**ORIGINAL ARTICLE**

# STATISTICAL REPORTING IN PHARMACEUTICAL PAPERS FROM ROMANIAN JOURNALS

DANIEL-CORNELIU LEUCUȚA[1], TUDOR DRUGAN[1]*, ANDREEA FARCAŞ[2], ANDREI ACHIMAŞ[1]

[1]*Department of Medical Informatics and Biostatistics, "Iuliu Haţieganu", University of Medicine and Pharmacy Cluj-Napoca, 6 Louis Pasteur, 400349, Cluj-Napoca, Romania*
[2]*Drug Information Research Centre Cluj-Napoca, "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, Pasteur Street No. 6, 400349, Romania*

**corresponding author: tdrugan@umfcluj.ro*

*Manuscript received: December 2014*

**Abstract**

The quality of the statistical reporting is poor in various biomedical journals, this being a problem stressed by numerous studies. Our aim was to evaluate all pharmaceutical papers published in six Romanian journals, in 2013, for statistical reporting. We used mainly the "Statistical Analyses and Methods in the Published Literature" – SAMPL - guideline to assess the statistical reporting in the respective original pharmaceutical papers. We found 170 original articles from which 134 (78.82%) reported statistical results. The most frequently used tests were ANOVA (50.9%), and t-test (37.7%). Confirming that the assumptions of the test were met by the data was not reported in 87.6% of the papers. The most important error found in the papers using hypothesis testing, was choosing the wrong statistical test (39.6%). The findings were similar to other published studies. The quality of reporting statistical analyses of pharmaceutical papers published in Romanian journals should be improved, in order to reduce the errors in studies, improve the methodology and clarity.

**Rezumat**

Calitatea raportării aspectelor statistice în diverse reviste medicale este redusă, aceasta fiind o problemă evidențiată de numeroase studii. Scopul acestei lucrări a fost evaluarea tuturor articolelor originale farmaceutice publicate în șase reviste românești pe parcursul anului 2013, în ceea ce privește raportarea aspectelor statistice. Am folosit în principal ghidul "Statistical Analyses and Methods in the Published Literature" - SAMPL, pentru a evalua calitatea raportării aspectelor statistice în articole originale farmaceutice respective. Am găsit 170 de articole originale, dintre care 134 (78,82%) au raportat rezultate statistice. Cele mai utilizate teste au fost: ANOVA (50,9%), și testul t (37,7%). Confirmarea faptului că prezumpțiile testelor au fost îndeplinite de date nu a fost raportată în 87,6% din reviste. Cea mai importantă eroare găsită în articolele care foloseau testarea ipotezei, a fost alegerea greșită a testului statistic (39,6%). Rezultatele au fost similare cu ale altor studii prezentate în literatură. Calitatea raportării aspectelor statistice în articolele farmaceutice publicate în revistele din România trebuie îmbunătățită, pentru a reduce erorile, a îmbunătăți metodologia și claritatea studiilor.

**Keywords**: statistical reporting, pharmacy, SAMPL guidelines

## Introduction

Statistics is a branch of mathematics that studies the collection, analysis, interpretation and presentation of data. Since variability of biomedical characteristics is inherent, the use of this methodology is compulsory. Insufficient understanding of statistical methods (their logic, use and interpretation) by researchers causes errors that get us further from the truth and diminishes the quality of the research conclusions. The quality of the statistical reporting is poor in various biomedical journals even though the problem was stressed by numerous studies. For example, studies assessing articles published in medical journals show statistical tests errors rate that vary from 20% to 70% [1-4].

Although many studies assessed statistical reporting in different countries and journals, we found no studies regarding statistical reporting of pharmaceutical papers published in Romanian journals.

Thus, our aim was to evaluate all original pharmaceutical papers published in six Romanian journals, in one year (2013), for statistical reporting.

## Materials and Methods

We used the "Statistical Analyses and Methods in the Published Literature" (the SAMPL Guidelines) [5], and a custom set of items (devised by us for several pharmaceutical methods) to assess the statistical reporting in pharmaceutical papers published in several Romanian Journals. From the SAMPL Guidelines [5] we used a selection of items to assess the quality of statistical reporting. The principles or concepts of reporting will be denoted as items from now on. The items that are identical

to the text written in the guideline were marked by * sign, those that were modified by us were marked by $ sign and the new ones that were added were marked by + sign. We assessed all original articles, published in 2013, from the online website, of three Thompson ISI indexed [6] journals: *Farmacia* (1.251 impact factor in 2013), *Studia Universitatis Babes-Bolyai Chemia* (0.136 impact factor in 2013), and *Revue Roumaine de Chimie* (0.393 impact factor in 2013), and three journals not indexed in Thompson ISI database: *The Medical-Surgical Journal*, *Acta Medica Marisiensis*, and *Clujul Medical*. We chose the journal whose publishing scope is the pharmaceutical field (*Farmacia*), and we had searched the papers in 2013 of some the most important Romanian journals in the biomedical (using the list of journals updated by a Romanian governmental research body that ranks national journals [7, 8]) field to identify those in the pharmaceutical field. Each article was analysed fully by at least two authors, disagreements being solved by discussion, without measuring agreement. For each item assessed, the number of papers in which it was reported, along with its percentage was presented. The percentages were computed by dividing the number of times the item was found/reported to the number of papers where the item was considered as relevant. The statistical analysis was performed in R core team environment for statistical computing and graphics, version 3.1.2 [9].

**Results and Discussion**

We found 170 original articles - pharmacy related papers, published in 2013, in six Romanian journals: *Farmacia* with 120 papers (70.6%), *Studia Universitatis Babes-Bolyai Chemia* with 6 (3.5%) papers, *Revue Roumaine de Chimie* with 15 (8.8%) papers, *The Medical-Surgical Journal* with 21 (12.4%) papers, *Acta Medica Marisiensis* with 7 papers (4.1%), and *Clujul Medical* with 1 paper (0.6%).

Out of all these papers 134 (78.8%) reported statistical results. The rest didn't report due to their aim or study design (ex. novel drug synthesis), or despite the reporting would have been necessary.

We didn't use all the items from the SAMPL guidelines, and some of them were modified, since these guidelines are meant for assisting the writing of the papers before publishing, but here we used them for assessing already published papers.

Due to the diversity of study designs, and statistical analyses, we considered important not to use as denominator the whole number of articles reviewed – which would have been the easiest task (and a common approach in this genre of articles), but to try to identify only those articles where the items were relevant to (function of the particularities of each article). This is why the totals for each of the reported item vary, from item to item, since the items are not applicable to all study types that were assessed. This leads to higher percentages of problems in the papers, compared to articles that use as denominator the whole number of papers reviewed.

*Primary analyses*

The reporting situation of the primary analyses is presented in Table I.

**Table I**

Reporting of primary statistical analyses

| Item | n/total (%) |
|---|---|
| summarize each variable with descriptive statistics $ | 101/104 (97.1) |
| verify that data conformed to the assumptions * | 14/113 (12.4) |
| skewed data were analysed with non-parametric tests * | 6/10 (60.0) |
| paired data were analysed with unpaired tests $ | 3/53 (5.6) |
| the study had multiple comparisons + | 25/101 (24.8) |
| indicate whether and how any allowance or adjustments were made for multiple comparisons * | 11/25 (44.0) |
| report how any outlying data were treated in the analysis $ | 4/82 (4.9) |
| say whether tests were one- or two-tailed | 4/51 (7.8) |
| how many were reported as one-tailed + | 1/4 (25) |
| justify the use of one-tailed tests * | 0/1 (0.0) |
| report the alpha level (e.g. 0.05) * | 40/53 (75.5) |
| alfa levels used: 0.01 + | 2/40 (5.0) |
| alfa levels used: 0.05 + | 38/40 (95.0) |
| name the statistical package or program used * | 44/134 (32.8) |

Not all papers had descriptive statistics summarized, some because it was not necessary (these were not taken into account), and only a few - 3 (2.9%) - didn't report it at all. A number of the papers – 12 (11.5%) -

presented the data only graphically. These findings are rather positive, but they can be improved.

No study identified the smallest difference considered to be clinically important (but also practically/biologically important/relevant) *. This is a worrisome finding since the clinical/practical/ biological importance should be sought after not the tyranny of the p-value. This smallest useful difference is not always an easy task to define. Never the less the clinical/practical/biological/scientifical importance of the differences should be always discussed, beside the statistical significance.

When assumptions of the statistical analyses, like skewedness/non normality of the data could be assessed in the papers (being stated by the authors or assessed from charts) 4 (40%) of the papers failed to use non-parametric tests. This is an important error. If the assumptions weren't met, the results are incorrect. An important number of papers – 99 (87.6%) - didn't report they verified that data conformed to the assumptions of the statistical analyses. Although this doesn't necessary mean the assumptions weren't properly checked, this is a big concern. Researchers have to check and also report this important analysis step. A similar high percentage was found in another study that reported 71.2% [1]. Another related error found was the use of unpaired statistical tests for paired data in three (17%) out of 17 papers where the type of test could be identified. The elimination of outliers is a subject of debate. Only a few articles reported their approach to handling outliers, which is normal since this should be reported only when relevant. There might be other articles that omitted to mention their handling that might have been important for the reader to know, so we cannot assess their correctness.

14 (66%) of the papers using multiple comparisons failed to indicate weather allowance or adjustments were made for multiple comparisons. Multiple comparisons inflate the type I error. This is why researchers have to correct this, using specific adjustments for multiple comparisons.

The name of the statistical package or program used was not reported in 44 (67.2%) of papers. This reduced reporting might be due to the fact many papers presented very simple statistics, like descriptive statistics. The reported ones were GraphPad Prism 9 times (20.5%), Microsoft Excel 6 times (13.6%), Matlab and Modde two times (4.5%), and only once (2.3% each) the following: AMIX, Bioinfoqsar, Epi Info, Gold software, Kinetica, Mathematica, MedCalc, MobyDigs, Phoenix, R, SAS, Simca-p+, Statistica, Stats Direct, Table Curve 2D, WinCats.

No supplementary analyses were reported as performed (sensitivity analyses, imputation of missing values) – which is normal since they are needed for specific cases of complex studies, analyses, data, and data particularities that are less frequently found in the analysed papers. There were studies that underwent subgroup analyses, but none of them identified it as explanatory. If the subgroup analyses were thought after the data collection, then these analyses should have been identified as explanatory. The reality of subgroup analyses is more complex than this, but a comprehensive discussion on this is beyond the scope of this article [10, 11].

*Reporting numbers and descriptive statistics*

The way numbers and descriptive statistics were reported is shown in Table II.

There are papers that use too many decimal places (3, 4 or 5 depending on what was measured, without using a rigid limit of say 2) – 30 (22.4%), that diminishes the ease of comprehension degree.

Reporting of total or group sample size for analyses was not applicable for 14 (10.4%) papers, like: structure activity studies, validation studies, characterization of specific formulations studies, biosensors, antioxidant activity. The studies that used groups were heterogeneous from the point of view of the units of observation (pharmaceutical formulations [tablets, solutions], patients, animals [mice, frogs], plants, cigarettes). The number of groups was between 1 and 24, with a median of 3, interquartile range from 1 to 5. The number of units per group was between 3 and 1100, with a median of 9, interquartile range from 3 to 24.

A number of 21 (15.7%) papers reported percentages. Out of those, 11 (52.4%) papers failed to report the numerators and denominators for all percentages - 6 (28.6%) papers, or for some of them - 5 (23.8%) papers. Without the absolute frequencies and the total number of observations, the image the reader gets is incomplete.

For those papers for whom the normality of the data was reported, the mean and standard deviation were correctly used generally, but for the data that was not normally distributed the median and ranges were poorly used (only 20%). The latter is an error that should be avoided. Also the use of medians (no matter the distribution) is relatively low – 4%, compared to the means – 86.1%. This is not necessarily a problem. It might be due to the fact data was indeed normally distributed, or if this was not the case, the habit of using means is still strong. Another clear error was the reporting of the mean along with a non-parametric test in 7 papers. The median should have been presented if the data was not normally distributed.

Beside papers reporting means and standard deviations, there were situations (9 papers) where numbers were presented, but it wasn't written if the value was a mean or a median or another descriptive statistics. There were situations (3) where data were presented as mean ± a value, but it

wasn't presented if the value after the ± sign was a standard deviation, or a standard error of mean. These are more likely involuntary omissions, but they should be avoided.

Reporting using the format of "mean ± standard deviation", which is not advised by the guideline, was found in 58 (85.3%) papers. This is more a formal thing, and not an error.

The size of the range without reporting the upper and lower boundaries of inter-percentile ranges and the minimum and maximum values of ranges was found in three (21.4%) papers out of 14. The range is less informative about possible problems that are present in the data.

The standard error of the mean (SE) to indicate the variability of a data set was reported in 12 (13.3%) of 86 papers. This should be avoided since SE, a measure of precision of the mean, is the standard deviation of the sampling distribution of the mean in the population, not a measure of variability of the observed sample [12]; instead the standard deviation should be reported. This result is much better than what was found in two pharmacology journals (90% of 190 articles) [2].

The most important errors in reporting numbers and descriptive statistics were the underreporting of data distribution.

**Table II**

Reporting of numbers and descriptive statistics

| Item | n/total (%) |
|---|---|
| round to a reasonable extent * | 104/134 (77.6) |
| report total or group sample size for analyses * | 96/120 (80.0) |
| report numerators and denominators for all percentages * | 10/21 (47.6) |
| the normality of the data was reported + | 11/113 (9.7) |
| summarize data that were stated as normally distributed with means $ | 7/7 (100.0) |
| summarize data that were stated as normally distributed with standard deviations $ | 5/7 (71.4) |
| the use of means (no matter the distribution) + | 93/108 (86.1) |
| the use of SD (no matter the distribution) + | 74/108 (68.5) |
| use the form: mean (SD), not mean ± SD + | 10/68 (14.7) |
| summarize data that was stated as not normally distributed with medians $ | 1/5 (20.0) |
| summarize data that was stated as not normally distributed with inter-percentile ranges, ranges, or both $ | 1/5 (20.0) |
| the use of medians (no matter the distribution) + | 4/101 (4.0) |
| report the mean along with a non-parametric test + | 6/32 (18.8) |
| report the upper and lower boundaries of inter-percentile ranges and the minimum and maximum values of ranges * | 11/14 (78.6) |
| do not use the standard error of the mean (SE) to indicate the variability of a data set * | 78/90 (86.7) |
| The use of standard deviations, inter-percentile ranges, or ranges * | 80/100 (80.0) |

*Reporting hypothesis tests*

53 papers reported hypothesis tests out of 134, and the details can be seen in Table III.

The most important reporting problems found were: the failure to confirm that the assumptions of the test were met by the data; the failure to report if the test was for paired or independent samples – if the choice between the two is incorrect the results can move away from the truth; failure to show precision for differences between groups especially by using confidence intervals – that helps to extrapolate the results of the study.

**Table III**

Reporting hypothesis tests

| Item | n/total (%) |
|---|---|
| identify the variables in the analysis * | 51/53 (96.2) |
| summarize the data for each variable with the appropriate descriptive statistics * | 42/53 (79.2) |
| report equivalence margin * | 0/1 (0.0) |
| report whether the test was for paired samples * | 3/53 (5.7) |
| report whether the test was for independent samples * | 4/53 (7.5) |
| confirm that the assumptions of the test were met by the data * | 7/53 (13.2) |
| precision for differences between groups $ | 2/40 (5.0) |
| precision for agreement between groups / for diagnostic measures / for regression slopes $ | - |
| do not use the standard error of the mean (SE) to indicate the precision of an estimate * | 12/18 (66.7) |
| 95% confidence coefficient instead * | 2/18 (11.1) |

The following statistical tests were stated in 53 papers: Kolmogorov-Smirnov: 2 (3.8%), Shapiro-Wilk: 2 (3.8%), D'agostino Pearson, 4 (7.5%), t-test not stating if paired or unpaired: 14 (26.4%): t-test unpaired: 5 (9.4%): t-test paired: 1 (1.9%): rank test: 1 (1.9%), Wilcoxon paired: 1 (1.9%), Mann Whitney U: 4 (7.5%), Wilcox not stating which type: 1 (1.9%), F-test, 1 (1.9%), analysis of variance (ANOVA) one way: 12 (20.8%), ANOVA one way repeated measures: 2 (4%), ANOVA not stating which type: 9 (18%) – though one could identify from other elements in the paper the type of ANOVA used, all ANOVA: 27 (50.9%), Kruskal-Wallis: 2 (3.8%): Chi square test, 4 (7.5%), Fisher Exact test, 1 (1.9%). In 4 (7.5%) cases the name of the test was not reported. In 2 (3.8%) cases reported names of test were not found to be used in results. Results were almost identical in respect of the most frequently used statistical tests (ANOVA – 50.9% and t-test – 37.7%), and non-parametric tests – 11.3%, were found in two Indian pharmacology journals in 2002-2010 (58.4%, 24.2%, and 10.7%) [2]. Non-parametric tests are less frequently used, as we showed before for the use of the median. When compared to papers in 6 western pharmacy journals the results were non-identical: ANOVA:18.1%, t-tests: 26.4%, non-parametric tests:33.3% (144 papers in 2001 from *American Journal of Health-System Pharmacy*, *The Annals of Pharmacotherapy*, *Canadian Journal of Hospital Pharmacy*, *Formulary*, *Hospital Pharmacy, and Journal of the American Pharmaceutical Association* [13]). When compared to medical journals the percentages were not exactly alike: ANOVA: 17%, t-tests: 26.4%, non-parametric tests: 21.7%, Chi-square: 48%, (106 papers from Journal of American Medical Association in 2010 [14]). The use of paired or independent sample tests was not clearly stated in 40 (85.1%) of 47 papers. The use of these types of tests could be understood by a reader from the name of the test, without the need of a clear statement (e.g. Kruskal Wallis test, Mann Whitney U test) in 13 (27.7%) papers. We found some errors: using an independent test when a paired one was needed in 3 (6.4%) papers, but not for the other way around. This result is similar to the one found in two pharmacology journals (8.1% of 5 articles) [2].

Out of 53 papers, the p values were reported as equalities in 19 (35.8%) papers, a mix of equalities and other ways (e.g. inequalities, or NS – non-significant) in 4 (7.5%) papers, or different ways that are not equalities, in 30 (56%) papers. P values were reported as inequalities (e.g. $p < 0.05$) in 32 (60.4%) papers or as "NS" (non-significant) in 14 (26.4%) papers. The maximum number of decimals of the p value reported as equality or inequality in 26 papers was 2 decimals in 4 (15.4%) papers, 3 decimals in 10 (38.5%) papers, 4 decimals in 10 (38.5%) papers, and 5 decimals in 5 (7.6%) papers. P values should be presented as equalities with not too many decimals. This is rather about style; the biggest problem is the abuse of p-values and the underutilization of confidence intervals [15, 16].

When taking into account important proven errors, related to choosing the wrong statistical test, like: skewed data analysed with non-parametric tests, using independent sample tests for paired data (or vice versa), not adjusting for multiple testing, 21 (39.6%) papers were found with problems. This is similar to what was found in other studies: 41.6% (*British Medical Journal* in 1977 [17]), 31.7% (2 Indian pharmacology journals in 2002-2010 [2]), 21.3%-67.9% (10 leading Chinese medical journals in 2008 [4]), 28.7% (80 papers in medical journals from Pakistan [1]).

*Reporting association analyses*

Association analyses were found in 6 (4.5%) papers out of 134 (see Table IV), and were well reported, except the lack of a measure of association. This can be useful to show the strength of association, but since descriptive statistics and contingency tables were used, they can help the reader understanding the data.

**Table IV**

Reporting association analyses

| Item | n/total (%) |
|---|---|
| reporting association is a primary analysis + | 4/6 (66.7) |
| describe the association of interest * | 5/6 (83.3) |
| identify the variables used * | 5/6 (83.3) |
| summarize each with descriptive statistics * | 5/6 (83.3) |
| identify the test of association used * | 4/6 (66.7) |
| for tests of association (e.g., a Chi-square test), report the P value of the test * | 5/6 (50.0) |
| present a measure of association * | 0/6 (.00) |
| for primary comparison, consider including the full contingency table for the analysis * | 3/4 (75.0) |

*Reporting correlation analyses*

Correlation analyses were found in 12 (9%) papers out of 134, and the reporting details are in Table V. The most important reporting problem here is the absence of reporting that the assumptions of the analysis were met. Describing correlation as low, moderate or high, without having these categories defined, is misguiding. Confidence intervals scatter plots, and additional info on the scatterplot, should be used more often for correlation analyses.

**Table V**

Reporting correlation analyses

| Item | n/total (%) |
|---|---|
| reporting correlation is a primary analysis + | 5/12 (41.7) |
| identify the correlation coefficient used in the analysis (e.g., Pearson, Spearman) * | 9/12 (75.0) |
| confirm that the assumptions of the analysis were met * | 0/12 (0.0) |
| report the value of the correlation coefficient * | 11/12 (91.7) |
| do not describe correlation as low, moderate, or high unless the ranges for these categories have been defined | 8/12 (66.7) |
| for primary comparisons, report the (95%) confidence interval for the correlation coefficient, whether or not it is statistically significant * | 1/5 (2.0) |
| for primary comparisons, consider reporting the results as a scatter plot * | 1/5 (2.0) |
| the correlation coefficient on the chart * | 1/3 (33.3) |
| the sample size, the confidence interval, the p value on the chart * | 0/3 (33.3) |

*Reporting regression analyses*

Regression analyses as primary analyses were found in 5 (3.7%) papers out of 134. Their reporting is shown in Table VI. The most important reporting problems here are: the lack of checking the assumptions of the analysis (linearity, collinearity) – that can induce error in results; failure to describe the variable selection process by which the final multiple regression model was developed; the absence of the p-values and of the confidence intervals for the coefficients – without which one can't assess if and how much the results are reliable.

**Table VI**

Reporting regression analyses

| Item | n/total (%) |
|---|---|
| describe the purpose of the analysis * | 4/5 (80.0) |
| summarize each with descriptive statistics * | 1/5 (20.0) |
| confirm that the assumptions of the analysis were met * | 1/5 (20.0) |
| for example, in linear regression indicate whether an analysis of residuals confirmed the assumptions of linearity * | 1/5 (20.0) |
| report how any outlying values were treated in the analysis * | 1/5 (20.0) |
| report how any missing data were treated in the analyses * | 0/5 (0.0) |
| for either simple or multiple (multivariable) regression analyses, report the regression equation * | 5/5 (100.0) |
| for multiple analyses: report whether the variables were assessed for collinearity * | 0/2 (0.0) |
| for multiple analyses: report whether the variables were assessed for interaction * | 0/2 (0.0) |
| for multiple analyses: describe the variable selection process by which the final model was developed (e.g. forward-stepwise; best subset) * | 1/2 (50.0) |
| report the regression coefficients (beta weights) of each explanatory variable * | 0/5 (0.0) |
| report the associated confidence intervals * | 0/4 (0.0) |
| report the p values * | 0/5 (0.0) |
| provide a measure of the model's "goodness-of-fit" to the data * | 5/5 (100.0) |
| provide the coefficient of determination, $R^2$, for simple regression * | 3/5 (60.0) |
| provide the coefficient of multiple determination $R^2$, for multiple regression * | 3/3 (100.0) |
| specify whether and how the model was validated * | 1/4 (25.0) |
| for primary comparisons analysed with simple linear regression analysis, consider reporting the results graphically, in a scatter plot showing the regression line * | 2/3 (66.7) |
| show the regression line confidence bounds * | 0/2 (0.0) |
| do not extend the regression line beyond the minimum and maximum values of the data * | 2/2 (100.0) |

*Reporting ANOVA*

ANOVA was found in 27 (20.1%) papers out of 134, that 20 (74.1%) were primary analyses. The most important reporting problems (see Table VII) here were: the lack of checking the assumptions of the analysis, like linearity or equality of variances - only one study reported it partially (3.8%) – that can induce error in results; the absence of the p-values for each explanatory variable – beside the four that reported it, and the three reported it only partially; the absence of a measure of the goodness-of-fit for the model – that helps the reader to assess how well the model fits a set of observations.

**Table VII**

Reporting ANOVA

| Item | n/total (%) |
|---|---|
| describe the purpose of the analysis * | 23/27 (85.2) |
| summarize each variable with descriptive statistics * | 25/27 (92.6) |
| confirm that the assumptions of the analysis were met * | 0/26 (0.0) |
| indicate whether an analysis of residuals confirmed the assumptions of linearity * | 0/27 (0.0) |
| report if outliers were found and how any outlying data were treated in the analysis $ | 0/27 (0.0) |
| report if there was missing data and how any missing data were treated in the analyses $ | 0/27 (0.0) |
| there is more than one variable so interaction could be tested $ | 8/22 (36.4) |
| specify whether the explanatory variables were tested for interaction * | 2/8 (25.0) |
| report the p value for each explanatory variable * | 4/11 (36.4) |
| report the test statistics * | 7/27 (25.9) |
| report the degrees of freedom for the analysis * | 4/27 (14.8) |
| provide an assessment of the goodness-of-fit of the model to the data, such as $R^2$ * | 2/26 (7.7) |
| specify whether and how the model was validated * | 0/8 (0.0) |

Not reporting the degrees of freedom is of reduced importance because having the sample and group size information helps in this respect.

The results in the following subchapters were assessed with custom made items, drawn for the most common types of analyses found in the analysed papers.

*Design of experiments*

Six papers reported the use of design of experiments (4.5%). The following concepts were reported: identify independent variables: 6 (100%); identify dependent variables: 6 (100%); present the experimental design matrix: 6 (100%); fitting experimental data: 5 (83.3%); present $R^2$: 5 (83.3%); present $Q^2$: 2 (33.3%); present ANOVA: 5 (83.3%); present dependence curves: 4 (66.7%); checking for interaction between factors: 4 (66.7%); equation coefficients numbers [18] or chart: 6 (100%). The reporting of experimental design studies for many items is very good, and for some items it can be improved, especially for checking interactions, and assessing goodness-of-fit.

*In vitro kinetic release analysis*

Eleven papers reported the use of *in vitro* kinetic release analysis (8.2%). The following concepts were reported: present kinetic equations used): 8/11 (72.7%); present equation coefficients: 3/9 (33.3%); present k: 5/11 (45.5%); present n: 7/11 (63.6%); present $R^2$: 9/11 (81.8%); [19] present Akaike Information Criterion - AIC: 2/11 (18.2%). AIC, Bayesian Information Criteria, followed by Mallow's Cp and adjusted $R^2$ should be used more when comparing different models, and to rely less only on the $R^2$.

*Validation*

Twenty two papers reported validation procedures (16.4%). The validation was the primary objective of the study in 14 (63.6%) papers. The following concepts were reported: coefficient of correlation or coefficient of determination: 21/22 (95.5%); standard error: 6/22 (27.3%); limit of detection: 14/22 (63.6%); limit of quantification: 16/22 (72.7%); intraday precision: 15/22 (68.2%); interday precision: 11/22 (50%); linearity assessment: 19/22 (86.4%); calibration curve: 17/22 (77.3%); relative standard deviation: 19/22 (86.4%); method accuracy % recovery: 17/22 (77.3%). The reporting of experimental design studies for many items is in the range 63.6%-86.4%, except the coefficient of correlation or coefficient of determination is almost always present, the interday precision and the standard error that have lower values. Therefore it can be improved for almost each item, especially for precision and limit of detection. These lower reporting values might be due to the fact many papers used the validation not as a primary objective, so they reported fewer qualities in the paper.

This study brings several original contributions to the scientific literature: it evaluates the statistical reporting of a corpus of papers in Romania that was not assessed before; it uses a comprehensive list of

items from the SAMPL guidelines; it adds new elements to assess the reporting of statistical analyses that are frequently found in the pharmaceutical literature.

## Conclusions

The distribution of reported statistical tests was similar to other pharmacology studies published in India, but different than in western pharmacy and some medical journals. The most frequently used tests were ANOVA, and t-test.

There were many aspects that were not well reported. Some were about style (not errors), some things that were not reported (this doesn't mean those steps were not dealt with during the study), could be found in other places in the article, some were important errors. The spirit of the guidelines is to have the information available, not necessarily in the results section.

Confirming that the assumptions of the test were met by the data is lacking in the papers. The most important error found was the wrong choice of statistical tests. The findings were similar to other studies. Another important problem is the reduced use of confidence intervals to express the results.

This paper is original by focus, use of SAMPL guidelines [5], and extensions for pharmaceutical papers.

The quality of reporting statistical analyses of pharmaceutical papers published in Romanian journals should be improved, in order to reduce the errors in studies, and improve methodology and clarity.

To achieve this, several paths should be followed: statisticians should be involved in studies from the planning phase of the research, researchers should update their statistical knowledge, journal editors should use statistical advisors to review submissions, journals should use statistical reporting guidelines and/or checklists like the one used in this paper.

## Acknowledgements

This paper was published under the frame of European Social Found, Human Resources Development Operational Programme 2007-2013, project no. POSDRU/159/1.5/S/138776.

## References

1. Hanif A., Ajmal T., Statistical Errors in Medical Journals (a Critical Appraisal). *Ann. King Edward Med. Univ.*, 2011; 17(2): 178-182.
2. Jaykaran J., Yadav P., Chavda N., Kantharia N.D., Some issues related to the reporting of statistics in clinical trials published in Indian medical journals: A survey. *Int. J. Pharmacol.*, 2010; 6(4): 484-489.
3. Gardner M.J., Altman D.G., Jones D.R., Machin D., Is the statistical assessment of papers submitted to the "British Medical Journal" effective? *Br. Med. J. (Clin. Res. Ed.)*, 1983; 286: 1485-1488.
4. Wu S., Jin Z., Wei X., Gao Q., Lu J., Ma X., Misuse of Statistical Methods in 10 Leading Chinese Medical Journals in 1998 and 2008. *Sci. World J.*, 2011; 11: 2106-2114.
5. Lang T.A., Altman D.G., Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or The "SAMPL Guidelines". Science editors' handbook. 2013: 29-32.
6. Journal Citation Reports® | Thomson Reuters.
7. Unitatea Executiva pentru Finantarea Invatamantului Superior, a Cercetarii, D. si I., Reviste romanesti recunoscute de CNCSIS categoria B plus.
8. REVISTE ROMÂNEȘTI INDEXATE/COTATE "ISI".
9. R Core Team., R: A Language and Environment for Statistical Computing. Vienna, Austria, 2014,
10. Rothwell P.M., Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*, 2005; 365(9454): 176-186.
11. Naggara O., Raymond J., Guilbert F., Altman D.G., The problem of subgroup analyses: an example from a trial on ruptured intracranial aneurysms. *AJNR - Am. J. Neuroradiol.*, 2011; 32(4): 633-636.
12. Everitt B.S., Skrondal A., The Cambridge Dictionary of Statistics. 4th ed., Cambridge University Press, Cambridge, 2010.
13. Lee C.M., Soin H.K., Einarson T.R., Statistics in the pharmacy literature. *Ann. Pharmacother.*, 2004; 38(9): 1412-1418.
14. Arnold L.D., Braganza M., Salih R., Colditz G.A., Statistical trends in the Journal of the American Medical Association and implications for training across the continuum of medical education. *PLoS One*, 2013; 8(10): e77301.
15. Nuzzo R., Scientific method: statistical errors. *Nature*, 2014; 506(7487): 150-152.
16. Smith, C.J., Fox Z.V., The use and abuse of hypothesis tests: how to present P values. *Phlebology*, 2010; 25(3): 107-112.
17. Gore S.M., Jones I.G., Rytter E.C., Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. *Br. Med. J.*, 1977; 1(6053): 85-87.
18. Popa G., Ochiuz L., Stoleriu I., Cojocaru I., Popovici I., Formulation and preparation of orally disintegrating tablets using innovative binder. *Farmacia*, 2013; 61(6): 1131-1136.
19. Mircioiu I., Anuta V., Purcaru S.O., Radulescu F., Miron D., Dumitrescu I.B., *In vitro* dissolution of poorly soluble drugs in the presence of surface active agents - *in vivo* pharmacokinetics correlations. II. Nimesulide. Farmacia, 2013; 61(1): 88-102.