A Teaching Experience in the Medical Department: From Concepts to Computation

Sunil K. Dhar
Department of Mathematical Sciences, Center for Applied Mathematics and Statistics, New Jersey Institute of Technology, Newark, NJ 07102, USA

Abstract: This work is based on the experience of teaching a biostatistics course to a group of medical professionals, including doctors, lab technicians, and medical students, with a minimal or rusty statistics background. The main goal of the course was to teach how to communicate in the language of statistics so that scientists can carry out their measurements, collect data according to a designed experiment, choose appropriate statistical analytic tools, and become proficient in statistical computation and interpretation of results. The topics that were covered and the computational resources that were used are discussed, with examples. The research is used to explore effective teaching tools and topics that are best suited for teaching statistics to those involved with the medical professions.
1. **INTRODUCTION**

Teaching statistics to medical research professionals is of importance for multiple reasons. Understanding statistics is needed for communicating the design of experimental trials in a lab, knowing the different available and appropriate methods of data analysis, accurately interpreting the results (a goal common with Svensson [1] and of significance to the investigators [2]). For example, application of an appropriate method affirmed the inverse relationship between fluoride concentration and dental caries, whereas no such relationship had been previously noted using the same data source [3]. One of the key steps in choosing the appropriate statistical methods is analyzing the measurement scales of the variables involved in data analysis [4]. It is quite well known that biostatistics is an important part of evidence-based medicine (EBM) [5]. The use of EBM as a motivating force for teaching biostatistics is a great asset [6] as evident in [7]. Cohn, et al. [8] note that in the field of quantitative nursing research, quality journal articles are statistically sound and have concrete foundation for evidence-based practice. Tantawi [9] notes that comprehension of biostatistics is important in literature evaluation and evidence-based practice in dentistry as well as for researchers wishing to have their publications accepted by international journals. The [10] compares standalone teaching method with integrated teaching method which incorporates clinical problem solving and journal clubs that give this type of teaching an edge over the former. The method of reading journal articles is used as an effective teaching technique in [12], this technique has also been implemented in [11] with great success. The article [13] suggests using a number of case study examples as a method of teaching biostatistics. The practice of applying biostatistical skills to problem
solving in the field of cell biology and molecular medicine will also help improve the quality of future research in the area [14] and [15]. Simpson [16] and [17] suggests using non-mathematical methods for teaching medical statistics unless displaying formulae is really essential. This strategy was also adopted for teaching the present course. Colditz et al. [18] surveyed 760 articles from a leading general medical journal to identify frequently used topics in biostatistics. Similar surveys to determine topics of importance in the field of biomedical informatics were conducted from the Journal of American Medical Informatics Association and the International Journal of Medical Informatics [19]. This trend of researching for topics in a field based on frequency in journal articles is a sound practice. A similar research could be conducted in the field of cell biology and molecular medicine. Some topics that one can guess based on experience of publishing in the field are design of experiment, randomization and survival analysis. Also, deepening the concepts and understanding requires a sustained effective teaching effort. There is no doubt that acquiring biostatistical skills is important for medical professionals. For example [20] suggests that statistics should be integrated over the entire length of the medical school curriculum. General Medical Council, UK, also requires doctors to have probability and statistics skills [21].

Biostatistics has been taught in American medical schools for over three decades [22]. However, in the last decade statistics courses have become better organized well-designed and contemporary [23]. While a history of teaching probability and statistics in years 1988 to 1994 can be found in [24], the future of biostatistical science and biostatisticians has been discussed in [25] and [26]. The latter also gives a list of major topics appearing in
Biometrics journal since the 80’s. There are many courses that have taught biostatistics in the traditional way for example [27]-[29]. Sahai [30] and [31] discusses some of the challenges faced in teaching biostatistics in medical and health sciences from the perspective of students as well as teachers. Means of communicating with students using internet resources such as blogging is discussed in [32], while and web-based resources for the analysis and interpretation of data can be found in Mullee [33].

Teaching biostatistics to help researchers arrive at the right conclusion after correctly designing and collecting data from an experiment is a lofty goal [17]. Accordingly, there is a demand for dedicated biostatistician, as observed by DeMets et al [34]. This demand is not limited to United States or other industrialized nations, but as the article [35] discusses, the need to produce the next generation of biostatisticians is also in the sub-Saharan Africa, as well as in other parts of world. Several methods of teaching biostatistics at the international level are discussed in [36].

Section 2 discuss the background of the participants taking the course. The format and content of the course are described in Section 3, together with references that can serve as a resource for teaching the course. A short discussion in Section 4 concludes the paper.

2. PARTICIPANTS

The student body taking the course comprised of participants with diverse backgrounds, such as financial managers, medical doctors, researchers, lab technicians and medical students studying cardiovascular cell biology and molecular medicine. All had some basic knowledge of statistics, but needed to know how to implement the statistical procedures
and deepen their understanding of statistical concepts. The class size was limited to 30 participants to allow for more interaction between participants and the instructor. The diversity in student background or in the level of understanding statistics did not create any problems in class. Classes were held for an hour and a half on a weekly basis during summer.

3. FORMAT AND COURSE OVERVIEW

The course titled “Medical and Health Statistics” was offered twice a week May through August, 2010. The course started with introductory statistics at the undergraduate level. The first quiz on basic concepts (please see Table 2 below) showed the lack of basic understanding among the participants who appeared to be biostatistically insecure. Similar observations were noted in [37] and [38]. The aptitude of students and problems faced in teaching such courses is well documented in [17]. It has been noted a point estimated of 11.8% orthodontic students failed to identify the appropriate Chi-squared test [39]. Based on the initial response from the students, the course was reset to a slower pace with greater focus on computation but sustained emphasis on concepts. The topics covered included: Population and sample, Basic statistics [variables, descriptive statistics], Stem-and-leaf plot, Distributions [data displays, normal, binomial], Estimation [point, interval], Hypothesis testing [level of significance, power, z-test, paired and two-sample t-test, F-test, $\chi^2$-test, p-value, parametric versus nonparametric], Multiplicity, Sample size computation, including Wilcoxon tests, Chi-squared Goodness-of-fit test and the test for homogeneity of populations, Correlation versus causation and regression, Randomization and bias, Basic
survival analysis, Effect size, Relative risk, Odds ratio, Proportional hazards, most of these topics are as in [11, p. 227, Table 1, modules 1-3]. The topics from the Table 1 below were covered next. Testing and confidence interval about the conditional mean of Y given $X_0$ and Prediction interval of Y given $X_0$ were covered using [40]. These topics covered so far, include all the top ten topics that are required in biostatistics courses as mentioned in [17]. Description of the source of these topics and the corresponding material are as in [11, p. 226-228, Section Format and Course Overview]. A take-home assignment on simple regression was given [40, problem 13, page 81], which involved interpreting a SAS output, the same output and model improvements were also demonstrated using Excel. The One and two-way repeated measure analysis of variance and chi-squared Cochran-Mantel-Haenszel test for comparing two treatments for given count data for a particular disease outcome with adjustment for the confounding variable were covered next using [41]. Multiple comparison procedures were covered using both [40] and [41]. The former gives a flow chart describing which multiple comparison procedure one could use when. Examples were taken from published research articles from the area of cell biology and molecular medicine to demonstrate the use of regression, equality of variance tests in ANOVA and the log rank test, which compares two survival curves. The tests for normality and qq-plot were taught using the R program. This program was also used to run the Wilcoxon tests. Use of Excel to implement simple regression and one-way analysis of variance was demonstrated. Such use of software has also been encouraged by [42]. The topics covered here include almost all (99%) of the ones listed in [43, Table 1] and almost all the topics in Indiana syllabus [44].
Importance of randomization in lab experiments that validates the independence assumptions for large sample size was emphasized (Table 2 question 3). In fact the lack of proper attitude towards randomized control trials among physicians has been observed in [45].

In the field of pediatrics a trend of using increasingly new and complex statistical techniques has been noted [46] and in term of its increased usage of multiple-regression by physicians has been noted by [47]. These, and several other topics, including independent Bayes method [48] that helps make decision using the probabilistic approach, could have been covered. However, due to lack of time only the basic topics could be covered.

4. DISCUSSION

Teaching biostatistics to medical students in order to help them decipher hard scientific data and translate it into real world medical application is a rewarding goal [17]. To achieve this goal one needs motivation, dedication, and perseverance as my own experience and that of other professionals (i.e. [49]) reflects. A survey of various effective teaching methods was given in the Introduction. It is important to note that when teaching medical statistics to researchers, it is not sufficient to give examples from the broad area of biostatistics. The key is to pull from practical examples or case studies particularly from scholarly journal articles as one of the vital steps in applying medical statistics is to be able to actually interpret and critique the quantitative results in relevant existing medical research, which was found important from the student’s perspective [50]. Additionally, the
use of EBM as a teaching tool is very effective, along with collaborative journal clubs and
discussion groups, which can keep lessons engaging, harmonious, and current with daily
medical progress via emerging studies and statistics. A method for teaching could also be
problem-based, where an unresolved issue or current problem is given and an open-ended,
interactive process is used to determine a resolution backed with data and statistic, as
mentioned in [51and 52]. Here, although the department of cell biology and molecular
medicine required participants to be taught effective statistical computing, as mentioned in
[53] emphasizing statistical concepts is essential for rationalizing the choice of appropriate
methods to be computed for analysis purposes. At the end of the day, regarding statistical
education, it is important to keep in mind that though biostatistical research is based in
number crunching and heavy data analysis, it is the overall concepts in biostatistics and
translation to action that makes this subject utilizable.

ACKNOWLEDGMENTS

I wish to thank Dr. Stephan Vatner for funding and providing logistical support for the
biostatistics teaching. I also wish to thank Ms. Radha Dhar for the help with the
compilation of literature search.

REFERENCES

1. Svensson E. Important considerations for optimal communication between
statisticians and medical research in consulting, teaching and collaborative research,
with a focus on the analysis of ordered categorical data. *Training Researchers in the
Use of Statistics*, 2001; 23-35.

2. Deutsch R. A seminar series in applied biostatistics for clinical research fellows,
Teaching in the Medical Department


11. Dhar SK, Kianifard F. Teaching statistics to clinical research staff in a
Teaching in the Medical Department


30. Sahai H. Some Comments on Teaching Biostatistics in Medical and Health

32. Tantawi MMAE. Blogging in a biostatistics and research design graduate dental course: For learning or interaction? *Journal of Dental Education* 2010; 74:410-416.


38. Windish DM, Huot SJ, Green ML. Medicine Residents’ Understanding of the biostatistics and results in the medical literature. *Journal American Medical*
Teaching in the Medical Department

Association 2007; 298(9):1010-1022.


42. Lee TR. Teaching Biostatistics to Medical Personnel with Computer Based Supplement. Training Researchers in the Use of Statistics 2001; 139-145.


48. Crichton NJ, Emerson PA. A probability-based aid for teaching medical


Table 1. Description for the syllabus for the Medical and Health Statistics course

- Testing and confidence interval about the conditional mean of $Y$ given $X_0$
- Prediction interval of $Y$ given $X_0$
- Fisher’s exact test to compare two populations with count data over several classes
- Log-rank test
- McNemar’s test for comparing paired binary data
- Cochran-Mantel-Haenszel test for comparing two treatments
- One and two-way Repeated measure analysis of variance
- Multiple comparison procedures
- Tests for checking equality of variances
Table 2. Administered and graded quiz 1.

<table>
<thead>
<tr>
<th>June 7, 2010</th>
<th>Medical and Health Statistics, Quiz 1</th>
<th>Name____________</th>
</tr>
</thead>
</table>

All problems are worth 1 point each unless specified otherwise.

In problems 1-2 (2 points each) circle the best answer.

1. A data set consisting of a random sample is necessary because:
   a. It will represent the population.
   b. To validate applicability of statistical methodology.
   c. To reduce the risk due to subjective selection bias.
   d. All the above.
   e. None of a, b, or c above.

2. When measuring a large set of measurements:
   a. The measurements can be close to the true value but need not be precise because we can adjust for the variability.
   b. The measurements can be biased but they have to be precise because we can correct for biased measurements.
   c. The measurements can be biased and not precise because we can correct for both biased measurements and for variability.
   d. The measurements must be unbiased and precise.

3. In laboratory experiments when comparing two treatments and a control, it is statistically justified to just select mice and assign them to the various treatments (which include control) because the mice provided to the lab are at random. Circle one True or False. Explain.

4. It is always true that parameter is the average value of the variable. Circle one True or False. Explain.

5. Interval scale data can be reduced to ordinal scale data and ordinal scale data can be further reduced to nominal scale data. Circle one True or False. Explain.

Consider the urinary concentration of lead in 8 children from a urban housing estate 0.6, 2.6, 0.1, 1.1, 0.4, 2.0, 0.8, 1.3. This data is applied to problems 6 - 9 below.

6. Compute the median

7. Show the method to compute the mean

8. Is the data symmetric, negatively skewed (long tail to the left) or positively skewed? Explain. (2 points)

9. The sample standard deviation computes out to be 0.839. Show how one would compute the estimate of the standard error of the sample mean.

10. Standard error of the sample mean is the amount of standard error made by the sample mean in estimating the population mean. Circle one True or False. Explain.

11. In case the data set follows a bell-shape or mount-shape distribution then within plus or minus one standard deviation of the mean there is about 95% of the data. Circle one True or False. Explain.

12. In case the data set follows a bell-shape or mount-shape distribution then a data value that lies outside plus or minus three standard deviations from the mean can be taken to be an outlier. Circle one True or False. Explain.

13. Binomial distribution can be used on any data set with “success” or “failure”, two category type data. Circle one True or False. Explain.

14. The standard error of the difference of two sample means is in terms of the square-root of sum of square of their individual standard errors. Circle one True or False.

15. A 95% confidence interval for $\mu_A - \mu_B$ is given by (-8.9, -1.10) can be interpreted to state that there is 95% probability that $\mu_A - \mu_B$ is in (-8.9, -1.10). Circle one True or False. Explain.

16. Power is the probability of not rejecting the null hypothesis when the null hypothesis is true. True or False. Explain.

17. To use the paired t-test the data sets for comparing the two treatments (before and after) should be independent of each other. Circle one True or False. Explain.