MISUSAGE OF STATISTICS IN MEDICAL RESEARCH

Ilker Ercan[1], Berna Yazıcı[2], Yaning Yang[3], Guven Özkaya[1],
Sengul Cangur[1], Bulent Ediz[1], Ismet Kan[1]

Uludag University, Medical Faculty, Department of Biostatistics[1], Bursa, Anadolu University, Science Faculty, Department of Statistics[2], Eskisehir, Turkey. University of Science and Technology of China, Department of Statistics and Finance, School of Business[3], Hefei, China.

The inappropriate use of statistical method and technique cause time and cost losts and it can be misleading to other scientific researhes. So in this study the main statistical error sources in medical research are discussed and aimed to be informative for researchers. The most common statistical error sources are determined examining the previous medical researches and taking errors into account occured in researches during statistical consulting. Inappropriate use of statistics can be found in every stage of a medical research related to data analysis; design of the experiment, data collection and pre-processing, analysis method and implementation, and interpretation. We listed several error sources that researchers easily commit if they are lack of solid statistical background. The mistakes in the studies mostly occur because of the researchers' lack of statistical knowledge and since they don't take statistical consulting. Unbiased, consistent, and efficient parameter estimates are made in statistics science. This can be provided using statistics from the planning until the end of the study. So it is neccessary to consult statisticians at each stage of the studies.

Key words: Statistics, Research, Medical research, Statistical mistakes, Statistical errors

INTRODUCTION

Statistics is needed at every stage of the research beginning from planning to the end, in order to gain scientifically importance and to obtain reliable results. The use of the inappropriate statistical method, technique and the analysis cause time and cost losts and most importantly thinking in the way of scientific ethics, it gives harm to science and humanity. Even if the study is carefully planned to conduct as a result of applications with errors, the misleading results might be obtained. That leads other mistakes who takes as a reference to those studies.

The increment of knowledge with the improvement of the tools used for obtaining knowledge and the complex structure of the knowledge require the necessity for the analysis of the data and we know that is only provided by statistics. With that development as mentioned by Sahai and Ojeda (1), physicians and other staff interested in medicine, notice that they need biostatistics principles and methods. Over the past decades, the use of statistics in medical journal has increased both in quantitiy and in sophistication (2, 3). The development of statistical software and computer are parallel with that improvement (4). A disadvantage of that development is, although not often recognized by consumer of research the statistical errors are so common that it is believed that almost 50% of medical literature have statistical flaws (5). Serious statistical errors were found in 40% of 164 articles published in a psychiatry journal and in 19% of 145 articles published in an obstetrics and gynecology journal (6, 7).

This study is prepared to put forward the mistakes that the researchers usually make, taking account the errors during the statistical consultings and the errors in some published works and to express the importance of statistical consulting.

Correspondence: *Dr. Ilker Ercan*
Department of Biostatistics, Medical Faculty, University of Uludag, Gorukle, Bursa, Turkey
Tel: 902244428200 (21028), GSM: 905053781903, Fax: 902244428666
E-mail adress: ercan@uludag.edu.tr

MAIN ERROR SOURCES IN RESEARCH

We enumerate some common mistakes in each stage of a research. We classify error sources in each stage of the research; stage of design, data collection and pre-processing, decide on methods and implementation, and interpretation.

Design of The Experiment
Description of the population

The population which the researcher will study on, must be defined in terms of time, location and at least one common particular characteristic (8). The clearness of the definition either provides clearly determined frame of the study or provides easiness in choosing the units that will be in the sample. The researchers have problems in choosing the units of the sample in case of badly defined population and this leads to increment in heterogeneity. Another benefit of a good definition of the population is to determine the variables clearly that will be analysed in the study (9, 10).

Sampling scheme
The errors in deciding the sampling technique

Each of the sampling techniques aims to make inference on the population parameter with the smallest error (11). More than one sampling technique can be used in a study. The subject of the study, the characteristics of the population, the length of the research and the cost must be taken into account in deciding the sampling technique. For all sampling procedure, particularly simple ramdom sampling is used unconsciously. Although making haphazardly sampling in many studies it is declared that simple random sampling technique is used (12). One source of using wrong sampling technique is the tendency of using the same sampling techniques that have been used in other similar studies. If the used techniques are not appropriate, the researchers run the risk of misinterpreting findings by using inappropriate, unpresentative and biased samples (13).

Williamson stated in his study, 89 (68%) studies misrepresented their samples as random although in fact they were either convenience samples or entire populations. This leaves a total of only 42 (32%) studies using genuine random sampling, or acceptable variant of sampling such as cluster sampling or stratified random sampling (13).

Sampling criteria

To represent the population by the sample, determining the subjects that will be included in the sample is the next stage that requires attention after deciding the appropriate sampling technique (14, 15). So the criterion of the selection must be clearly determined. One of the most common errors in selection of the subject is collecting the units by different researchers who are not in the research group. Especially that occurs as a result of work of the reserchers who do not have enough knowledge about the research at data collection stage. If the selection criteria are not well known in selection of the subjects, one, unconsciounously can be biased (12, 16).

In the studies, eligibility criteria are often not reported adequately (17). For example, 25% of 364 reports of randomized, controlled trials in surgery did not specify the eligibility criteria (18).

Selection the type of the sampling

Selection type of the units to the sample is also defined due to the research subject. The misreprensentation of nonprobability sampling as random sampling has important implications (13). Nonprobability samples often reflect selection biases of the person doing the study and do not fulfill the requirements of randomness needed to estimate sampling errors. Random sampling methods are used when a sample of subjects is selected from a population of possible subjects in observational studies, such as cohort, case-control, and cross-sectional studies (19). Especially in the studies that are made in order to get the knowledge about the population, probability sampling is inevitable. But in some cases researchers make mistakes by not using probability techniques. So constructing probability sampling or nonprobability sampling due to the research subject should be examined very carefully.

Defining the number of subjects

The representation ability of the sample increases as the number of subject increases. Appropriate sample size should be obtained examining the previous studies, with an error and at a significance level. But some researchers, although they have some information (mean/proportion, standard deviation/standard error of mean etc) to define the appropriate sample size that they can get a reference, they define the sample

size without referring to other sources.

Power analysis must be also used in defining the sample size (16, 19). In particular, if there are similar studies, the power of the study in question must be compared with the power of similar studies.

Another point about the sample size is that the researchers take fewer number of subjects than the planned ones, in order to prepare the paper sooner to the conference or the publication.

Study design

Some researchers do not have enough knowledge about study designs. If the researchers choose inappropriate study design, they will get the results with low presicion of estimation.

Each study has some advantages and disadvantages. Randomized, conrolled clinical trials are the most powerful designs possible in medical research, but they are often expensive and time-consuming well-designed observational studies are in contrary much quicker and less expensive. Cross-sectional studies provide a snapshot of a disease or condition at one time, and we must be cautious in inferring disease progression from them. Surveys, if properly done, are useful in obtaining current opinions and practices. Case-series studies should be used only to raise questions for further research (19).

Information about the variables

The researchers must have the adequate information examining the previous publications that they consider about the variables they will take or not take in the study. All possible sources of variation should be listed and controlled or measured to avoid their being confounded with relationships among those items that are of primary interest (10). Cause and effect relationship can be seen between some variables (20). If the researchers do not know this, they can make wrong interpretations by not examining the variables they should examine.

The risk factors researches of hip fracture may be given as an example to cause and effect relationship. It must be kept in mind that while researching the effect of both lack of calcium and osteroporosis on hip fracture, the lack of calcium (cause) is an important risk factor of osteroporosis (effect).

As an example to confounding factor, the study on the relationship between alcohol and lung cancer can be thought. When the result is significant the researcher will think that question; Has the smoking habit which is usually being used with alcohol been taken into account? If the smoking habit hasn't been taken into account, it can be thought as a confounding factor.

Heterogeneity of the groups

In case of having both control and treatment groups in the study, it is required to have homogeneity of the variables which are not being examined (21). If there are repeated measurements in the study, the baseline values must be homogeneous. If homogeneity is not provided, the statistical result at the end of the study may not reflect the actual situations since there are uncontrolled heterogeneous effects of control and treatment groups. Even if the experiment animals are the same race from the same environment, still there can be heterogeneity between the groups. So the homogenity of control and treatment groups must be examined at the beginning of the study.

An example on cardiovascular disease, related with the subject, can be given. If the family history factor on cardiovascular disease is being researched, there are two groups; the ones who has cardiovascular disease in her/his family and the ones who do not. In order to examine the effect of family history factor, the two groups must be homogenous in other factors like age, daily physical activities, and diet.

Data Collection and Pre-processing
Inappropriate measurements

Some of the researchers measure the variables with inappropriate methods. Data obtained this way may give unuseful or misleading results (12). A failure to reject may result from insensitive or inappropiate measurements, or too small a sample size (10).

For example while examining the effect of smoking to a disease, some researchers classify the subjects as smoker-nonsmoker. In this case how long the subject has been smoking, how much he/she smokes a day can not be observed. Here, it is more informative to observe the variable as packet-year in order to measure the duration of smoking and the amount of smoking (ie for a subject who had been smoking for 6 years and smokes 10 cigarettes a day the observation value would be 10*6/20=3). Different examples can be given for the situation.

Compiling the data

At the stage of compiling the data, firstly the data source should be decided, afterwards the inclusion of that data source, completeness and reliability should be examined carefully (12). At the stage of compiling (obtaining) data, one of the most common error occurs while using the data previously recorded and compiling them during secondary compiling. Reseachers may not find the exact variable they will examine in the recordings or they may find them measured in different scales. In that case the researchers may struggle to increase the number of data or try to change the structure of the data.

An example assume that the researcher has collected raw data for cholesterol values. If the researcher is obtaining the data from the records and if the values are noted as normal (143-200), some researchers may take the average values (171.5) of the lowest and highest limit values to use this record. That causes systematic error. To prevent this kind of errors, clear definitions of the variables should be made and be agree strictly until the end of the study.

Censored, truncated data

One of the most common error source in the studies is, some of the subjects' drop out the research or can not be getting knowledge from some of them at the stage of data collection. If there are that type of subjects in the data set, information about those subjects should be given and if they are in the evaluation, it should be mentioned at which stage they have been dropped out (9). Dropping out some subjects from the study is a factor that reduces the power which is aimed at the beginning of the study (22).

Converting the continuous data to categorical ones

In statistics the scales are examined in four classes; rational, interval, ordinal and nominal. Statisticians would like to study with the interval and rational scales because of their mathematical properties, however this is impossible in some cases. But some researchers categorize their data, convert them to nominal scale and analyse although they have data measured in interval or rational scale. Reducing the level of measurement in this way also reduces the precision of the measurement (23). That situation causes the loss of information and lead to wrong interpretations.

For example, instead of comparing the cholesterol values obtained before and after the use of the drug presumed to reduce the cholesterol, if categorizing the data as "low-normal-high" takes place, this may cause the loss of knowledge and valid results won't be obtained. Because the variations in the categories (low-normal-high) won't be taken into account.

Graphical demonstrations

The graphs are plotted to get a demonstrative knowledge about variables in data set. There are different types of graphics to demonstrate the distribution and the tendency of the variable in detail (24). Nevertheless most of the researchers do not know which graphic type is suitable for which type of data and objective, so they plot the graphs at random, resulting in wrong impression of the true nature of the data.

Graphic representations can be misleading, and large differences between groups that come with large variability might not be significant, no matter how they look (25). Another error made about the graphical exploration is that the researchers tend to change data set and the test they used, to match the results of the tests and the graphical display. It should be kept in mind that graphics give only subjective results.

Analysis Method and Implementation
Tendency of using the same analysis, method or test for similar studies

One of the most common errors that are made by the reseachers who do not consult a statistician is, if they are making a similar study with some previous ones, they have a tendency in using the same statistical analysis, methods and tests that are used in those previous studies (26). The statistical method that will be used for a certain data set is decided by examining some statistical criteria like the number of the data, the type of the scale, variability and theoretical distribution. The same method is not neccessarily used only since the subject of the research is the same.

Statistical softwares

If the researchers do not consult a statistician and if they don't have adequate statistics knowledge, one of the most common errors is the error sourced from the statistical softwares which makes the statistical analysis easier. After entering the whole data to a statistical software, the researchers who don't take statistical consulting choose an

analysis method which is convenient for them, regardless of the special features of the current project, and they obtain a p-value. Since they get a p-value, they think the analysis they made is true. It should be kept in mind that whatever the sample size is, whatever the scale is, whichever the data type is, or whichever the analysis type is, chosen statistical softwares give a p-value.

Sometimes different softwares may use different representation of the same model and if the researchers don't know about that, it may lead wrong interpretations. For example, for exponential regression in survival analysis, some software use the proportional hazards representation (λ $(t/z) = \lambda e^{\beta'z}$) and some others use log-linear model ($\log (T) = -\alpha + \beta*'z$) which results with opposite sign ($\beta^*=-\beta$). Sometimes the researchers must reproduce results because software programs might differ in how they do calculations, and different programs might give you slightly different results (25).

Making the comparisons independently from the baseline values in repeated measurements

One of the most common errors in repeated studies is made in the comparison of groups. While comparing the means of the groups, the statistical tests are conducted whitout taking into account the baseline values. Researchers directly compare the post test observations that are measured after the baseline observations. Nevertheless, those values are dependent to the baseline values. Comparisons conducted directly and after adjustment give different p-values. The percentage change (=[(last value)-(baseline value)]/(baseline value)) due to the baseline values should be taken into account for rational measurements which takes the 0 point to mention real absence and difference between the scores (=(last score)-(baseline score)) due to the baseline values should be taken into account for ordinal and interval measurements which takes the 0 as not a real absence.

The distribution of the percent change from the baseline is not normal; rather it's a Cauchy distribution whose moments such as mean and variance do not exist. As a result, no inference about the mean or variance can be made. In this case it is recomended that nonparametric methods be employed to obtain inference of the treatment effect based on the median (16).

Compelling the results for the expected ones

Some researchers get anxious about their results being different from the similar studies. In such cases researchers have an idea that the reason of getting those different results is the insufficient number of subjects and they increase the number of subjects until getting the same results or they take some of the subjects out of the study.

Another mistake that the researchers do about compelling the results is that researchers choose the test due to their expectations. In case of accepting the null hypothesis, researchers try the other test. This mistake is especially seen in post hoc comparison tests which are applied without taking account the criteria of use for those different tests after analysis of variance. In case of accepting the null hypothesis, researchers try the other one.

Interpretation
Statistical expressions

Some of the researchers do not mention the meaning of the numerical values they wrote. Some other researchers do not know what they should write and how they should write at the end of the study while they are interpreting the results. So misleading statistical expressions can be occured. Researchers should consult a statistician to check the statistical expressions they used before publishing the study.

Estimate of the population standard deviation and estimate of standard deviation of the sample means may be given as a simple example to that subject. Many researchers don't know the difference between the standard deviation and the standard error. In some studies in the literature, sample means are reported "±" a second value, is not indicated if the second value is a sample standard deviation, standard error, or some other measure of dispersion (27).

Another simple example is indicating the computer outputs of p value as 0.000 exactly the same with the output as 0.000. That can be misunderstood as the p value is equal to zero. In fact that is given as the output of the program as the number of the digits. So it must be given as <0.001.

Wrong descriptive statistics given in interpretation in case of missing data after the comparison of the dependent groups

During the comparison of dependent groups, if there are missing data for one or more variables, some statistical softwares conduct the analyse taking missing data out of the study and make comparisons. But some

researchers do not realise that, and in the interpretation of the dependent groups they take into account the descriptive values of the whole data set.

Contradictory interpretations about the significance test

Some researchers, although they found insignificant results, they use expressions such as, 'the result is not statistically significant but the mean of x is bigger than the mean of y'. This expression absolutely does not reflect the truth and has a contradictary with the significance test that is conducted. As a result of the statistical analysis, significance test denotes that there is not significant difference between two variables. That means the absolute difference between the two means is not significant. If the study is repeated with the same number of subjects and under the same conditions, the mean of x can be found smaller, although obtaining insignificant results again ($p > 0.05$).

CONCLUSION

It is seen that statistical mistakes in most studies in medical journals have been made (6, 7, 13, 17, 18, 25, 26). The errors that are made in the studies are mainly sourced from lack of statistical knowledge and not consulting to a statistician.

The errors on an original subject will cause worse results since there are only a few studies on that subject. Statistics must be used to not let scientists and whom will use the scientific findings in their lives to exposed to that kind of negativeness. But unfortunately some of the researchers are in a cycle of mistakes without noticing the statistical mistakes they have made.

In this study we investigated possible misuse of statistics at every stage of a research. We listed several error sources researchers easily commit if they are lack of solid statistical background. We emphasized the importance of statistical consulting from the very beginning to the end of a medical research.

Unbiased, consistent, and efficient parameter estimates are made in statistics science. This can be provided using statistics from the planning until the end of the study. So it is neccessary to consult statisticians at each stage of the studies.

REFERENCES

1. Sahai H, Ojeda MM. Teaching biostatistics to medical students and professionals: Problems and solutions. Int J Math Educ Sci Technol 1999;30(2):187–96
2. Altman DG. Statistics: Necessary and important. Br J Obstet Gynaecol 1986;93: 1–5
3. Altman DG. Statistics in medical journals: Some recent trends. Stat Med 2000;19:3275–89
4. Applegate KE, Crewson PE. An introduction to biostatistics. Radiolog 2002;225:318–22
5. Altman DG, Bland JM. Improving doctors' understanding of statistics. J R Stat Soc Ser A 1991; 154(2):223-67
6. McGuigan SM. The use of statistics in the British Journal of Psychiatry. Br J Psychiatry 1995;167:683-8
7. Welch GE II, Gabbe SG. Review of statistics usage in the American Journal of Obstetrics and Gynecology. Am J Obstet Gynecol 1996;175(5):1138-41
8. Comlekci N. Basic statistics principals and techniques (in Turkish). Eskisehir: Bilim Teknik Publications, 1989;13:17
9. Feinstein AR. Common faults in biostatistics in medical research. Scientific writing, editing and auditing in medicine symposium. Ankara: Tubitak Health Sciences Research Group, 1994: 49-55
10. Good PI, Hardin JW. Common errors in statistics (and how to avoid them). New York: John Wiley & Sons, 2003:3-4, 105
11. Serper O, Aytac M. Sampling (in Turkish). Istanbul: Filiz Publications, 1988:1-14
12. Sumbuloglu K, Sumbuloglu V. Consciously use of biostatistics principals and methods in scientific researches (in Turkish). Pfizer Ltd. Co, 2002:7-40
13. Williamson GR. Misrepresenting random sampling? A systematic review of research papers in the Journal of Advanced Nursing. J Adv Nurs 2003;44(3): 278-88
14. Kan I. Biostatistics (in Turkish). Bursa: Uludag University Publication, 1998: 55.
15. Tokol T. Marketing research (in Turkish). Bursa: Vipas Co, 2000:19
16. Chow SC, Liu JP. Design and analysis of clinical trials. New York: Wiley-Interscience Publication, 1998:93
17. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. Ann

Intern Med 2001;134(8):663-94

18. Halls JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. Surgery 1996;119:466-72

19. Dawson B, Trap RG. Basic & clinical biostatistics (3th ed.). Lange Medical Books/Mcgraw-Hill International Editions, 2001:21,71,107

20. Sumbuloglu V, Sumbuloglu K. Research methods in health sciences (in Turkish). Ankara: Hatiboglu Publications, 1988: 30

21. Devita VT, Hellman JRS, Rosenberg SA. Principles and practice of oncology (5th ed.). Philadelphia: Lippincott-Raven Publishers, 1997:519

22. Ellenberg JH, Armitage P, Chalmers TC, Gehan EA, O'Fallon JR; Pocock SJ; Zelen M. Biostatistical collaboration in medical research. Biometrics 1990; 46(1): 1-32

23. Lang T. Twenty statistical errors even you can find in medical research articles. Croatian Med J 2004; 45(4): 361-70

24. Ozdamar K. Statistical data analysis using softwares-1 (in Turkish). Eskisehir: Kaan Publications, 1999: 80-102,193

25. Murphy JR. Statistical errors in immunologic research. J Allergy Clin Immunol 2004;114:1259-63

26. Altman DG. Poor-quality medical research-What can journals do? The JAMA 2002;287(21): 2765-7

27. Holmes TH. Ten categories of statistical errors: A guide for research in endocrinology and metabolism. Am J Physiol Endocrinol Metab 2004;286: 495-501