

Application of biostatistics in research by teaching faculty and final-year postgraduate students in colleges of modern medicine: A cross-sectional study

AD Gore, YR Kadam, PV Chavan, GB Dhumale

Department of Community Medicine, Bharati Vidyapeeth Deemed University Medical College and Hospital, Sangli, Maharashtra, India

ABSTRACT

Context: Biostatistics is well recognized as an essential tool in medical research, clinical decision making, and health management. Deficient basic biostatistical knowledge adversely affects research quality. Surveys on this issue are uncommon in the literature. **Aims:** To study the use of biostatistics in research by teaching faculty and postgraduate students from colleges of modern medicine. **Settings and Design:** Cross-sectional study in colleges of modern medicine. **Materials and Methods:** A pretested proforma was used to collect information about the use of biostatistics by teaching faculty and final-year postgraduate students from colleges of modern medicine. The study period was 6 months. **Statistical Analysis:** Chi-square test, Spearman rank correlation coefficient, and multivariate analysis were used for analysis of data. **Results:** With this questionnaire, the maximum possible score for appropriate use of biostatistics in research was 20. The range of scores obtained by the study subjects was 1–20 and the median was 11. Appropriate use of biostatistics was independent of sex, designation, and education ($P > .05$). Spearman coefficient showed low—but significant—correlation between the score and the number of papers presented and published ($P = .002$ and $P = .000$, respectively). **Conclusions:** The study showed that nearly half of the respondents were not using statistics appropriately in their research. There was also lack of awareness about the need for applying statistical methods from the stage of planning itself.

Key words: Awareness, biostatistical knowledge, PG students, teaching faculty

INTRODUCTION

Biostatistics is a branch of applied statistics and it must be taught with the focus being on its various applications in biomedical research.^[1] It is an essential tool for medical research, clinical decision making, and health management.^[2] Statisticians have long expressed concern about the slow uptake of statistical ideas by the medical profession and the

frequent misuse of statistics when these methods are used. On the other hand, doctors have been worried about the increasing pressure to make use of techniques that they do not fully understand.^[3] The biostatistical literacy of medical students is a problem all over the world.^[1]

Research is an important activity for not only postgraduate (PG) medical students but for all medical professionals. Deficient basic biostatistical knowledge adversely affects research quality. Inappropriate statistical methods, techniques, and analysis results in time and cost lost and, most importantly, from the perspective of scientific ethics, does harm to science and humanity.^[4] Writing on the teaching and learning of medical statistics in South Africa, Stander remarked that ‘medical practitioners were totally intimidated by the idea of statistics.’^[5] Surveys on this issue are uncommon in the literature.^[2]

This study was designed to find out the problems associated with biostatistical usage in research done by medical professionals in medical colleges. The aim of the study was to examine the use of biostatistics in research by

Address for correspondence: Mrs. Alka Dilip Gore, Department of Community Medicine, Bharati Vidyapeeth Deemed University Medical College and Hospital, Sangli, 416 414, Maharashtra, India.
E-mail: gorealkadilip@yahoo.com

Access this article online	
Quick Response Code:	Website: www.ijabmr.org
	DOI: 10.4103/2229-516X.96792

the teaching faculty and PG students of colleges of modern medicine.

MATERIALS AND METHODS

A cross-sectional study was conducted amongst all teaching faculty and final-year PG students from five colleges of modern medicine in three adjacent districts of the south-western region of Maharashtra state, India, from June 2010 to November 2010. Data collection was done using a pretested questionnaire. A pilot study was done to validate the questionnaire and the proforma was modified as necessary. Permission for data collection was taken from the deans of the respective medical colleges. Data was collected by paying a visit to final-year PG students and teaching faculties of every department. They were briefed about the study. Proforma were distributed and filled in proforma were collected.

Final-year PG students are required to finish research work on some topic before obtaining their PG degree. Hence, they were chosen for this study as they can be expected to have relatively better understanding of biostatistics than junior residents. Those who were willing to participate in the study were explained the nature of the study. Information was collected by using a pretested self-administered questionnaire that was designed to elicit information on personal and professional characteristics and knowledge of basic biostatistics. Those study subjects who were not available during the first visit were visited again and administered the proforma. Those who could not be contacted despite five visits, as well as those who failed to return the filled-in proforma, were excluded from the study. About 3–5 visits were paid to each college for collection of the data.

Scoring was based on the responses to 20 questions on biostatistical knowledge. There were 11 closed-ended and 8 open-ended questions, and the maximum possible score was 20. Study subjects were classified into four groups according to the score obtained, as follows: <25%, 25%–50%, 50%–75%, and >75%.

Data was analyzed by calculating percentages. The chi-square test was applied to check the association of sex, education, and designation with the score. Spearman rank correlation coefficient was used to check the degree of association between the score and age, teaching experience, and number of papers presented and published. Multivariate regression was used to get an advanced model for the highly significant independent factors and score. The analysis was done with the help of MS® Excel® and the trial version of SPSS® 17.

Ethical consideration

The institutional ethical committee approved this study.

We explained the nature and purpose of the study to the participants and assured confidentiality before obtaining voluntary informed consent.

RESULTS

Of the 600 proformas that were distributed, 310 filled-in proformas were returned, giving a response rate of 51.67%. Twenty-nine respondents (9.35%) failed to mention their designation, gender, and/or age. Among the 310 respondents, there were 46 (14.84%) professors, 43 (13.87%) associate professors, 122 (39.35%) lecturers, and 75 (24.19%) final-year PG students. The average age of the participants was 38.3 ± 11.06 years (range: 22–70 years). Among the 310 respondents, there were 175 males and 130 females [Figure 1].

Of the 310 respondents in the present study, 305 (98.39%) agreed that biostatistics is important for research. For 118 (38.06%) respondents biostatistics was easy to understand, while for 167 (53.87%) it was difficult. Of these latter 167 respondents, 16 (9.58%) said that all topics in biostatistics were difficult. However, 9 (56.25%) of these 16 respondents had not consulted a biostatistician for help with their research work despite facing problems with understanding biostatistics.

Two hundred and sixty-three (84.8%) respondents took the help of the statistician for data analysis, whereas 36 (11.6%) felt that such help was not necessary; 11 (3.5%) respondents did not answer this question. Only 97 (31.29%) respondents felt that the use of statistics is required from the stage of planning itself; the remaining respondents sought the help of a statistician after data collection, after collating the data in tabular form, or after analysis for interpretation and to check the significance of findings.

Half of the respondents (158; 50.97%) did not calculate

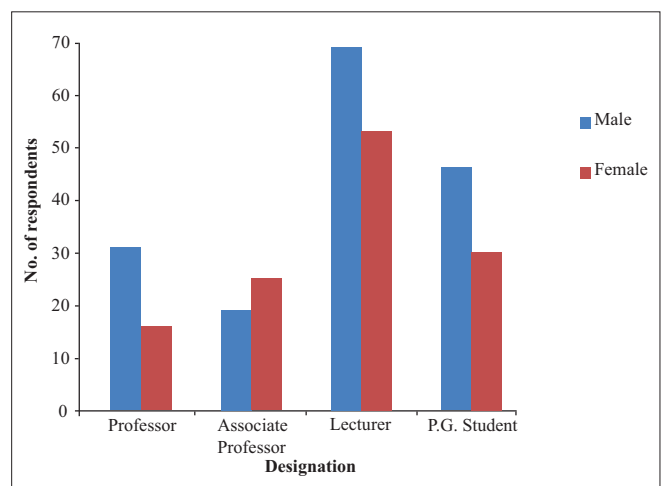


Figure 1: Distribution of study subjects according to gender and designation

sample size appropriately. These respondents used either all available study subjects or a figure of convenience (27.74% and 26.13%, respectively), and some (21.94%) decided the sample size according to previously published articles. Only 152 respondents (49.03%) made the effort to calculate sample size correctly, either by using a standard formula (13.87%) or by asking for the help of statistician (35.16%). Thirteen (4.19%) respondents did not answer the question related to calculation of sample size.

Various options were chosen by subjects in response to the question on the factors upon which data analysis depends: namely study design, sample size, type of data, and aim and objectives. Only 124 (40%) of the respondents mentioned all the factors that can influence data analysis. Twelve (3.87%) respondents did not have any knowledge about this. They responded as 'don't know'.

The most commonly mentioned use of a test of significance was 'to find out the association' and in general the respondents had very little knowledge about the other uses of test of significance. Three (0.9%) respondents had no idea whatsoever about the uses of tests of significance, and 16 (19.4%) did not respond to the question at all. None of the respondents was able to mention all the applications of tests of significance [Table 1].

The majority of the respondents (172; 55.5%) were unaware about different sampling techniques, and those who claimed about biostatistical knowledge, could not mention the various sampling techniques correctly. Irrelevant names of sampling techniques were given by 45 (14.51%) respondents, means they

were totally unaware about sampling techniques. 74 (23.87%) mentioned the correct names, and 191 (61.61%) could not mention any of the names also.

Two hundred and three (65.5%) of the respondents admitted to preparing dummy tables in their research project. Two hundred and sixty-five (85.5%) of the respondents felt that they would need the help of a statistician for proper presentation of data, whereas the remaining respondents considered themselves capable of doing this without help.

Table 1: Various factors for which medical professionals seek the help of the statistician

Factors	Number of study subjects taking the help of statisticians	
	n (%)	95% CI
Stages		
At planning level	97 (31.29)	26.02 to 36.55
After data collection	125 (40.3)	34.75 to 45.89
After preparing tables	52 (16.8)	12.53 to 21.03
After analysis	63 (20.3)	15.75 to 24.89
Dependent factors		
Study design	61 (19.68)	15.16 to 24.19
Sample size	88 (28.39)	23.27 to 33.51
Type of data	76 (24.52)	16.63 to 29.40
Aim and objectives	61 (19.68)	15.16 to 24.19
All above factors	124 (40)	34.44 to 45.56
Don't know	12 (3.87)	1.68 to 6.06
Uses of tests of significance		
To find association	111 (35.81)	30.36 to 41.25
To find difference	75 (24.19)	19.33 to 29.06
To find the effect	31 (10)	6.59 to 13.41
To find the relationship	83 (26.77)	21.74 to 31.80
Don't know	3 (0.9)	-0.14 to 2.08

Table 2: Score of study subjects according to personal characteristics

Personal determinants	Score				Total	Significance
	< 25%	25%– 50%	50%– 75%	> 75%		
Age (years)						
20–30	4 (4.0)	35 (35.4)	55 (55.6)	5 (5.1)	99 (100)	$\chi^2 = 20.671$ $P = .014$
30–40	10 (10.2)	40 (40.8)	43 (43.9)	5 (5.1)	98 (100)	
40–50	6 (9.5)	24 (38.1)	21 (33.3)	12 (19.0)	63 (100)	
>50	3 (6.7)	22 (48.9)	16 (35.6)	4 (8.9)	45 (100)	
Sex						
Male	12 (6.9)	79 (45.1)	71 (40.6)	13 (7.4)	175 (100)	$\chi^2 = 4.53$ $P = .21$
Female	11 (8.5)	43 (33.1)	64 (49.2)	12 (9.2)	130 (100)	
Education						
MD/MS	12 (6.8)	74 (42.0)	73 (41.5)	17 (9.7)	176 (100)	----
PhD	0 (0)	3 (25.0)	9 (75.0)	0 (0)	12 (100)	
MSc	1 (16.7)	1 (16.7)	4 (66.7)	0 (0)	6 (100)	
Diploma	3 (9.1)	20 (60.6)	9 (27.3)	1 (3.0)	33 (100)	
PG Student	4 (8.3)	11 (22.9)	28 (58.3)	5 (10.4)	48 (100)	
Designation						
Professor	4 (8.7)	18 (39.1)	15 (32.6)	9 (19.6)	46 (100)	$\chi^2 = 12$ $P = .22$
Assoc Prof	4 (9.3)	20 (46.5)	15 (34.9)	4 (9.3)	43 (100)	
Lecturer	9 (7.4)	47 (38.5)	59 (48.4)	7 (5.7)	122 (100)	
PG student	6 (8.0)	27 (36.0)	37 (49.3)	5 (6.7)	75 (100)	

Standard deviation (SD) is a measure of dispersion. It measures the degree variation in the data. The majority of the respondents (197; 63.55%) mentioned the correct meaning of standard deviation. Of the 310 respondents, 53 (17.1%) said that SD is a measure of central tendency, 11 (3.55%) stated that it is a measure of skewness, and 47 (15.16%) respondents did not even answer the question.

In this study, we scored each respondent for appropriate use of biostatistics. The maximum possible score was 20. The range of the scores obtained by the respondents was 1–20, and the median score was 11. We found that the score was independent of designation ($P=.22$); however, higher scores were obtained by professors than by associate professors and lecturers. The score of PG students was high in comparison to that of MD or MS degree holders, diploma holders, and MSc holders. Female respondents scored more than males, though the difference was not statistically significant ($P=.21$) [Table 2].

The Spearman rank correlation coefficient was calculated for different parameters, including age, years of teaching experience, and number of papers presented and published. A very low (nonsignificant) degree of correlation was found between score and age. There was low but significant correlation of score with number of papers presented and published ($R = 0.002$ and $R = 0.000$, respectively) [Table 3].

Personal and professional determinants, which were significantly associated with score ($P<0.01$); were considered for binary logistic regression. Wald's backward method was used to

Table 3: Spearman rank correlation coefficient for score and other parameters

Parameters	Cor coe	P value
Age	-0.003	0.960
Experience of teaching postgraduates	0.118	0.060
Experience of teaching undergraduates	0.107	0.067
Number of papers presented	0.174	0.002
Number of papers published	0.208	0.000

Cor coe: Correlation coefficient

Table 4: Logistic regression

Parameter	B	S.E.	Wald	df	Sig	Exp (B)	95% CI for Exp (B)
No. of papers published	0.271	0.157	2.960	1	0.085	1.311	0.963 - 1.784
Education			14.116	5	0.015		
MD / MS	-1.006	0.493	4.163	1	0.041	0.366	0.139 - 0.961
PhD	-0.978	0.381	6.587	1	0.010	0.376	0.178 - 0.794
MSc	-0.269	0.788	0.116	1	0.733	0.764	0.163 - 3.583
Diploma	-0.501	0.945	0.281	1	0.596	0.606	0.095 - 3.864
PG student	-1.845	0.521	12.558	1	0.000	0.158	0.057 - 0.438
Constant	0.889	0.342	6.758	1	0.009	2.433	

B: Coefficient of regression; SE: Standard error of mean; Wald: Wald statistic; df: degree of freedom; Sig: significance, Exp (B): Odds ratio

find out the most significant factors. Education, experience in teaching undergraduates, and number of paper publications were the significant factors at this level. Logistic regression showed that the score was highly dependent on the level of education of the respondents ($P=.009$ for PG student and $P=.01$ for PhD) [Table 4].

In this study only 9 (2.9%) respondents gave the correct meaning of 'P value,' 164 (52.9%) could not give the correct answer, and 115 (37.10%) did not respond to the question at all. More than half of the respondents (204; 65.81%) felt that the results of their research project need not be positive or concordant with that of the references used, while 43 respondents (13.87%) felt that the results should agree with that of the references mentioned. Two hundred and forty-seven (79.68%) respondents said that they wished to upgrade their knowledge, whereas 18 (5.81%) did not want to upgrade it.

DISCUSSION

Of the 600 distributed proformas, 310 filled-in proformas were returned, a response rate of 51.67%. This is relatively high in comparison to other studies; for example, in the study by Khan *et al.* the response rate was only 44.7%, and in the study by Laopaiboon *et al.* the response rate was 40.0%.^[6,7]

It is important to understand biostatistical concepts to read the literature intelligently. The majority of the respondents in this study (305; 98.39%) agreed that biostatistics is important for research. Swift *et al.* and Windish *et al.* found that 79% and 95%, respectively, of the participants in their studies considered statistics as important for their work.^[8,9] According to 118 (38.06%) respondents in our study, biostatistics was easy to understand, but for 167 (53.87%) it was difficult subject. Windish *et al.* mentioned that 75% of their respondents did not understand all of the concepts in statistics.^[9] This difference from our findings regarding the understanding level may be because they considered only residents in their study, whereas we included final-year PG students as well as teaching faculty

members. Seventy-seven (46.1%), respondents who found biostatistics difficult mentioned analysis, calculation, application of tests, or advanced biostatistics as complex topics; an equal number of respondents did not specify the difficult topics. Twenty-one of the respondents (6.77%) did not respond to the question.

Teachers of medical statistics have recommended that the focus should be on interpretation and understanding of concepts, and that mathematical formulae and calculation must be kept to a minimum.^[10-12] Doctors engaging in research are expected to perform statistical analyses themselves or consult with a statistician right from the beginning of the research project.^[13] Two hundred and sixty-three (84.8%) respondents in this study said that they took the help of the statistician for data analysis. The respondents gave various responses to the question on the stage at which they would seek a statistician's help. Doctors' statistical training needs may have changed due to advances in information technology and the increasing emphasis on evidence-based medicine.^[13]

Biostatistical methods make research scientific if they are used from the stage of planning of the research itself. Unbiased, consistent, and efficient parameter estimates are provided by correct use of statistics. This is possible by applying statistics from the planning stage until the end of the study. So it is necessary to consult statisticians at each and every stage of the study. Only 97 (31.29%) respondents in this study felt that the use of statistics is required from the stage of planning of the proposal, the remaining respondents felt that the help of a statistician is required only after data collection is completed, after tabulating the data, or after analysis—for interpretation and to check the significance of findings. Those who would not seek the statistician's help from the stage of planning seemed to be more interested in the 'P value.'

The respondents mentioned various reasons for not seeking a statistician's help, of which the most common were lack of awareness regarding the need for consulting a statistician from the beginning of the research and the nonavailability of a statistician at their institute. Some of the respondents mentioned that they would be capable of doing it themselves by referring to books and the internet and by discussion with colleagues. Harry Robinson *et al.* found in their study that students who preferred learning by self-instruction did as well or better in terms of exam grades than their colleagues taking lectures.^[14]

Actually researcher have to calculate sample size appropriately, either himself/herself or with the help of statistician by examining previous studies (i.e. references or review of

literature), with suitable error; with certain significance level and suitable power of the test; but some researchers take 25, 30, 50 or 100 as the sample size without referring to other studies. In this study, half of the respondents (158; 50.97%) did not calculate sample size appropriately. Only 152 respondents (49.03%) calculated sample size correctly, either by using standard formulae (13.87%) or with the help of a statistician (35.16%). The subject of the study, the characteristics of the population, the length of the research, and the cost of the research must all be taken into account when deciding the sampling technique. Unfortunately, irrespective of the demand of the study design, some researchers use simple random sampling technique, without thinking, as they know only this method.^[4] The majority of the respondents (172; 55.5%) were unaware of the different sampling techniques, and those who said they were aware, did not mention the sampling techniques correctly. Two hundred and sixty-five (85.5%) respondents felt that they would need the help of a statistician for the presentation of data, whereas the remaining felt that they were capable of doing it themselves.

Internal medicine residents had low scores in a test of knowledge of biostatistics, and about three-fourths of the residents surveyed indicated that they were not confident about their understanding of the statistics they encountered in medical literature. The poor knowledge of biostatistics and difficulty experienced in interpretation of study results among the residents in the study likely reflects insufficient training.^[15]

The score of respondents in this study was independent of designation; however, higher scores were obtained by professors compared to associate professors and lecturers. The score of PG students was high in comparison to that of MD or MS degree holders, diploma holders, and MSc degree holders. This may be due to the fact that the PG students were currently involved in research for their dissertation. The score of female respondents was more than that of males; however, the observed difference was not statistically significant. Khan *et al.*, have also reported that gender did not show any significant effect on responses.^[6] Windish *et al.* reported higher scores for male respondents,^[9] whereas Asif *et al.* found that females had higher scores.^[16]

The Spearman rank correlation coefficient showed that the senior teaching faculty members had lower scores than the younger faculty members; the seniors claimed that this was because they were not taught biostatistics as a part of their undergraduate curriculum. The remaining parameters like number of years of teaching experience and number of research papers presented and published had only low degree of correlation with the score. There was low but significant correlation of the score with the number of papers presented and published. This may be

due to the fact that scientifically correct research papers, wherein appropriate statistical methods are applied, are more likely to be published than those that lack appropriate application of statistics.

Most researchers are interested mainly in deriving the P value, without having a clear understanding of its meaning. In this study also, only 9 (2.9%) respondents could give the correct meaning of 'P value.' One of the most common errors made by the researchers who do not consult a statistician is that, when conducting a study similar to a previous published study, they tend to use the same methods of statistical analysis and the same tests that were used in the previous study.^[15] This reveals an indifference on the part of the researchers towards statistics and also research as a whole.

From the above observations it is evident that the majority of the teaching faculty and postgraduate students do not apply biostatistical concepts in a scientific manner while conducting research. Although they are aware that the proper use of biostatistical methods is important for scientific research, they lack the required knowledge. Most of the respondents in the present study wished to upgrade their knowledge of biostatistics and suggested refresher training programs, workshops, Continued Medical Education, and self-learning as the means of achieving this. Many respondents were reluctant to fill up the proforma and preferred to leave it blank. Improvements in teaching statistics to medical students should improve their understanding of statistical concepts and reduce the incidence of misconceptions among clinicians and medical researchers.^[17] The poor knowledge of biostatistics and the consequent difficulty faced when interpreting study results among study subjects in the present study reflects insufficient training. Nearly one-third of the study subjects indicated that they never received biostatistics teaching at any point in their career and suggested the need for more effective training in biostatistics in undergraduate or postgraduate education. Zuger had also reported similar findings.^[18] To conclude, it is essential for medical professionals to upgrade biostatistical knowledge frequently to improve research quality.

REFERENCES

1. Sami W. Biostatistics education for undergraduate medical students.

- Biomedica 2010;26:80-4.
2. Adeleye OA, Offili AN. Difficulty in understanding statistics: Medical students' perspectives in a Nigerian University. *Int J Health Res* 2009;2:233-42.
 3. Altman D, Bland JM. Improving doctors understanding of statistics. *Stat Soc* 1991;154:223-67.
 4. Ercan I, Yazıcı B, Yang Y, Özkaya G, Cangur S, Ediz B, *et al.* Misusage of statistics in medical research. *Eur J Gen Med* 2007;4:128-34.
 5. Stander I. Teaching conceptual vs theoretical statistics to medical students. International Statistical Institute, 52nd Session; 1999. Available from: <http://www.stat.auckland.ac.nz/~iase/publications/52stan0219.pdf> [Last accessed on 2011 Dec 29].
 6. Khan N, Mumtaz Y. Attitude of teaching faculty towards statistics at a medical university in Karachi, Pakistan. *Pakmedinet* 2009;21:166-71.
 7. Laopaiboon M, Lumbiganon P, Walter SD. Doctor's statistical literacy: A survey at Srinagarind Hospital, Khon Kaen University. *J Med Assoc Thai* 1997;80:130-7.
 8. Swift L, Miles S, Price GM, Shepstone L, Leinster SJ. Do doctors need statistics? Doctors' use of and attitude to probability and statistics. *Stat Med* 2009;28:1969-81.
 9. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA* 2007;298:1010-22.
 10. Freeman JV, Collier S, Staniforth D, Smith KJ. Innovations in curriculum design: A multi-disciplinary approach to teaching statistics to undergraduate medical students. *BMC Med Educ* 2008;8:6920-8.
 11. Evans SJ. Statistics for medical students in the 1990's: How should we approach the future? *Stat Med* 1990;9:1069-75.
 12. Campbell MJ. Statistical training for doctors in the UK. Sixth International Conference on Teaching Statistics Cape Town, South Africa; 2002. Available from: http://www.stat.auckland.ac.nz/~iase/publications/1/4f3_camp.pdf [Last accessed on 2011 Dec 26].
 13. Miles S, Price GM, Swift L, Shepstone L, Leinster SJ. Statistics teaching in medical school: Opinions of practising doctors. *BMC Med Educ* 2010;10:75.
 14. Robinson H, Burke R, Stahl SM. Self-instructional teaching of biostatistics for medical students. *J Community Health* 1976;1:249-55.
 15. Altman DG. Poor-quality medical research-What can journals do? *JAMA* 2002;287:2765-7.
 16. Asif H, Asim B, Awais SM. Importance and understanding of biostatistics among post graduate students at king edward medical university Lahore – Pakistan. *Annals* 2009;15:107-10.
 17. Mahmood Z. Uses and abuses of biostatistics in medical research in Pakistan. *J Pak Med Assoc* 1990;40:270-1.
 18. Zuger A. Survey finds significant statistical insecurity: Most physicians have no confidence in their own ability to use medical statistics. *J Watch Gen Med* 2007 Aug;82:939-43.

How to cite this article: Gore AD, Kadam YR, Chavan PV, Dhumale GB. Application of biostatistics in research by teaching faculty and final-year postgraduate students in colleges of modern medicine: A cross-sectional study. *Int J App Basic Med Res* 2012;2:11-6.
Source of Support: Nil. **Conflict of Interest:** None declared.