Application of Bio-Statistical methods in Public Health Research

Mr.K.S.Sridhara

Assistant Professor cum Statistician, Vijayanagara Institute of Medical Sciences, Ballari, Karnataka.

Abstract
Objective:
To find the application of Bio-Statistical methods used in the field of Public Health research.
Methods:
Sampling methods used for data collection, graphical technique used for data presentation and Bio-Statistical methods used for data analysis in the International journal of Public Health of the WHO, is reviewed from the year 2010 to the year 2015.
Findings:
For data collection multistage stratified cluster sampling method is used by 13.79% and for presentation bar diagram with 95% confidence interval and trend line are the common procedure adopted. Multiple regression analysis technique is used by 50.57% of the researchers. It is found that 37.93% of the researcher used 95% confidence interval for their estimated value of mean, proportion or other estimates.
Conclusion:
Based on the findings of the study the topics of training in biostatistics need to include the methods applied by the Public Health researchers.

I.   Introduction

Biostatistics is considered as one of the important tool in medical research. It is applied in designing presentation and analysis of the research. Lifelong learning as a physician, demands facility in the assessment and application of clinical evidence from medical literature. Appraised of an article, methodological rigor depends on an understanding of the study design and analysis used by the author (1).

It is also not clear which statistical concepts are most necessary and useful to readers to become familiar with. There has been a well documented trend towards the use of new and increasingly complex statistical techniques in published article (2). An earlier study demonstrated that reader of pediatrics who understood descriptive statistics (for example mean and SD) and three inferential statistical procedure (student t test, chi-square test, Pearson's r) could understand the statistical analysis in 97% of research article published in 1952 but only in 49% of article inl982 (3).

The present study finds Bio-Statistical methods that are used for the research in the field of public health.

II.   Objectives

1. To find the common Bio-Statistical methods for data collection, data presentation and data analysis applied for the research in Public Health
2. To know the statistical software most commonly used for data analysis in Public Health research.

III.   Materials And Method

This study conducted by reviewing the research article published by the WHO's international journal of Public Health. These include About 87 research articles that used Bio-Statistical methods from the year 2010 to the year 2015.In each research article the sampling methods used for data collection, various graphical methods used for data presentation and different Bio statistical methods used for data analysis were compiled. After grouping the different methods percentage of use is calculated. The statistical software most commonly used for the research in Public Health was also listed.

IV. Results
Table 1: The list of journals reviewed month and year wise.

Sl.No.	Month and year
1	Jan 2015
2	Feb 2015
3	Apr 2015
4	July 2015
5	Aug 2015
6	Mar 2014
7	May 2014
8	Aug 2014
9	Oct 2014
10	Nov 2014
11	May 2013
12	Aug 2013
13	Oct 2013
14	Jan 2012
15	Aug 2012
16	Dec 2012
17	Nov 2011
18	Dec 2011
19	Feb 2010
Total no of article reviewed	87

Table 2: Showing the sampling methods used for data collection

Sampling methods	No. of research article	Percentage
Multistage cluster sampling method	12	13.79
Randomized control trial	04	4.59
Multilevel Poisson model sampling approach	01	1.15
Complex sampling method (stratified clustering and sample weights)	01	1.15
Retrospective ecological study	01	1.15
Population based retrospective cohort study	01	1.15
Non sampling method	21	1.15
Total		24.13

Sampling methods specified in 24.13% of the research articles. Among the sampling methods multistage cluster sampling method is used by 13.79% of the research, followed by Randomized control trial by 4.59% of the research.

Table No. 3: Showing the Graphical Methods used for the research.

Graphical Method	No. (%)
Multiple Bar diagram or component Bar diagram	10 (11.49)
Trend line	4 (4.59)
Bar diagram with 95% confidence interval	14 (16.09)
Trend line with 95% confidence interval	04 (4.59)
Comparison trend line with two variable on Y-axis	06 (6.89)
Multiple Bar diagram with trend line	02 (2.29)
Scatter plot diagram	03 (3.45)
Box plot diagram	01 (1.15)
Funnel plot graph	01 (1.15)
Receiver operating characteristic curve	01 (1.15)
Kaplan-Meir survival curve	03 (3.44)
Spot Map	01 (1.15)
Total	50 (57.47)

For graphical presentation either bar diagram or trend line is used by 16.08%. Trend line with 95% confidence interval and bar diagram with 95% CI is used by 21.68% of the research article. Comparison trend line graph by taking two variables on y axis is used by 6.89% while use of bar diagram and trend line together found in 2.29%.
Over all 57.47% used graphical presentation.

Table No. 4: Methods of Analysis

Analysis Methods	No.	%
Used statistical value of Mean, Median, percentage, proportion, Relative risk, odds Ratio, Incidence Rate, Prevalence Rate	18	20.68
Used statistical value mean. median etc + 95% confidence Interval	33	37.93
Statistical test of significance	43	49.42
Pearson correlation coefficient	02	2.29
Weight for age Z Score, weight for height Z Score, height for age Z Score, stunt Z < -2	04	4.59
Regression Analysis	44	50.57

Estimated value of statistic is presented as it is in 20.68% of the research but 37.93% of research presented with statistic + 95% confidence interval. About 49.42% uses statistical significance test and regression analysis is used by 50.57% of the research.

Table No. 5: Statistical test of Significance used for the study

Statistical Test	No.	Percentage
Pearson Chi-square Test	17	39.53
Fishers Exact test	06	13.95
Wilicoxen rank sum test	04	9.30
Kruskalwallis test	02	4.65
Person correlation coefficient	02	4.65
Wald's test	01	2.32
ANOVA test	04	2.32
Z -Test	01	2.32
Man- Whitney U Test	02	4.65
Student 't' test	04	9.30

Statistical test of significance usage is 39.53% with chi-square test and Fisher exact test is 13.95% of researchers. Application of new statistical test also found in 2.32 to 4.65% of the research article.

Table No. 6: Regression Method used for Data analysis.

Regression	No.
Multivariable regression model	14
Multivariate logistic regression model	11
Kaplan Meir survival analysis	04
Poisson regression model	03
Bayesian mixed effect regression	01
Bayesian random effect regression	01
Fuzzy regression discontinuity design	01
Lanier Mixed model	01
Maximum likelihood estimation	01
Least absolute shrinkage and selection operator model	01
Taylor series linearization method	01
Random effect model I^2 statistics	01
Principal component Analysis	01
Non parametric Boot strapping method	01
Generalized Binomial regression model	01
Logril model	01
Parametric survival model with Weibull distribution	01

The above table shows the regression methods of different type are used in 50.57% of the research.

Table 7: Showing the Statistical Software most commonly used for the Statistical Research

Statistical Software	No. (Percentage)
STATA SE ver. 11.1.12.10 (Stata Corp, Lp.College Station USA)	21 (24.13)
SAS ver. 4.5,9.3,9.2,9.1	08 (09.19)
R stat Package (R foundation for stat Comp. Vienna, Australia)	06 (06.89)
SPSS ver. 13.0 ver. 20 (SPSS Inc Chicago USA)	06 (06.89)
EPI info (Center for Disease Control and Prevention Atlanta USA,(11)	06 (06.89)
Minitab Stat Software ver. 14 (Minitab Ltd. Covenly, England)	01 (01.14)
Sample size PASS Software ver. 11 (NCSS LCC Kysville USA)	01 (01.14)

Table (7) shows statistical software used for the research. It is found 24.13% of the research's used STATA version. SAS version is used by 9.19% of the research and SPSS by 6.89%. Also R foundation for stat is used by 6.89% of the research.

V. Discussion

The application of Bio-statistical methods in the public health research is by reviewing the WHO's International Journal of Public health from the year 2015. A total of 87 journal article where Bio-Statistical methods applied is used. (Table 1)

In the study designing stage itself 24.13% of the research articles used different statistical sampling methods. Multistage sampling methods are used by maximum 13.79% of the researchers. Randomized control trial is used by 4.59%. Multilevel Poisson model sampling approach in 1.15%, complex sampling method (Stratified clustering and sample weights) in 1.15%. Other methods like retrospective ecological study, population based retrospective cohort study and non-sampling method also used by 1.15%.

Most common graphical methods used for presentation is by Bar diagram with 95% confidence interval (16.09%) followed by multiple Bar diagram or component Bar diagram by 11.49% of the research article. Trend lines with 95% confidence interval comparison trend line with two variables on Y-axis were also used. Other graphical methods used are tabulated in Table-3. Overall 57.4% of articles used graphical presentation.

For data analysis regression analysis is used by maximum times (50.57%) followed by statistical test of significance (49.42%) and used statistical value mean, median with 95% confidence interval by 37.93%. Application of mean, median percentage, proportion, relative risk, odds ratio, incidence rate, prevalence rate by 20.68%. All other methods are tabulated in Table-4.

Pearson Chi-square test is the most common test of significance 39.53% followed by Fishers Exact test by 13.95%. Other test like Wilicoxen rank sum test and student 't' test by 9.30%. Kruskal Wallis test, Pearson correlation coefficient and Man-Whitney U test by 4.65% in the study. Wald's test, ANOVA test and Z-test for testing the statistical significance used by 2.32% in the study. (Ref. Table-5)

Different methods of regression methods are used for data analysis. Multi variable regression model is the highest followed by Multivariate logistic regression model. Different types of regression methods are used for analysis by 50.57% of research. (Ref. Table-6)

Statistical software STATA SE ver. 10, 11.1, 12 is used by highest researchers i.e. by 24.13% and followed by 9.19% used SAS ver. 4.5, 9.1, 9.2, 9.3. R stat package, SPSS and EPI info software is used by 6.89%. Minitab stat software ver. 14, sample size PASS software ver. 11 is used by 1.14% of users.

VI. Conclusion

Bio-Statistical methods applied by the Public Health research show among the random sampling methods, multistage cluster sampling is used by 13.79% of the research article. Graphical methods used by the research are more of combination and comparison type. Both bar-diagram and trend line combination used for data presentation. Statistical estimates were used with 95% confidence interval and newer statistical test were applied.

Acknowledgement

Reference

[1]. Martha. A. Hellens, Mathew. J. GrkaGregony, Hayden, statistical literacy for readers of pediatrics, a moving target.
[2]. Heyden G.F. Bio-Statistical trends in pediatric implication for the future pediatrics 1983, 72:84-87.
[3]. Altman.D.G. Statistics in medical Journals some recent trends Stat Med 2000. 19; 3275-3289.