistički značajno učestalija u studijama koje su prethodno prikazale procenu normalnosti raspodele podataka ($p < 0,001$).

Zaključak: Rezultati pokazuju da se statističke metode u opservacionim neurološkim studijama uglavnom primenjuju adekvatno i u skladu sa karakteristikama podataka. Ipak, postoji potreba za češćom proverom statističkih pretpostavki i preciznijim planiranjem veličine uzorka radi unapređenja kvaliteta i pouzdanosti budućih istraživanja.

Ključne reči: normalnost raspodele, veličina uzorka, *t*-test, hi-kvadratni test, regresiona analiza

ABSTRACT

Introduction/Objective: Statistical methods play a key role in the planning, analysis, and interpretation of results in medical research. Due to the complexity of neurological diseases and the heterogeneity of study populations, the proper application of statistical methods in neurology is particularly important for drawing reliable clinical conclusions. This study aims to evaluate the application of statistical methods in observational studies in neurology, with special emphasis on the frequency of the use of different statistical techniques, the adequacy of their application, and the interpretation of results.

Materials and methods: A literature search of the *MEDLINE/PubMed* database was conducted for all publications available up to December 1, 2024, using the term "neurology" with the application of the filter "observational study." After applying the inclusion and exclusion criteria, 95 observational studies published between 2013 and 2024 were randomly selected for analysis. Two researchers independently extracted the data using a previously designed data collection form.

Results: A retrospective design was identified in 53.7% of studies, while 46.3% were prospective. Ethical approval was reported in 78.9% of studies. Assessment of the normality of distribution was presented in 29.5% of studies. The most commonly used statistical tests were the chi-square test (46.3%), *t*-test (36.8%), Fisher's exact test (36.8%), and the Mann–Whitney test (33.7%). Regression analysis was applied in 33.7% of studies. Sample size planning was clearly reported in only 10.5% of studies. The application of the *t*-test was statistically significantly more frequent in studies that reported a previous assessment of the normality of distribution ($p < 0.001$).

Conclusion: The results indicate that statistical methods in observational neurological studies are generally applied adequately and in accordance with data characteristics. However, greater attention should be devoted to testing statistical assumptions and proper sample size planning in order to further improve the quality and reliability of future research.

Keywords: normality of distribution, sample size, *t*-test, chi-square test, regression analysis

Autor za korespondenciju:
Jelena Tomašević
Katedra za preventivnu medicinu, Medicinski fakultet, Univerzitet u Prištini
Anri Dinana bb, 38220 Kosovska Mitrovica, Srbija
Elektronska adresa: jelenakm0@gmail.com

Corresponding author:
Jelena Tomašević
Department of Preventive Medicine, Faculty of Medicine, University of Pristina in Kosovska Mitrovica
bb Anri Dinana Street, 38220 Kosovska Mitrovica, Serbia
E-mail: jelenakm0@gmail.com

Primljeno • *Received*: May 21, 2026; **Revidirano • *Revised*:** June 11, 2026; **Prihvaćeno • *Accepted*:** June 12, 2026; ***Online first*:** June 25, 2026

DOI: 10.5937/smclk7-67463

UVOD

Statističke metode omogućavaju istraživačima da objektivno procene hipoteze, izbegnu pristrasnost i donesu zaključak na osnovu koga mogu da izrade preporuke koje se mogu generalizovati na širu populaciju [1]. Statistika ima ključnu ulogu u planiranju istraživanja, određivanju adekvatne veličine uzorka, izboru odgovarajućih statističkih metoda, kao i u pravilnom predstavljanju i tumačenju dobijenih rezultata [2].

S obzirom na raznovrsnost i kompleksnost kliničkih stanja koje neurologija proučava, kao i na heterogenost populacije obuhvaćene studijama, istraživanja u neurologiji zahtevaju pažljivo planiranje i primenu statističkih metoda, kako bi se u dovoljnoj meri obuhvatile razlike među pacijentima i izbegla pogrešna generalizacija rezultata. Brojne neurološke bolesti odlikuju se varijabilnim kliničkim slikama, različitim oblicima progresije i često multifaktorskom etiologijom [3]. Mnoge neurološke studije zasnivaju se na longitudinalnim vrstama dizajna i kompleksnim mernim alatima (npr. nuklearna magnetna rezonanca (NMR), elektroencefalogram (EEG), skale funkcionalne sposobnosti), što iziskuje kompleksniju statističku obradu podataka [4]. Jedan od značajnih izazova u neurološkim istraživanjima, naročito u studijama koje se bave retkim bolestima ili specifičnim sindromima, jeste ograničena veličina uzorka, što može da utiče na statističku snagu i pouzdanost rezultata. Na sve to, treba dodati i izazove u definisanju ishoda – mnogi neurološki ishodi su subjektivni ili se mere putem složenih skala čija se pouzdanost i validnost mora prethodno proveriti [5]. Nepravilno primenjene statističke metode u medicinskim istraživanjima mogu da imaju ozbiljne posledice po kliničku praksu, posebno u oblastima kao što je neurologija, gde statistički zaključci predstavljaju osnovu za donošenje dijagnostičkih i terapijskih odluka.

Pregled prethodnih evaluacija jasno pokazuje da je dosadašnja praksa često nedovoljno rigorozna, te da postoji značajan potencijal za unapređenje kroz usvajanje međunarodno priznatih standarda i smernica [1–5].

Cilj rada je da se izvrši evaluacija primene statističkih metoda u naučnoistraživačkim radovima iz oblasti neurologije, sa posebnim osvrtom na učestalost korišćenja različitih statističkih tehnika, adekvatnost (ili pravilnost) njihove primene i način interpretacije rezultata.

MATERIJAL I METODE

Strategija pretraživanja

Sprovedena je temeljna pretraga literature u okviru bibliografske baze podataka *MEDLINE/PubMed* obuhvativši sve dostupne literaturne jedinice do 1. decembra 2024. Pretraga je izvršena na osnovu ključne reči „*neu-*

INTRODUCTION

Statistical methods allow researchers to objectively evaluate hypotheses, avoid bias, and draw conclusions based on which they can develop recommendations that can be generalized to a wider population [1]. Statistics plays a crucial role in research planning, determining an adequate sample size, selecting appropriate statistical methods, and correctly presenting and interpreting the results obtained [2].

Given the diversity and complexity of the clinical conditions studied in neurology, as well as the heterogeneity of the populations included in studies, neurological research requires careful planning and the application of statistical methods to adequately account for differences between patients and avoid incorrect generalization of results. Numerous neurological diseases are characterized by variable clinical presentations, different forms of progression, and often multifactorial etiology [3]. Many neurological studies are based on longitudinal designs and complex measurement tools (e.g., nuclear magnetic resonance (NMR), electroencephalogram (EEG), functional ability scales), which require more complex statistical data processing [4]. One of the significant challenges in neurological research, particularly in studies involving rare diseases or specific syndromes, is limited sample size, which can affect the statistical power and the reliability of results. There are also additional challenges in defining outcomes – many neurological outcomes are subjective or are measured using complex scales whose reliability and validity must be verified beforehand [5]. Improperly applied statistical methods in medical research can have serious consequences for clinical practice, especially in fields such as neurology, where statistical conclusions form the basis for diagnostic and therapeutic decisions.

A review of previous evaluations clearly shows that current practice is often insufficiently rigorous and that there is significant potential for improvement by adopting internationally recognized standards and guidelines [1–5].

This article aims to evaluate the application of statistical methods in scientific research articles in neurology, with special emphasis on the frequency of use of various statistical techniques, the adequacy (or correctness) of their application, and the manner in which results are interpreted.

MATERIALS AND METHODS

Search strategy

A thorough literature search of the *MEDLINE/PubMed* database was conducted, including all available literature published up to December 1, 2024. The search

rology" u naslovu i sažetku, a dizajn pretraženih literaturnih jedinica je podrazumevao opservacione studije: *neurology[Title/Abstract]* uz korišćenje limitatora „*observational study*".

Kako bi osigurali sveobuhvatan pristup literaturi, proces pretraživanja nije obuhvatao jezička ili vremenska ograničenja publikovanja.

Kao kriterijumi za isključivanje iz analize navedeni su: istraživanja na životinjama, ili *in vitro*, pregledi literature, uvodnici, opisi budućih studija ili opisi pilot projekata.

Izvođenjem literaturnog pretraživanja identifikovano je 569 radova objavljenih u periodu od 2013. do 2025. godine. Preuzeta je lista sa numerisanim referencama i rasporedom publikacija od najranije do najskorije publikovanih radova. Od 569 radova, za dalju analizu iskorišćeno je 100 publikacija izabranih metodom slučajnog izbora uz korišćenje alata *Research Randomizer*.

Na osnovu uzoračkog okvira publikovanih radova i prethodno definisane veličine uzorka, generisan je izbor literaturnih jedinica, odnosno lista rednih brojeva numerisanih referenci. Po završenom pretraživanju bibliografske baze podataka pristupilo se lociranju i prikupljanju radova u celini (*in extenso*). Prikupljeno je svih 100 radova od kojih je pet predstavljalo opise budućih studija ili pilot projekata pa su isključeni iz dalje analize.

Dva istraživača su nezavisno ekstrahovala podatke koristeći unapred pripremljen i dizajniran obrazac za prikupljanje podataka. U slučaju neslaganja istraživači su ponovo pregledali sporne publikacije da bi neslaganja razrešili konsenzusom.

was performed using the keyword "neurology" in the title and abstract, while the design of the searched literature was limited to observational studies: neurology[Title/Abstract] with the application of the "observational study" filter.

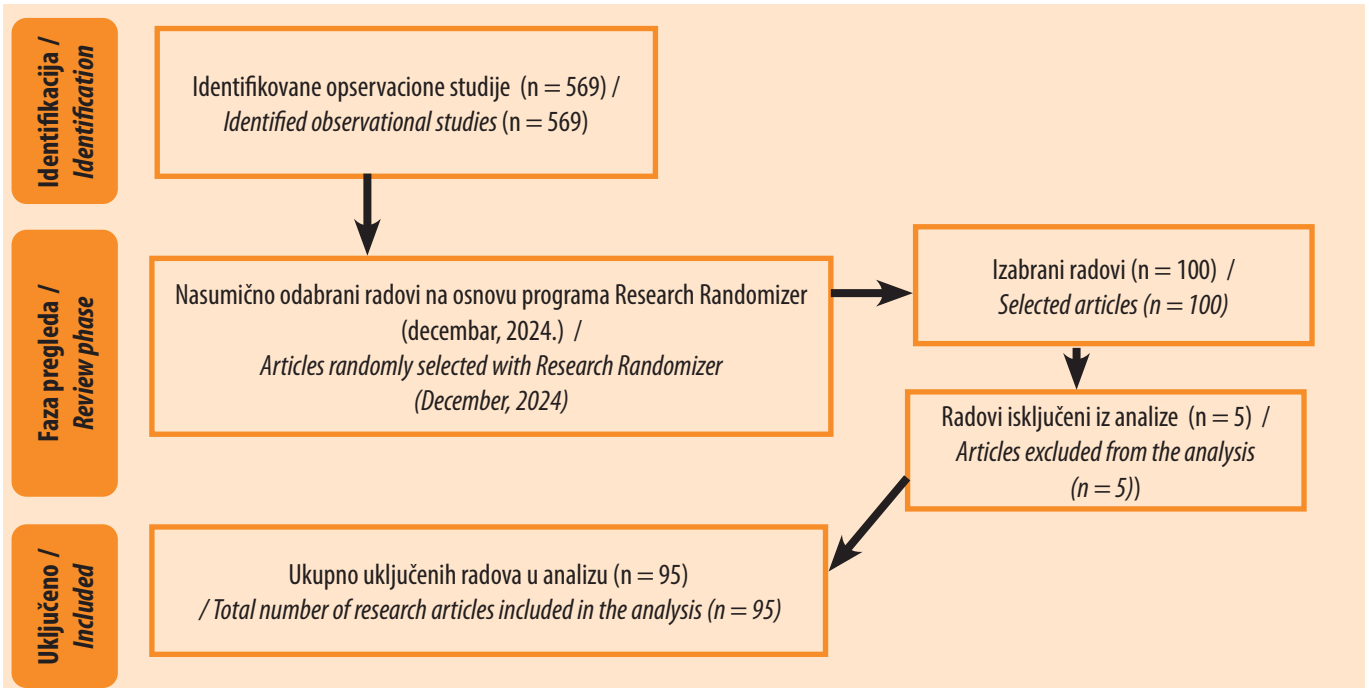
To ensure a comprehensive approach to literature, the search process did not include language or publication date restrictions.

The exclusion criteria for the analysis were as follows: animal or *in vitro* studies, literature reviews, editorials, descriptions of future studies, or descriptions of pilot projects.

The literature search identified 569 articles published between 2013 and 2025. A list with numbered references, arranged from the earliest to the most recently published articles, was downloaded. Out of the 569 papers, 100 publications were randomly selected for further analysis using the *Research Randomizer* tool.

Based on the sampling frame of published articles and the predefined sample size, a selection of literature units (i.e., a list of serial numbers for the referenced publications) was generated. Upon completion of the database search, the articles were located and collected in full (*in extenso*). All 100 papers were collected; however, five were descriptions of future studies or pilot projects and were therefore excluded from further analysis.

Two researchers independently extracted data using a pre-prepared and pre-designed data collection form. In the event of disagreement, the researchers re-examined the disputed publications to resolve discrepancies by consensus.



Slika 1. PRISMA dijagram odabranih radova uključenih u analizu primene statističkih metoda u neurološkim istraživanjima

Graph 1. PRISMA diagram of selected articles included in the analysis of the application of statistical methods in neurological research.

Statistička analiza

U ovom istraživanju, podaci su prikazani deskriptivnim statističkim metodama i analizirani odgovarajućim statističkim testovima. Od deskriptivnih metoda korišćeni su apsolutni i relativni brojevi. Za poređenje kategorijalnih podataka između dve grupe primenjen je hi-kvadratni test ($\chi 2$ test) ili Fišerov test i Man-Vitnijev test.

Svi testovi su bili dvostrani, a nivo statističke značajnosti je bio postavljen na $p < 0,05$. Analiza je izvršena korišćenjem statističkog softvera *SPSS Statistics v. 22* (*IBM Corp., Armonk, NY*).

Statistical Analysis

In this study, data were presented using descriptive statistical methods and analyzed using appropriate statistical tests. Absolute and relative frequencies were used as descriptive methods. The chi-square test ($\chi 2$ test) or Fisher's exact test and the Mann-Whitney test were applied to compare categorical data between two groups.

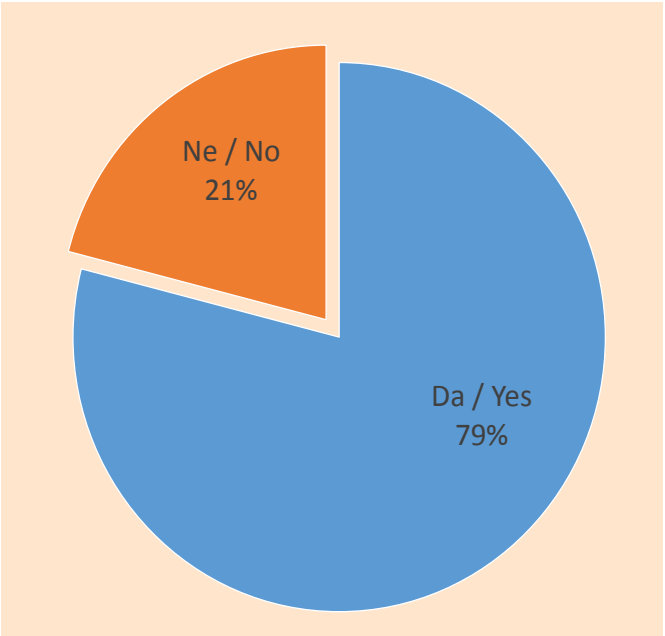
All tests were two-sided, and the level of statistical significance was set at $p < 0.05$. The analysis was performed using SPSS Statistics v. 22 (IBM Corp., Armonk, NY) statistical software.

REZULTATI

Studijom je obuhvaćeno 95 od 569 opservacionih studija iz oblasti neurologije publikovanih u periodu od 2013. do 2024. godine. U metodologiji publikovanih studija, navedeno je da je nešto više od polovine studija (51; 53,7%) imalo retrospektivni dizajn, dok su 44 (46,3%) studije imale prospektivni dizajn. Vrsta studije je jasno navedena u 54 naučno-istraživačka rada, od toga su trećina bile kohortne studije, njih 20 (21,1%) su bile studije preseka, dok su dve (2,1%) studije bile slučaj-kontrole (Tabela 1).

Dozvola Etičke komisije za izvođenje istraživanja u opservacionim studijama iz oblasti neurologije je prikazana u 75 (78,9%) studija (Grafikon 1).

Normalnost raspodele podataka je bila prikazana u 28 (29,5%) studija, dok u 67 (70,5%) radova nije navedena. Od 28 radova gde je navedena normalnost raspodele podataka, test za procenu normalnosti ras-



Grafikon 1. Prikazivanje dozvole etičke komisije u opservacionim studijama iz oblasti neurologije

Graph 1. Reporting approval by an ethics committee in observational studies in neurology

Tabela 1. Tip i vrsta studije, navođenje kriterijuma za uključivanje i isključivanje u studiju, opis podataka i analize, prikaz p-vrednosti i korišćenog softvera u opservacionim studijama iz oblasti neurologije

Table 1. Type of study and study design, presentation of inclusion and exclusion criteria, description of data and analysis, and presentation of p-values and the software used in observational studies in neurology

Parametri / *Parameters*	**Broj / *Number* (%)**
Tip studije / *Study type*	
Retrospektivna / *Retrospective*	51 (53.7%)
Prospektivna / *Prospective*	44 (46.3%)
Vrsta studije / *Study design*	
Studija preseka / *Cross-sectional study*	20 (21.1%)
Slučaj-kontrola / *Case-control study*	2 (2.1%)
Kohortna studija / *Cohort study*	32 (33.7%)
Kriterijumi za uključivanje / *Inclusion criteria*	
Ne / *No*	19 (20.0%)
Da / *Yes*	76 (80.0%)
Kriterijumi za isključivanje / *Exclusion criteria*	
Ne / *No*	37 (38.9%)
Da / *Yes*	58 (61.1%)
Opis podataka / *Data description*	
Ne / *No*	5 (5.3%)
Da / *Yes*	90 (94.7%)
Opis analize / *Analysis description*	
Ne / *No*	23 (24.2%)
Da / *Yes*	72 (75.8%)
p-vrednost / *p value*	
Ne / *No*	29 (30.5%)
Da / *Yes*	66 (69.5%)
Softver / *Software*	
Ne / *No*	24 (25.3%)
Da / *Yes*	71 (74.7%)

podele podataka nije bio opisan u 9 (9,5%) studija. Od onih koje su navele primenjeni test, najčešće je korišćen Kolmogorov-Smirnovljev test, odnosno u 12 (12,6%) studija, a sledi ga Šapiro-Vilkov test, koji je naveden u 7 (7,4%) studija (Tabela 2).

Tabela 2. Prikaz normalnosti raspodele i testa za procenu normalnosti raspodele u opservacionim studijama iz oblasti neurologije

Table 2. Reporting data distribution normality and the normality test used in observational studies in neurology.

Parametri / *Parameters*	**Broj / *Number (%)***
Prikaz normalnosti raspodele / *Presentation of distribution normality*	
Ne / *No*	67 (70.5%)
Da / *Yes*	28 (29.5%)
Test za normalnost raspodele / *Distribution normality test*	
Nije opisan / *Not presented*	9 (9.5%)
Kolmogorov-Smirnovljev test *Kolmogorov-Smirnov test*	12 (12.6%)
Šapiro-Vilkov test / *Shapiro-Wilk test*	7 (7.4%)

U većini studija (48; 67,4%), rezultati su predstavljeni primenom kombinovanog prikaza, kroz tabele i grafikone, dok je 39 (41,1%) studija koristilo isključivo tabelarni prikaz. Od ukupnog broja radova, u 89 (93,7%) su jasno navedeni nazivi tabela i grafikona, dok je pravilno prikazivanje osnovnih statističkih parametara (srednja vrednost, standardna devijacija, medijana) zabeleženo u 75 (78,9%) radova.

U analiziranim studijama, *t*-test je korišćen u 35 (36,8%) radova, dok kod preostalih 60 (63,2%) nije primenjivan. Man-Vitnijev test je korišćen u 32 (33,7%), a nije korišćen u 63 (66,3%) studija. Vilkoksonov test je najmanje korišćen – primenjen je u svega 5 (5,3%) studija. Hi-kvadratni test je primenjen u 44 (46,3%) studije, dok u 51 (53,7%) studiji nije korišćen. Fišerov test je korišćen u 35 (36,8%) studija, dok u 60 (63,2%) radova nije bio primenjen. MekNemarov test je najmanje korišćen i primenjen je u samo dve (2,1%) studije. Statističke metode za procenu razlike između tri i više nezavisnih i zavisnih uzoraka su korišćene sa manjom učestalošću. Kada su u pitanju neparametarski testovi, Fridmanov test je primenjen u jednoj (1,1%) studiji a Kruskal-Volisov test je korišćen u 5 (5,3%) studija. Što se tiče parametarskih testova, jednofaktorska *ANOVA* je korišćena u 7 (7,4%) studija dok je *ANOVA* ponovljenih merenja korišćena u dve (2,1%) studije.

RESULTS

The present study included 95 of the 569 observational studies in neurology published between 2013 and 2024. The methodology of the published studies showed that slightly more than half of the studies (51; 53.7%) had a retrospective design, while 44 (46.3%) studies had a prospective design. The type of study was clearly stated in 54 research papers; of these, one-third were cohort studies, 20 (21.1%) were cross-sectional studies, and two (2.1%) were case-control studies (Table 1).

Approval by an ethics committee to conduct research in observational studies in neurology was reported in 75 (78.9%) of the analyzed research articles (Graph 1).

The normality of data distribution was reported in 28 (29.5%) studies, while it was not stated in 67 (70.5%) articles. Of the 28 articles where the normality of data distribution was mentioned, the test used to assess normality was not described in 9 (9.5%) studies. Among those that specified the applied test, the Kolmogorov-Smirnov test was the most frequently used, appearing in 12 (12.6%) studies, followed by the Shapiro-Wilk test, which was cited in 7 (7.4%) studies (Table 2).

In most of the analyzed studies (48; 67.4%), results were presented using a combination of tables and graphs, while 39 (41.1%) studies used exclusively tabular presentation. Of the total number of articles, 89 (93.7%) clearly stated the titles of tables and graphs, while the correct presentation of basic statistical parameters (mean, standard deviation, median) was noted in 75 (78.9%) papers.

In the analyzed studies, the *t*-test was used in 35 (36.8%) articles, while it was not applied in the remaining 60 (63.2%). The Mann-Whitney test was used in 32 (33.7%) studies and was not used in 63 (66.3%). The Wilcoxon test was the least utilized, being applied in only 5 (5.3%) studies. The chi-square test was applied in 44 (46.3%) studies, while it was not used in 51 (53.7%) studies. Fisher's exact test was used in 35 (36.8%) studies and was not applied in 60 (63.2%) articles. McNemar's test was the least frequently used, applied in only two (2.1%) studies. Statistical methods for assessing differences between three or more independent and dependent samples were used with a lower frequency. Regarding non-parametric tests, the Friedman test was applied in one (1.1%) study, and the Kruskal-Wallis test was used in 5 (5.3%) studies. As for parametric tests, one-way ANOVA was used in 7 (7.4%) studies, while repeated measures ANOVA was used in two (2.1%) studies. Correlation analysis was applied in 14 (14.7%) studies, while it was not used in 81 (85.3%) studies. Of the

Tabela 3. Učestalost korišćenja statističkih testova u opservacionim studijama iz oblasti neurologije

Table 3. Frequency of statistical test usage in observational studies in neurology

Primenjeni testovi / *Applied tests*	**Broj / *Number* (%)**
t-test /*t-test*	35 (36.8%)
Man-Vitnijev test / *Mann-Whitney test*	32 (33.7%)
Vilkoksonov test / *Wilcoxon test*	5 (5.3%)
Hi-kvadrat test / *Chi-square test*	44 (46.3%)
Fišerov test / *Fisher's test*	35 (36.8%)
McNemar's test	2 (2.1%)
Fridmanov test / *Friedman test*	1 (1.1%)
Kruskal-Volisov test / *Kruskal-Wallis test*	5 (5.3%)
Jednofaktorska analiza varijanse (*ANOVA*) / *One-way analysis of variance (ANOVA)*	7 (7.4%)
Korelacija /*Correlation*	14 (14.7%)
Tip korelacije / *Type of correlation*	
• Spirmanova / *Spearman*	6 (6.3 %)
• Pirsonova / *Pearson*	8 (8.4 %)
Regresiona analiza / *Regression analysis*	32 (33.7%)
• Logistička / *Logistic*	18 (18.9%)
• Linearna / *Linear*	8 (8.4%)
• Koksov regresioni model / *Cox regression model*	1 (1.1%)
• Poasonova / *Poisson*	2 (2.1%)
• Navedeno / *Not reported*	3 (3.2%)

Tabela 4. Određivanje normalnosti raspodele u odnosu na korišćenje statističkih testova u opservacionim studijama iz oblasti neurologije

Table 4. Determining the normality of distribution in relation to the use of statistical tests in observational studies in neurology

Primenjen test / *Applied test*	Procena normalnosti raspodele podataka / *Assessment of the normality of data distribution*		*p*-vrednost / *p value*
	No	Yes	
	67 (70.5%)	28 (29.5%)	
t-test / *t-test*	15 (22.4%)	20 (71.4%)	< 0.001*
Man-Vitnijev test / *Man-Whitney test*	15 (22.4%)	17 (60.7%)	< 0.001*
Vilkoksonov test / *Wilcoxon test*	1 (1.5%%)	4 (14.3%)	0.011*

Podaci su prikazani kao broj (%), *statistički značajna razlika /
*Data are presented as numbers (%), *statistically significant difference*

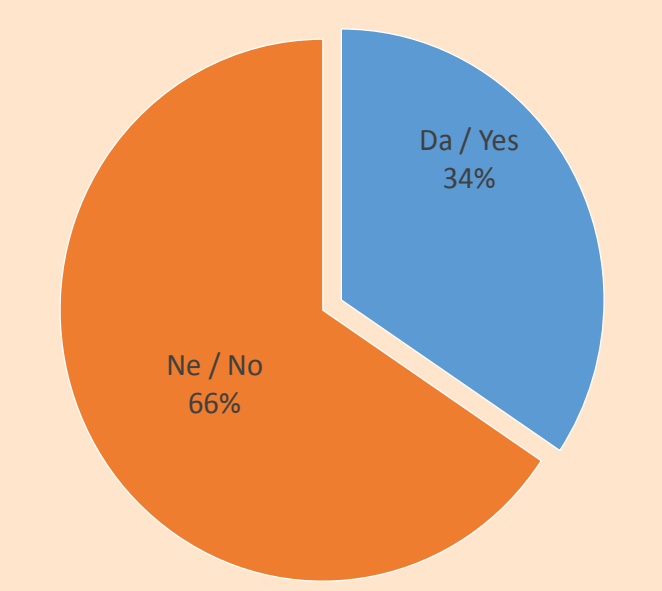
correlation analyses, Spearman's correlation was used in 6 (6.3%) articles, and Pearson's correlation in 8 (8.4%) studies. Regression analysis was the method of choice for data analysis in 32 (33.7%) studies (Table 3).

Sample size planning was clearly described in 10 studies (10.5%), while the remaining 85 (89.5%) articles lacked information regarding a prior assessment of the required number of participants (Graph 2). Nevertheless, sample size as a descriptive parameter was reported in 91 (95.8%) studies, while this information was missing in only 4 (4.2%) articles.

The frequency of *t*-test application in studies that previously reported an assessment of data distribution normality is 71.4%, while the *t*-test was also applied in 22.4% of studies that did not report an assessment of data distribution normality. The application of the *t*-test is statistically significantly more frequent in studies that report the normality of data distribution ($p < 0.001$).

The Mann-Whitney test was applied in 60.7% of studies that reported an assessment of data distribution normality, while in articles where authors did not report an assessment of data distribution, the Mann-Whitney test was used for analysis in 53.1% of studies. The application of the Mann-Whitney test is statistically significantly more frequent in studies that report the normality of data distribution ($\chi 2 = 13.0$, $p < 0.001$).

The application of the Wilcoxon test is statistically significantly more frequent in studies that reported an assessment of data distribution normality ($p = 0.011$), (Table 4).



Grafikon 2. Određena veličina uzorka u opservacionim studijama iz oblasti neurologije

Graph 2. Defining sample size in observational studies in neurology

Tabela 5. Veličina uzorka prema primenjenom testu u opservacionim studijama iz oblasti neurologije

Table 5. Sample size according to the test applied in observational studies in neurology

Test / *Test*	Primenjen /*Applied*	Broj /*Number*	Medijana /*Median*	Min-Maks /*Min-Max*	p
Hi-kvadratni test / *Chi-square test*	Da / *Yes*	44	262.0	(29–2,749,872)	0.003*
	Ne / *No*	49	93.0	(4–3,400,000)	
Fišerov test / *Fisher's test*	Da / *Yes*	35	138.0	(23–7,969)	0.837
	Ne / *No*	58	117.5	(4–3,400,000)	

*statistički značajna razlika / **statistically significant difference*

Korelacija je primenjena u 14 (14,7%) studija, dok u 81 (85,3%) studiji nije korišćena. Od korelacionih analiza, Spirmanova korelacija je korišćena u 6 (6,3%) radova, a Pirsonova korelacija u 8 (8,4%) studija. Regresiona analiza je bila metoda izbora za analizu podataka u 32 (33,7%) studije (Tabela 3).

Planiranje veličine uzorka bilo je jasno navedeno u 10 studija (10,5%), dok je u preostalih 85 (89,5%) radova izostala informacija o prethodno sprovedenoj proceni potrebnog broja ispitanika (Grafikon 2). Ipak, veličina uzorka kao deskriptivni parametar je bila prijavljena u 91 (95,8%) studiji, dok je u samo 4 (4,2%) rada ovaj podatak izostao.

Učestalost primene *t*-testa u studijama koje su prethodno prikazale procenu normalnosti distribucije podataka iznosi 71,4%, dok je *t*-test primenjen i kod 22,4% studija koje nemaju navedenu procenu normalnosti distribucije podataka. Primena *t*-testa je statistički značajno učestalija u studijama koje imaju prikazanu normalnost raspodele podataka ($p < 0,001$).

Man-Vitnijev test je primenjen kod 60,7% studija koje su imale prikazanu procenu normalnosti raspodele podataka, dok je u radovima gde autori nisu prikazali procenu distribucije podataka Man-Vitnijev test korišćen za analizu u 53,1% studija. Primena Man-Vitnijevog testa je statistički značajno učestalija kod studija koje imaju prikazanu normalnost raspodele podataka ($p < 0,001$).

Primena Vilkoksonovog testa je statistički značajno učestalija kod studija koje su prikazale procenu normalnosti distribucije podataka ($p = 0,011$), (Tabela 4).

Hi-kvadratni test je primenjen u 44 studije. U ovim studijama medijana veličine uzorka iznosila je 262,0 (opseg: 29–2.749.872), dok je u preostaloj 51 studiji, gde test nije primenjen, medijana bila 93,0 (opseg: 4–3.400.000). Postoji statistički značajna razlika u veličini uzorka između studija u kojima je primenjen hi-kvadratni test i onih u kojima nije. Studije u kojima je test primenjen imale su značajno veće uzorke u poređenju sa studijama u kojima test nije primenjen ($p = 0,003$). Fišerov test tačne verovatnoće primenjen je u 35 studija, u kojima je medijana veličine uzorka iznosila 138,0 (opseg: 23–7.969),

The chi-square test was applied in 44 studies. In these studies, the median sample size was 262.0 (range: 29–2,749,872), while in the remaining 51 studies where the test was not applied, the median was 93.0 (range: 4–3,400,000). There is a statistically significant difference in sample size between studies wherein the chi-square test was applied and those wherein it was not. Studies wherein the test was applied had significantly larger samples compared to studies wherein the test was not applied ($p = 0.003$). Fisher's exact test was applied in 35 studies, wherein the median sample size was 138.0 (range: 23–7,969), while in 58 studies where the test was not applied, the median sample size was 117.5 (range: 4–3,400,000). There is no statistically significant difference in sample size between these two groups of studies ($p = 0.837$), (Table 5).

DISCUSSION

In the available literature, we have not found a study that systematically analyzes the application of statistical and methodological techniques in published observational articles in the field of neurology. This study, therefore, represents a unique attempt to critically evaluate the quality and frequency of the application of statistical methods in this field of medicine through a review of relevant scientific literature, which may contribute to raising methodological standards in future research.

Our evaluation included 95 out of 569 identified observational studies in neurology. The distribution of designs showed an almost equal representation of retrospective and prospective studies. A similar structure can be found in a recently published meta-analysis that included 126 studies on neurological manifestations following COVID-19 infection. The authors of this meta-analysis included 48 (38.1%) prospective cohort studies and 42 (33.3%) retrospective cohort studies, which also indicates an almost equal application of both approaches in the analysis of neurological outcomes [6].

dok je u 58 studija u kojima test nije primenjen medijana uzorka bila 117,5 (opseg: 4–3.400.000). Ne postoji statistički značajna razlika u veličini uzorka između ove dve grupe studija (U = 989,0; p = 0,837), (Tabela 5).

DISKUSIJA

U dostupnoj literaturi nismo pronašli studiju koja se na sistematičan način bavi analizom primene statističkih i metodoloških tehnika u objavljenim opservacionim radovima iz oblasti neurologije. Ovo istraživanje, stoga, predstavlja jedinstveni pokušaj da se kroz pregled relevantne naučne literature kritički sagleda kvalitet i učestalost primene statističkih metoda u ovoj oblasti medicine, što može da doprinese podizanju metodoloških standarda u budućim istraživanjima.

Evaluacijom je obuhvaćeno 95 od 569 naučnoistraživačkih radova iz oblasti neurologije, čije je istraživanje sprovedeno po tipu opservacionih studija. Raspodela dizajna je pokazala skoro jednaku zastupljenost retrospektivnih i prospektivnih studija. Sličnu strukturu nalazimo i u nedavno objavljenoj meta-analizi koja je obuhvatila 126 studija o neurološkim manifestacijama nakon KOVID-19 infekcije. Autori ove meta-analize su uključili 48 (38,1 %) prospektivnih kohortnih studija i 42 (33,3 %) retrospektivne kohortne studije, što takođe ukazuje na gotovo ujednačenu primenu oba pristupa u analizi neuroloških ishoda [6].

U našem istraživanju, etičko odobrenje je bilo navedeno u 75 (78,9%) od 95 analiziranih radova, što ukazuje na relativno visok nivo svesti o značaju etičkih standarda u biomedicinskim istraživanjima. Ovaj rezultat je ohrabrujući i u skladu je sa globalnim trendom jačanja etičkog okvira, ali ipak zaostaje za nekim drugim oblastima. U sistematskoj analizi objavljenoj u oblasti nege, etičko odobrenje je bilo navedeno u čak 93,7% studija. Još značajnije, 95,1% navedenih odobrenja je bilo dobijeno pre početka studije, što ukazuje na visok nivo metodološke i etičke discipline. Takođe, prospektivne studije u tom radu imale su nešto veću učestalost izveštavanja u odnosu na retrospektivne (94,6% prema 82,8%) [7].

Prospektivne studije se planiraju za buduće sprovođenje i zato je preciziranje metodoloških principa jasnije i pravilnije postavljeno, pa tako i traženje etičke dozvole. Sličan rezultat je zabeležen i u analizi studija o fizikalnoj terapiji nakon moždanog udara, gde je etičko odobrenje bilo navedeno u 81,2% radova [8]. Isto navode i Ng i saradnici [9]. U našem istraživanju, kriterijumi za uključivanje u istraživanje su jasno definisani u 80% radova, što ukazuje na relativno dobru metodološku praksu i usklađenost sa važećim smernicama za dizajn kliničkih studija. Međutim, kriterijumi za isključivanje iz istraživanja su navedeni u 61% slučajeva, što ostavlja prostor za unapređenje transparentnosti u izveštavanju.

In our study, ethics committee approval was reported in 75 (78.9%) of the 95 analyzed articles, which indicates a relatively high level of awareness regarding the importance of ethical standards in biomedical research. This result is encouraging and consistent with the global trend of ethical framework strengthening, yet it still falls behind some other fields. In a systematic analysis published in the field of nursing, ethical approval was reported in as many as 93.7% of studies. Even more significantly, 95.1% of these approvals were obtained before the commencement of the study, which indicates a high level of methodological and ethical discipline. Also, prospective studies in this article had a slightly higher reporting frequency, as compared to the retrospective studies (94.6% versus 82.8%) [7].

Prospective studies are planned to be conducted in the future, and therefore, the specification of methodological principles is clearer and more properly established, as is the application for ethical approval. A similar result was found in an analysis of studies on physical therapy following a stroke, where ethical approval was reported in 81.2% of articles [8]. The same has been reported by Ng et al. [9]. In our study, inclusion criteria were clearly defined in 80% of the articles, indicating rather sound methodological practice and compliance with current guidelines for clinical study design. However, exclusion criteria were stated in 61% of cases, which leaves room for improvement in reporting transparency.

In a systematic review by Miranda et al. [10], which analyzed 40 randomized controlled trials on pressure ulcer treatment, only 12.5% of papers fully described all inclusion and exclusion criteria. This result falls significantly short of the results in our study, which indicates potential progress in neurological literature when it comes to the precision of methodological description [10]. Transparent reporting on inclusion criteria and the sample selection procedure is considered an important element of the methodological quality of research, as it enables the assessment of bias and a more reliable interpretation of results [11].

In our study, the sample size was determined in advance and clearly stated in only 10 out of the 95 analyzed articles (10.5%). Button et al. [12] particularly emphasize that insufficient statistical power represents one of the greatest challenges in neurosciences, as it leads to the overestimation of effects, reduced reproducibility, and an increased risk of false-positive findings. A study by Lee et al. [13], which included 451 randomized controlled trials (RCTs), states that the sample size was determined before the commencement of research in 58.1% of the articles, which represents a sig-

U sistematskom pregledu Mirande i saradnika [10], u kojem je analizirao 40 randomizovanih kontrolisanih studija u oblasti lečenja dekubitusa, samo 12,5% radova je u potpunosti opisalo sve kriterijume za uključivanje i isključivanje u istraživanje. Taj rezultat značajno je niži u odnosu na rezultate našeg rada, što ukazuje na potencijalni napredak u neurološkoj literaturi kada je reč o preciznosti metodološkog opisa [10]. Transparentno izveštavanje o kriterijumima uključivanja ispitanika i postupku selekcije uzorka smatraju se važnim elementom metodološkog kvaliteta istraživanja, jer omogućava procenu pristrasnosti i pouzdaniju interpretaciju rezultata [11].

U našem istraživanju, veličina uzorka je unapred bila određena i jasno navedena u samo 10 od 95 analiziranih radova (10,5%). Baton i saradnici [12] posebno naglašavaju da nedovoljna statistička snaga predstavlja jedan od najvećih izazova u neuronaukama, jer dovodi do precenjivanja efekata, smanjenja reproduktivnosti i povećanog rizika od lažno pozitivnih nalaza. Istraživanje Lija i saradnika [13], koje je obuhvatilo 451 randomizovanu kontrolisanu studiju (engl. *randomized controlled trial – RCT)*, navodi da je u 58,1% studija pre početka istraživanja određena veličina uzorka, što predstavlja znatno viši procenat u odnosu na naš rezultat [13]. Razlika u dizajnu studija delimično može objasniti ovaj disbalans, s obzirom na to da se za randomizovane kontrolisane studije češće zahtevaju stroži metodološki standardi, uključujući formalnu procenu veličine uzorka. Sa druge strane, veličina uzorka, kao deskriptivni parametar, prijavljena je u 91 (95,8%) studiji, što ukazuje da autori ipak prepoznaju značaj predstavljanja osnovnih statističkih karakteristika uzorka, iako izostaje formalna procena adekvatnosti uzorka u fazi planiranja. U 72 (75,8%) rada je uključen opis statističke analize, dok je u 71 (74,7%) radu naveden softver koji je korišćen za statističku obradu podataka.

Ovi rezultati ukazuju na zadovoljavajući nivo transparentnosti u odnosu na prethodne studije. Na primer, u jednom istraživanju o kvalitetu eksperimentalnog dizajna i statističke analize u radovima koji koriste životinje, Kilekni i saradnici [14] su otkrili da je 74% radova navelo opis statističkih analiza, ali je takođe primećeno da su mnogi radovi imali nedostataka u pružanju detaljnih podataka o korišćenim statističkim procedurama i softverskim alatima.

Blejkli i saradnici [15] u svom istraživanju navode da 33% radova iz oblasti otorinolaringologije nije navodilo softver koji je korišćen. U našoj analizi radova iz oblasti neurologije, testiranje normalnosti raspodele izvršeno je u svega 29,5% studija, što ukazuje na nedovoljnu primenu ove ključne statističke provere. Ovaj rezultat je u skladu sa rezultatima Hoekstre i saradnika

nificantly higher percentage than our result [13]. The difference in study design can partially explain this imbalance, given that stricter methodological standards are more often required for randomized controlled trials, including a formal sample size assessment. On the other hand, sample size as a descriptive parameter was reported in 91 (95.8%) studies, which indicates that authors do recognize the importance of presenting the basic statistical characteristics of the sample, even though a formal assessment of sample adequacy in the planning phase is lacking. In our analysis, 72 (75.8%) articles included a description of statistical analysis, while 71 (74.7%) articles stated the software used for statistical data processing.

These results indicate a satisfactory level of transparency compared to previous studies. For example, in a study on the quality of experimental design and statistical analysis in articles involving animal subjects, Kilkenny et al. [14] found that 74% of articles included a description of statistical analyses. However, it was also noted that many studies displayed shortcomings in providing detailed data on the statistical procedures and software tools used.

In their study, Blakley et al. [15] found that 33% of articles in otorhinolaryngology did not report the software used. In our analysis of articles in neurology, normality testing was performed in only 29.5% of studies, which indicates insufficient application of this key statistical test. This result is consistent with the results of Hoekstra et al. [16], who found that only about one-third of papers precisely check the assumptions of frequently used statistical methods. A similar problem was observed in orthopedic research. Namely, Parsons et al. [17] state that about 30% of papers reported checking for normality, while in other cases, parametric statistics were used without prior testing, which can lead to methodological errors and inaccurate conclusions.

Ghasemi and Zahediasl [18] highlight the importance of normality testing as a prerequisite for selecting appropriate statistical tests, warning that neglecting this test, especially in studies with small samples, can lead to incorrect interpretation of results. Blakley et al. [15] have also confirmed that a similar practice of neglecting normality testing is observed in other medical fields, such as otorhinolaryngology, where it can compromise the reliability of statistical conclusions.

In our analysis of neurological observational studies, 69.5% of articles clearly displayed *p*-values, which indicates a high level of statistical transparency and appropriate application of the significance level. This percentage is significantly higher than the frequency recorded in the study by To and Jukes [19], whose anal-

[16], koji su u svojoj studiji ustanovili da samo oko trećine radova precizno proverava pretpostavke često korišćenih statističkih metoda. Sličan problem je uočen i u ortopedskim istraživanjima. Naime, Parsons i saradnici [17] navode da je oko 30% radova navelo proveru normalnosti, dok je u ostalim slučajevima korišćena parametarska statistika bez prethodne provere, što može dovesti do metodoloških grešaka i netačnih zaključaka.

Gasemi i Zahediasl [18] ističu značaj testiranja normalnosti kao preduslova za izbor odgovarajućih statističkih testova, upozoravajući na to da zanemarivanje ove provere, naročito u studijama sa manjim uzorcima, može da dovede do pogrešne interpretacije rezultata. Blejkli i saradnici [15] potvrđuju da je slična praksa zanemarivanja provere normalnosti raspodele uočena i u drugim medicinskim oblastima, kao što je otorinolaringologija, gde to može ugroziti pouzdanost statističkih zaključaka.

U analizi neuroloških opservacionih studija koju smo sproveli, u 69,5% radova jasno su prikazane *p*-vrednosti, što ukazuje na visok nivo statističke transparentnosti i adekvatne primene nivoa značajnosti. Ovaj procenat je značajno veći od učestalosti zabeležene u studiji čiji su autori To i Džuks [19], koji su u svojoj analizi apstrakata neurohirurških radova utvrdili da je samo 23% studija sadržalo jasno prikazane *p*-vrednosti. S druge strane, studija Ginsela i saradnika [20] ukazuje da je čak 82% prijavljenih *p*-vrednosti u analiziranim medicinskim radovima bilo statistički značajno, što može sugerisati postojanje selektivnog izveštavanja rezultata.

U analiziranim radovima iz oblasti neurologije, 67,4% je prikazalo rezultate u vidu tabela ili grafikona, dok je pravilno i dosledno obeležavanje tabela i grafikona zabeleženo u 93,7% slučajeva. Slične rezultate u vezi sa prikazom rezultata zabeležili su Blejkli i saradnici [15] koji su u uzorku radova iz otorinolaringologije ustanovili da je oko 70% radova adekvatno prikazivalo i opisivalo tabele i grafikone, što potvrđuje opšti trend u medicinskoj literaturi. Kilkejni i saradnici [14] u studiji pregleda kvaliteta izveštavanja u eksperimentalnim studijama na životinjama navode da između 60% i 75% radova prikazuje rezultate na jasan i dosledan način, što je u skladu sa našim rezultatima.

U analizi korišćenja statističkih testova u ispitivanim radovima, uočava se da je *t*-test bio najčešće primenjivani parametarski test, korišćen u 36,8% radova. Man-Vitnijev test, kao neparametarski ekvivalent, primenjen je u 33,7% studija, dok je Vilkoksonov test korišćen u manjem broju radova (5,3%). Među drugim testovima, hi-kvadratni i Fišerov test su korišćeni u 46,3%, odnosno 36,8% radova, dok je MekNemarov test bio najređe primenjivan (2,1%). Ovi rezultati pokazuju različit pristup izboru statističkih metoda u analizi-

ysis of neurosurgical abstracts found that only 23% of studies clearly displayed *p*-values. On the other hand, a study by Ginsel et al. [20] indicates that as many as 82% of reported *p*-values in analyzed medical articles were statistically significant, which may suggest selective reporting of results.

In the analyzed neurology articles, 67.4% presented results in tables or graphs, while correct and consistent labeling of tables and graphs was recorded in 93.7% of cases. Similar findings regarding the presentation of results were recorded by Blakley et al. [15]. In a sample of otorhinolaryngology articles, they found that about 70% of studies adequately presented and described tables and graphs, confirming a general trend in medical literature. In a study reviewing the quality of reporting in experimental studies on animals, Kilkenny et al. [14] report that between 60% and 75% of articles present results clearly and consistently, which is in keeping with our results.

In the analysis of the use of statistical tests in the examined articles, it has been observed that the *t*-test was the most frequently applied parametric test, used in 36.8% of the studies. The Mann-Whitney test, as a non-parametric equivalent, was applied in 33.7% of studies, while the Wilcoxon test was used in a smaller number of articles (5.3%). Among other tests, the chi-square and Fisher's tests were used in 46.3% and 36.8% of papers, respectively, while McNemar's test was the least frequently applied (2.1%). These results show a different approach to the selection of statistical methods in the analyzed papers. Similar trends in the application of statistical methods have been reported by Parsons et al. [17], who determined that the *t*-test was used in about 40% of papers, the Mann-Whitney test in about 30%, and the chi-square test in 45%–50% of studies. These data indicate that both parametric and non-parametric tests have significant applications, but that there is a significantly higher application of tests for independent samples compared to tests for testing dependent samples. This can also be explained by the nature of the disease, where many medical emergencies do not allow for research to develop over a longer time frame, where repeated measurements would be performed.

In this study, the association between reporting the assessment of data distribution normality and the choice of statistical test was analyzed. Although at first glance it may be concluded that many articles (70.5%) did not report normality testing, a more detailed analysis reveals that in cases where normality was reported, the choice of test shows a conscious adjustment of the analysis to the data, which indicates responsible statistical practice. The *t*-test was applied in 35 articles, of

ranim radovima. Slične trendove u primeni statističkih metoda navode Parsons i saradnici [17], koji su utvrdili da je *t*-test korišćen u oko 40% radova, Man-Vitnijev test u oko 30%, a hi-kvadratni test u 45%–50% studija. Ovi podaci ukazuju da i parametarski i neparametarski testovi imaju značajnu primenu, ali značajno je učestalija primena testova za nezavisne uzorke u odnosu na testove za testiranje zavisnih uzoraka. To se može objasniti i prirodom bolesti, gde mnoga urgentna stanja ne dozvoljavaju da se razvije istraživanje u dužem vremenskom okviru gde bi se vršila ponovljena merenja.

U ovoj studiji analizirana je povezanost između prikazivanja procene normalnosti raspodele podataka i izbora statističkog testa. Iako se na prvi pogled može zaključiti da veliki broj radova (70,5%) nije prijavio proveru normalnosti raspodele, detaljnija analiza otkriva da u slučajevima kada je normalnost prikazana, izbor testa pokazuje svesno prilagođavanje analize podacima, što ukazuje na odgovornu statističku praksu. Primena *t*-testa zabeležena je u 35 radova, od toga u 71,4% studija sa navedenom procenom normalne distribucije podataka. Upotreba neparametrijskih testova je takođe bila povezana sa prisustvom procene raspodele. Man-Vitnijev test je korišćen u 60,7% studija sa navedenom procenom normalnosti i u 22,4% studija gde ona nije bila navedena, dok je Vilkoksonov test bio manje zastupljen, ali prisutan uglavnom u radovima sa navedenom procenom raspodele.

U poređenju sa ovim rezultatima, studija Kumara i saradnika [21], koja je analizirala 300 radova objavljenih u vodećim biomedicinskim časopisima, utvrdila je da je test normalne distribucije podataka bio jasno naveden u manje od 1,5% slučajeva, iako je *t*-test i dalje bio među najčešće primenjivanim analizama. Ovo ukazuje da je u širem biomedicinskom kontekstu, kao i kod nas, *t*-test često primenjivan bez transparentne dokumentacije o pretpostavkama, ali naša analiza pokazuje bolji nivo izveštavanja i primene testova u skladu sa karakteristikama podataka. Rezultati studije Jina i saradnika [22], koja je analizirala 2.913 kineskih medicinskih radova, pokazuju da je više od polovine radova sadržalo statističke greške, uključujući nepravilnu primenu *t*-testa.

Naša studija pokazuje da, iako nije uvek navedena procena normalnosti, izbor statističkih metoda u neurološkim opservacionim studijama uglavnom prati osnovne statističke pretpostavke i daje nadu u kontinuirano unapređenje metodološkog pristupa u ovoj oblasti.

Rezultati naše analize pokazuju da je primena hi-kvadratnog i Fišerovog testa u opservacionim studijama iz oblasti neurologije u velikoj meri u skladu sa teorijskim preporukama ali i da postoje određena odstupanja.

which 71.4% were studies with a reported assessment of normal data distribution. The use of non-parametric tests was also associated with the presence of a distribution assessment. The Mann-Whitney test was used in 60.7% of studies with a reported normality assessment, as well as in 22.4% of studies where it was not reported, while the Wilcoxon test was less represented, but was present mainly in papers with a reported distribution assessment.

Compared to these results, the study by Kumar et al. [21], which analyzed 300 papers published in leading biomedical journals, determined that the test for normal data distribution was clearly reported in less than 1.5% of cases, although the *t*-test remained among the most frequently applied analyses. This indicates that in the wider biomedical context, as was the case in our study, the *t*-test is often applied without transparent documentation of assumptions, but our analysis shows a better level of reporting and application of tests in accordance with data characteristics. The results of the study by Jin et al. [22], which analyzed 2,913 Chinese medical articles, show that more than half of the studies contained statistical errors, including incorrect application of the *t*-test.

Our study shows that, although the assessment of normality is not always reported, the choice of statistical methods in neurological observational studies generally follows basic statistical assumptions, indicating continuous improvement of the methodological approach in this field.

The results of our analysis show that the application of the chi-square and Fisher's tests in neurological observational studies is largely in keeping with theoretical recommendations, but also that certain deviations are present.

Studies included in our research that used the chi-square test for data analysis had a significantly larger sample, as compared to studies where the test was not used. This indicates that the chi-square test, as recommended by statistical rules, is more often used in studies with larger samples, where expected frequencies are large enough for the reliable application of this test.

A similar conclusion was presented by Campbell [23], who showed that the chi-square test can yield unreliable results when the expected frequencies are below 5, and that in such situations the application of Fisher's test is recommended. The author emphasizes that the choice between these two tests is conditioned not only by the total number of participants but also by the distribution of data in contingency tables, and recommends that Fisher's test be used whenever at least one expected frequency falls below 5, regardless of the total sample size.

Studije u okviru našeg istraživanja, koje su kao metod analize podataka koristile hi-kvadratni test, imale su značajno veći uzorak u odnosu na studije u kojima test nije korišćen. Ovo ukazuje na to da se hi-kvadratni test, kao što i statistička pravila preporučuju, češće koristi u studijama sa većim uzorcima, gde su očekivane frekvencije dovoljno velike za pouzdanu primenu ovog testa.

Sličan zaključak izneo je Kembel [23], koji je pokazao da hi-kvadratni test može da daje nepouzdane rezultate ako su očekivane frekvencije manje od 5, i da je u takvim situacijama preporučena primena Fišerovog testa. Autor naglašava da je izbor između ova dva testa uslovljen ne samo ukupnim brojem ispitanika, već i distribucijom podataka u tabelama kontigencije, te preporučuje da se Fišerov test koristi uvek kada najmanje jedna očekivana frekvencija padne ispod 5, bez obzira na ukupnu veličinu uzorka.

Fišerov test je u našoj studiji primenjen u 35 slučajeva. Iako razlika nije statistički značajna, uočava se da je Fišerov test bio metoda izbora kod studija sa manjom veličinom uzorka u poređenju sa hi-kvadratnim testom. Ovo je u skladu sa rezultatima Oliveire i saradnika [24], koji su zaključili da je Fišerov test konzervativniji u proceni značajnosti u odnosu na hi-kvadratni test, posebno u studijama sa manjim uzorcima. Autori su uporedili više statističkih testova (uključujući hi-kvadratni, Fišerov, Jejtsovu korekciju i druge) na simuliranim podacima sa malim uzorcima. Njihovi rezultati su pokazali da je Fišerov test davao najstabilnije *p*-vrednosti, dok je hi-kvadratni test u više od 20% slučajeva davao lažno značajne rezultate, posebno u tabelama sa disbalansiranim raspodelama. Autori su istakli da upotreba hi-kvadratnog testa bez procene očekivanih vrednosti predstavlja značajan metodološki propust.

Sera i saradnici [25] su sproveli simulacionu analizu i pokazali da hi-kvadratni test daje statistički značajan rezultat u većem procentu slučajeva (23,29%) u odnosu na Fišerov test (17,72%). Ova razlika ukazuje na to da hi-kvadratni test ima veću osetljivost ali i veću verovatnoću lažno pozitivnih rezultata u uslovima malih očekivanih frekvencija. Ovi rezultati potkrepljuju naš rezultat da se Fišerov test koristi u studijama sa manjim brojem ispitanika, jer je u takvim uslovima pouzdaniji.

Iako je u našoj analizi medijana uzorka u studijama koje su koristile Fišerov test relativno visoka nije moguće sa sigurnošću zaključiti kakve su bile očekivane frekvencije u tabelama. Moguće je da je Fišerov test bio primenjivan i u slučajevima malog broja događaja, uprkos velikom ukupnom uzorku. Ovo ukazuje na potrebu da se, prilikom izbora statističkog testa, ne oslanjamo samo na ukupnu veličinu uzorka, već i na konkretnu distribuciju podataka u okviru kategorija posmatranih obeležja.

Fisher's test was applied in 35 cases in our study. Although the difference is not statistically significant, it has been noted that Fisher's test was the method of choice in studies with a smaller sample size, as compared to the chi-square test. This is in accordance with the results of Oliveira et al. [24], who concluded that Fisher's test was more conservative in assessing significance compared to the chi-square test, especially in studies with smaller samples. The authors compared several statistical tests (including chi-square, Fisher's, Yates's correction, and others) on simulated data with small samples. Their results showed that Fisher's test gave the most stable *p*-values, while the chi-square test gave falsely significant results in more than 20% of cases, especially in tables with unbalanced distributions. The authors pointed out that the use of the chi-square test without assessment of expected values represented a significant methodological oversight.

Serra et al. [25] conducted a simulation analysis and showed that the chi-square test gave a statistically significant result in a higher percentage of cases (23.29%), as compared to Fisher's test (17.72%). This difference indicates that the chi-square test has higher sensitivity but also a higher probability of false-positive results in the case of small expected frequencies. These results support our finding that Fisher's test is used in studies with a smaller number of participants because it is more reliable in such conditions.

Although the median sample size in studies that used Fisher's test in our analysis is relatively high, it is not possible to conclude with certainty what the expected frequencies were in the tables. It is possible that Fisher's test was also applied in cases with a small number of events, despite the large total sample. This indicates the need, when choosing a statistical test, not to rely only on the total sample size, but also on the specific distribution of data within the categories of observed characteristics.

In essence, the results support the recommendations from the literature: the chi-square test is more appropriate for larger samples and large frequencies, while Fisher's test is recommended for small expected values.

Although the study provides significant insight into the application of statistical methods in observational neurological studies, it is necessary to highlight several limitations. First, the analysis was limited to articles indexed in the *PubMed* database, which potentially excludes relevant publications from other sources. Second, the analysis focuses only on observational studies, which limits the possibility of comparison with experimental research, which often uses different methodological approaches.

U suštini, dobijeni rezultati podržavaju preporuke iz literature: hi-kvadratni test je primereniji za veće uzorke i velike frekvencije, dok se Fišerov test preporučuje kod malih očekivanih vrednosti.

Iako studija pruža značajan uvid u primenu statističkih metoda u opservacionim neurološkim radovima, potrebno je istaći nekoliko ograničenja. Prvo, analiza je bila ograničena na radove indeksirane u bazi podataka *PubMed,* što potencijalno isključuje relevantne studije iz drugih izvora. Drugo, istraživanje je usmereno samo na opservacione studije stoga se ograničava mogućnost poređenja sa eksperimentalnim istraživanjima, koja često koriste drugačije metodološke pristupe.

ZAKLJUČAK

Na osnovu analize 95 opservacionih studija iz oblasti neurologije, utvrđeno je da je primena statističkih metoda u većini radova zadovoljavajuća i u skladu sa karakteristikama obrađenih podataka. Najčešće su korišćeni hi-kvadratni test, *t*-test, Fišerov test tačne verovatnoće i Man-Vitnijev test, pri čemu je njihov izbor najčešće bio prilagođen rezultatima provere normalnosti raspodele i usklađen sa veličinom uzorka.

Rezultati pokazuju da su istraživači u većini slučajeva pravilno prepoznavali statističke pretpostavke i odabirom odgovarajućih testova adekvatno prilagodili analizu, što svedoči o dobrom nivou znanja i primene statističkih metoda u neurološkim studijama. Izveštavanje o korišćenim metodama bilo je pretežno jasno i precizno, što doprinosi transparentnosti i mogućnosti reprodukcije rezultata.

Međutim, i pored ovog pozitivnog trenda, preporučuje se da se posveti dodatna pažnja temeljnijoj proveri pretpostavki i planiranju veličine uzorka radi daljeg unapređenja pouzdanosti i kvaliteta istraživanja. Poboljšanje standarda izveštavanja i kontinuirana edukacija istraživača u oblasti statistike predstavljaju važne korake ka unapređivanju naučne vrednosti i primenljivosti rezultata u kliničkoj praksi.

Rezultati pokazuju da se u neurološkim opservacionim studijama statističke metode primenjuju na adekvatan i profesionalan način, što predstavlja dobru osnovu za dalji razvoj i napredak naučnih istraživanja u ovoj oblasti.

Sukob interesa: Nije prijavljen.

CONCLUSION

Based on the analysis of 95 observational studies in neurology, it was determined that the application of statistical methods in the majority of papers is satisfactory and consistent with the characteristics of the processed data. The most frequently used tests were the chi-square test, *t*-test, Fisher's exact test, and the Mann-Whitney test, with their selection generally being adapted to the results of normality distribution testing and aligned with the sample size.

The results show that, in most cases, researchers correctly recognized statistical assumptions and adequately adapted their analysis by selecting appropriate tests, which testifies to a good level of knowledge and application of statistical methods in neurological studies. Reporting on the methods used was predominantly clear and precise, thus contributing to transparency and the reproducibility of results.

However, despite this positive trend, it is recommended that additional attention be paid to more thorough verification of assumptions and sample size planning so as to further improve the reliability and quality of research. Improving reporting standards and the continuous education of researchers in statistics represent important steps toward enhancing the scientific value and applicability of results in clinical practice.

The results show that statistical methods are applied in an adequate and professional manner in neurological observational studies, which provides a sound foundation for the further development and progress of scientific research in this field.

Conflict of interest: None declared

LITERATURA / *REFERENCES*

1. Altman DG. Statistics in medical journals: developments in the 1980s. Stat Med. 1991 Dec;10(12):1897-913. doi: 10.1002/sim.4780101206.
2. Dawson B, Trapp RG. Basic & clinical biostatistics. 4th ed. New York: Lange Medical Books-McGraw-Hill, Medical Pub. Division; 2004. 438 p.
3. Campese N, Fanciulli A, Stefanova N, Haybaeck J, Kiechl S, Wenning GK. Neuropathology of multiple system atrophy: Kurt Jellinger`s legacy. J Neural Transm (Vienna). 2021 Oct;128(10):1481-94. doi: 10.1007/s00702-021-02383-3. Erratum in: J Neural Transm (Vienna). 2021 Oct;128(10):1495. doi: 10.1007/s00702-021-02412-1.
4. Jellinger KA, Lantos PL. Papp-Lantos inclusions and the pathogenesis of multiple system atrophy: an update. Acta Neuropathol. 2010 Jun;119(6):657-67. doi: 10.1007/s00401-010-0672-3.
5. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess. 2009 Feb;13(12):iii, ix-x, 1-177. doi: 10.3310/hta13120.
6. Giussani G, Westenberg E, Garcia-Azorin D, Bianchi E, Yusof Khan AHK, Allegri RF, et al. Global COVID-19 Neuro Research Coalition. Prevalence and Trajectories of Post-COVID-19 Neurological Manifestations: A Systematic Review and Meta-Analysis. Neuroepidemiology. 2024;58(2):120-33. doi: 10.1159/000536352.
7. Biggin F, Emsley HCA, Knight J. Routinely collected patient data in neurology research: a systematic mapping review. BMC Neurol. 2020 Nov 27;20(1):431. doi: 10.1186/s12883-020-01993-w.
8. Ferrarello F, Viligiardi M, Di Bari M. Ethics reporting practices in randomized controlled trials of physical therapy interventions after stroke. Arch Physiother. 2018 Jun 5;8:8. doi: 10.1186/s40945-018-0049-9.
9. Ng LC, Hanlon C, Yimer G, Henderson DC, Fekadu A. Ethics in global health research: the need for balance. Lancet Glob Health. 2015 Sep;3(9):e516-7. doi: 10.1016/S2214-109X(15)00095-9.
10. Miranda JS, Abbade LPF, Abbade JF, Thabane L, Mbuagbaw L, Pascon GC, et al. Deficiencies in reporting inclusion/exclusion criteria and characteristics of patients in randomized controlled trials of therapeutic interventions in pressure injuries: a systematic methodological review. Int Wound J. 2023 Oct 30;21(2):e14351. doi: 10.1111/iwj.14351.
11. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull World Health Organ. 2007 Nov;85(11):867-72. doi: 10.2471/blt.07.045120.
12. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Erratum in: Nat Rev Neurosci. 2013 Jun;14(6):451.
13. Lee PH, Tse ACY. The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed. Eur J Intern Med. 2017 May;40:16-21. doi: 10.1016/j.ejim.2016.10.008.
14. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One. 2009 Nov 30;4(11):e7824. doi: 10.1371/journal.pone.0007824.
15. Blakley BW, Janzen B. Statistical Considerations in Otolaryngology Journals. Otolaryngol Head Neck Surg. 2017 Nov;157(5):745-7. doi: 10.1177/0194599817704395.
16. Hoekstra R, Kiers HA, Johnson A. Are assumptions of well-known statistical techniques checked, and why (not)? Front Psychol. 2012 May 14;3:137. doi: 10.3389/fpsyg.2012.00137.
17. Parsons NR, Price CL, Hiskens R, Achten J, Costa ML. An evaluation of the quality of statistical design and analysis of published medical research: results from a systematic survey of general orthopaedic journals. BMC Med Res Methodol. 2012 Apr 25;12:60. doi: 10.1186/1471-2288-12-60.
18. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 2012 Spring;10(2):486-9. doi: 10.5812/ijem.3505.
19. To MS, Jukes A. Reporting trends of p values in the neurosurgical literature. J Neurosurg. 2019 Feb 8;132(2):662-70. doi: 10.3171/2018.8.JNS172897.
20. Ginsel B, Aggarwal A, Xuan W, Harris I. The distribution of probability values in medical abstracts: an observational study. BMC Res Notes. 2015 Nov 26;8:721. doi: 10.1186/s13104-015-1691-x.
21. Kumar A, Kishun J, Singh U, Gaur D, Mishra P, Pandey CM. Use of appropriate statistical tools in biomedical research: Current trend & status. Indian J Med Res. 2023 Apr;157(4):353-7. doi: 10.4103/ijmr.IJMR_809_20.
22. Jin Z, Yu D, Zhang L, Meng H, Lu J, Gao Q, et al. A retrospective survey of research design and statistical analyses in selected Chinese medical journals in 1998 and 2008. PLoS One. 2010 May 25;5(5):e10822. doi: 10.1371/journal.pone.0010822.
23. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med. 2007 Aug 30;26(19):3661-75. doi: 10.1002/sim.2832.
24. Oliveira NL, Pereira CAB, Diniz MA, Polpo A. A discussion on significance indices for contingency tables under small sample sizes. PLoS One. 2018 Aug 2;13(8):e0199102. doi: 10.1371/journal.pone.0199102.
25. Serra N. A significant minimization of Pearson's chi square statistics in 2x2 contingency tables: preliminary results for small samples. Epidemiol Biostat Public Health. 2018;15(3):e12949. doi: 10.2427/12949.