INTEGRATING REAL MEDICAL STUDIES INTO TEACHING BIOSTATISTICS
A Hands-On Experience

Marissa P. JUSTAN
AMA Computer University
AMA Education System, 59 Panay Avenue
Quezon City, Philippines
e-mail: mpjustan@amaes.edu.ph

ABSTRACT

This paper describes an innovative way of teaching Biostatistics (or Biostat) at the undergraduate level. Statistics is a fundamental subject in all courses. In particular, senior students taking up pre-med courses enrol in the subject Biostat. However, there is not much difference between the methods of teaching Biostat and the fundamental statistics. The course content (or curricula) is the same for both except for the case studies. To make this difference strikingly clear to the students, they were asked to do Biostat with medical practitioners. Notably, students experienced the applications of statistics software package SPSS® and learned diagnostic tests and other statistical analysis tools which are not found in their Biostat curriculum We summarize their studies and the proposed changes to the curriculum of Biostat that their collaborations with medical doctors brought about.

1. Introduction

The ongoing challenge of learner-centered curriculum is to help students learn by active inquiry rather than by memorizing facts. This is opposed to the traditional design or subject-centered curriculum [9]. The emphasis of the later is on making the learner absorb as much knowledge as possible on the subject matter.

In the former, learning is built upon the activities students engage in. Under this design, learning activities may be based on the actual (or presupposed) needs and interests of the students. They choose what they want to learn and the teacher serves as guide, pointing where to get the necessary information. After the learner has completed his investigation of the problem that he has chosen, he makes a presentation to the teacher or takes a test on the subject matter.

The purpose of this paper is to report such experience for a group (n=8) of senior biology students, 6 girls and 2 boys. The students were grouped into 3 teams. Team I consisted of 3 girls and worked with a female Obstetrician. The other 3 girls were grouped as Team II and collaborated with a female Gynaecologist. The two boys formed team III and cooperated with a male Pediatrician. The doctors were in their second year of residency in the same hospital. They have already collected their data and would just need assistance in applying statistical tools. The doctors' studies were all due in two weeks. This circumstance provided the cap to the extent of time and work that the students would have to spend with the doctors.

In Sections 2, we tell the experiences of Team I in their journey to learning clinical diagnostic tests such as sensitivity, specificity, etc. Team II explored the flexible statistical analysis and data management system of SPSS® in Section 3. In Section 4, Team III made clear the importance of graphical representations. Finally, we give the conclusions of these learning activities in Section 5.

2. Team I: Diagnostic Tests

Often, medical doctors want to know whether the tests that they perform match the actual findings. Team I worked with an Obstetrician who wanted to know whether ultrasound (USG) test on expectant mothers can detemine anomalies (harelip, sunset eyes, hydrocephalus, etc.) in their babies. Commonly used diagnostic tests that measure the accuracy of such procedure are the sensitivity and specificity analysis. The data are shown in Table 1a.

Table 1a. Distribution of cases according to USG test against the outcome

Anomaly Outcome				
USG Test		Present	Absent	Total
	Present	16	5	21
	Absent	4	73	77
	Total	20	78	98

Among the 98 cases, 16 anomalies detected by the USG test were observed in the babies delivered. Five anomalies detected by the USG test were not found in the babies delivered. Four anomalies were found in the babies but not detected by the USG test. Seventy-three cases were detected by USG test as anomaly-free and not found in the babies.

The team was not familiar with the diagnostic tests required by the doctor. They included a glossary of the terms in their report, which is found in Appendix A. The summary of the results of diagnostic tests is found in Table 1b.

Table 1b. Diagnostic Tests

Sensitivity	80.00%
Specificity	93.59%
False Positive	6.41%
False Negative	20.00%
Positive Predictive Value	76.19%
Negative Predictive Value	94.81%
Overall Accuracy	90.82%
Prevalence	20.41%
*p value	0.00000<0.05 S

*Fisher's Exact Test, 2-Tail, 95% confidence interval

Based on the formula in Appendix B, the computations were as follows:

Sensitivity = 16 / 20 = 0.8 or 80%
Specificity = 73 / 78 = 0.9359 =93.59%
False Positive = 5 / 78 = 0.0641 = 6.41%
False Negative =4 / 20 = 0.2 or 20%
Positive Predictive Value = 16 / 21 = 0.7619 or 76.19%
Negative Predictive Value = 73 / 77 = 0.9481 or 94.81%
Overall Accuracy =(16+73) / 98 = 0.9082 or 90.82%
Prevalence = 20 / 98 = 0.2041 or 20.41%

Here, they got a very high sensitivity, specificity and overall accuracy rates. This led them to the conclusion that USG test can detect anomalies in babies before they are born.

The team also learned from another doctor, a Gynaecologist, about a study that required diagnostic tests for a 3x3 distribution table. The doctor wanted to know the accuracy of the frozen section test in determining the actual stage of cancer in 339 patients. Table 2a gives the distribution and Table 2b summarizes the results of the desired diagnostics tests.

Table 2a. Distribution of benign, borderline, malignant cases according to the Frozen Section test against the final diagnosis

		Final Diagnosis			
		Benign	Borderline	Malignant	Total
Frozen Section Test	Benign	267	4	0	271
	Bordeline	2	13	3	18
	Malignant	0	0	50	50
	Total	269	17	53	339

Table 2b. Diagnostic tests

	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Benign	99.3%	94.3%	98.5%	97.1%
Bordeline	76.5%	98.4%	72.2%	98.8%
Malignant	94.3%	100.0%	100.0%	99.0%

The formulas used were just derived from the results of a previous study (source unknown). The computations were as follows:

(i) Sensitivity:

Benign = 267 / 269 =0.993 or 99.3%
Borderline = 13 / 17 = 0.765 or 76.5%
Malignant = 50 / 53 = 0.943 or 94.3%

(ii) Specificity:

Benign = (13+0+3+50) / (17+53) = 0.943 or 94.3%
Borderline = (267+0+0+50) / (269+53) = 0.984 or 98.4%
Malignant = (267+2+4+13) / (269+17) = 1 or 100%

(iii) Positive Predictive Value:

Benign = 267 / 271 = 0.985 or 98.5%
Borderline = 13 / 18 = 0.722 or 72.2%
Malignant = 50 / 50 = 1 or100%

(iv) Negative Predictive Value:

Benign = (13+0+3+50) / (13+0+3+50+2+0) = 0.971 or 97.1%
Borderline = (267+0+0+50) / (267+0+0+50+4+0) = 0.988 or 98.8%
Malignant = (267+2+4+13) / (267+2+4+13+0+3) = 0.99 or 99.0%

Here, the students learned that the benign stage has the highest sentivity rate and the malignant stage has the highest specificity rate when using the frozen section test.

3. Team II: Estimating Risk in a Case-Control Study

Team II worked with a Gynaecologist. It is reported that premature rupture of fetal membrane (PROM) occurs in 4.5 –7-6% of pregnancies. The doctor wanted to evaluate the clinical usefulness of a new bedside test, called PROM test, for the detection of ruptured fetal membrane (ROM).

Among the 28 patients evaluated for suspected ROM, the PROM was positive in 8 cases and negative in 20 cases. Among the PROM test- positive group, 6 patients had preterm delivery while among the PROM test- negative group, 2 had preterm delivery. Table 3a summarizes the number of cases.

Table 3a. The number of patients who had preterm delivery in the PROM Test groups

PROM Test			
	+ Group	- Group	Total
Cases (Preterm delivery)	6	2	8
Control	2	18	20
Total	8	20	28

Using the statisctial software SPSS®, the Student's unpaired t-test was used for continuous variables (age, weeks of gestation) and differences in the distribution of discrete variables were computed using Fisher's exact test. The result was compared with that of the Likelihood ratio shown in able 3b. They also estimated the relative risk using the odds ratio (OR) shown in Table 3c.

Table 3b. Chi-Square results

	Value	DF	Significance
Likelihood Ratio	8.85838	1	0.00070
Fisher's Exact Test (2-Tail)			0.00176

Table 3c. Relative Risk Estimate

	Value	95% Confidence Bounds	
Case Control (odds ratio)	27.0000	3.09261	235.7233
+ group risk	7.5000	1.89761	29.64250
- group risk	0.27778	0.08291	0.93067

Women with suspected ROM and a positive test result had a 7.5 relative risk, odds ratio 27, 95% confidence interval (CI) 1.89-29.64, p value < .05, of preterm delivery. The 95% CI does not include 1 so we can conclude that the two incidence rates are significantly different.

Here, the students learned about the difference between p value and 95% CI. They contain the same kernel of information, but the 95% CI contains more information. The following was the discussion that occurred:

Suppose there were other two studies that showed the same odds ratio (OR). They showed different CIs, however, at p value < .05.

Study #2 showed an OR of 27 with a 95% CI from 1.45-36.21

Study #3 showed an OR of 27 with a 95% CI from 0.4-25.6

What can be said from these results?

Studies 1 and 2 were "statistically significant," with a p value < .05 because the CI does not include 1.

Study 3 included an OR of 1 in the 95% CI, and therefore the p value was not < .05.

Study 1 had a more precise estimate of the true OR, with a very small 95% CI.

4. Team III: Testing Hypotheses about Mean Differences

Team III worked with a Pediatrician. The subjects were 243 full term infants born from the periods of January 1999 to January 2000. The records were reviewed by the researcher-pediatrician and the following data were recorded: sex, type of milk feeding, i.e. purely breast-fed versus purely formula-fed, weight, height, head circumference and illnesses encountered from

birth, 6 months of age and at 1 year. These ages were chosen since not all the records contained complete data for ages between birth to 6 months and from 6 months to 1 year.

The mean and standard deviations of their birth weight, height and head circumference were shown in Table 4. However, due to constraints, the presentation here is limited to the results on weights.

Of the 243 infants reviewed, 140 were male and 103 were female. Furthermore, out of the total population only 37 subjects (15.2%) were purely breastfed from birth to 1 year of age while the remaining 206 subjects (84.8%) were given milk formulas.

Table 4. Population Characteristics

	Breastfed	Formula-fed	p Values
Birth weight	3.18 +/- 0.63	3.10 +/- 0.54	p=.000 (S)
Birth length	49.52 +/- 3.48	49.14 +/- 2.79	p=.000 (S)
Head circumference	34.12 +/- 1.96	34.03 +/- 1.71	p=.000 (S)
Sex ratios (M:F)	3:2	4:3	p=.000 (S)

*mean +/- standard deviation

A t-test between breast-fed and formula-fed infants with 95% confidence interval for difference was made. At value $p<0.05$, the null hypothesis that there is no significant difference between the growth parameters of babies given breastmilk and milk formula from birth to 12 months was accepted. As such, there is evidence that milk formulas are comparable to breastmilk in terms of affecting weight measurements in infants below 1 year. This study also compared the results of the subjects' growth curves with existing growth tables such as the Food and Nutrition Research Institute & Philippine Pediatric Society Anthropometric Tables and Charts for Filipino Children (FNRI-PPS) [2] and the National Center for Health Statistics Percentiles Tables and Charts (NCHS) [1]. The graphs are shown in Figures 1a & 1b.

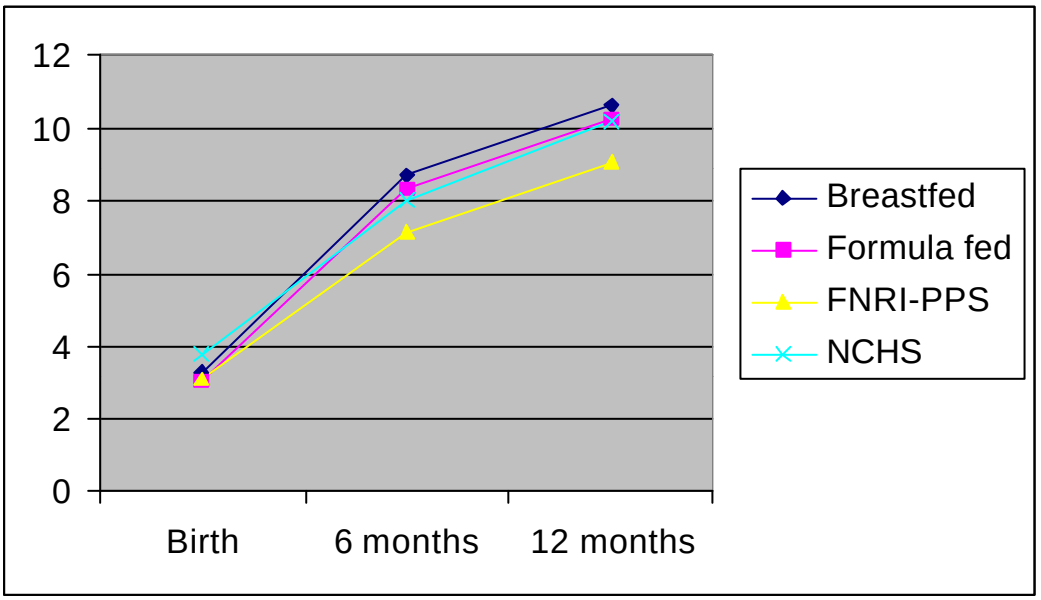

Figure 1a. Comparison of weights between types of milk feeding vs established growth curves in male infants. The values were plotted with those of FNRI-PPS and NCHS tables of boys 0-12 months.

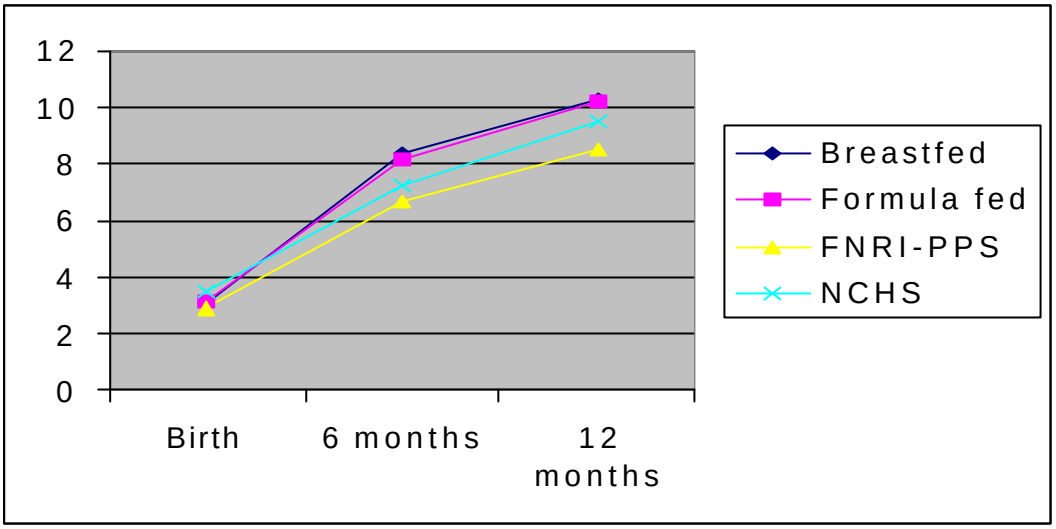

Figure 1b. Comparison of weights between types of milk feeding vs established growth curves in female infants. The values were plotted with those of FNRI-PPS and NCHS tables of girls 0-12months

Finally, a chi-test to determine the relationship between the type of feeding and occurrence of common illnesses was formulated. All chi-test showed the value $p=0.000<0.05$, thus rejecting the null hypothesis. Therefore, there is evidence to show that the occurrence of common illness is dependent on the type of milk feeding (Figure 2).

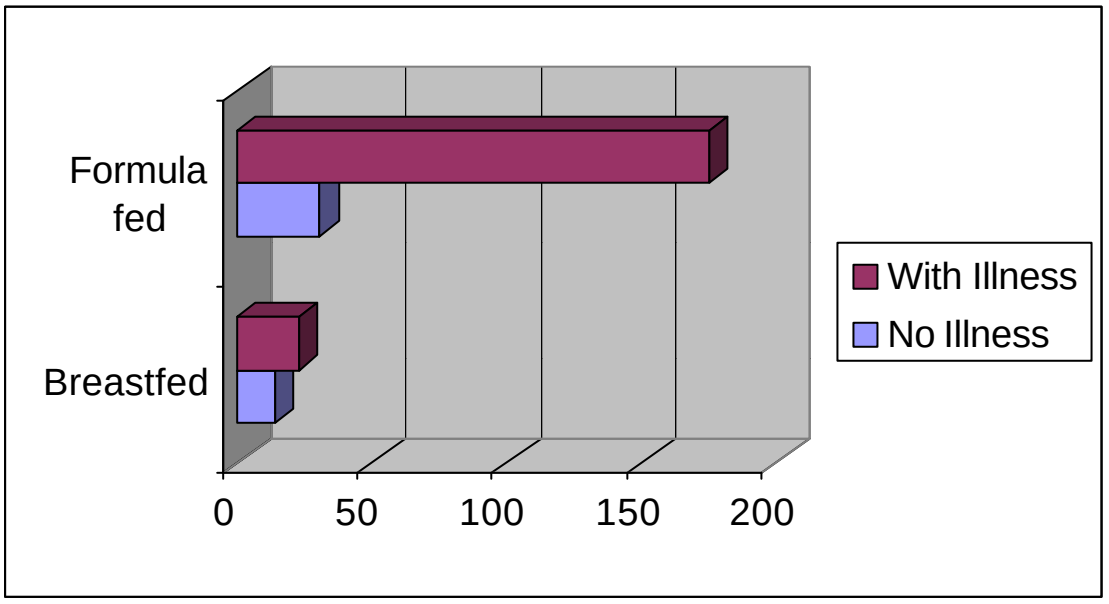

Figure 2. Comparison of type of feeding versus existence of illnesses during the 1st year of life

This study showed that present day milk formulas are comparable to that of breastmilk as to weight gain at least for the first year of life. The growth curves of breast-fed infants versus formula-fed infants did not differ significantly as opposed to previous studies that state that breast-fed infants are leaner. Formula-fed infants are, however, more prone to develop illnesses compared to their breastfed counterparts. The FNRI-PPS growth tables may need further examination in

terms of applicability to the Filipino population since even growth curves of breast-fed babies, specifically the weights were significantly different from existing weight values.

5. Conclusion

For most statistics classes, the projects have been simply to conduct surveys which concepts most closely matched the lessons. In contrast, the above learning activities had made a concerted effort to create lessons directly aligned with biostatistics concepts. The lessons were unique because learners worked on real-medical data from a respected medical center that promotes research, and were classroom-ready.

The 8 students involved in this learning activites signified their intention to continue their study in medicine. Two factors, intentionally designed into this particular course, may have contributed to this disposition.

- The On-The-Job Experience. Collaborating with medical practitioners and working with real data, helped them become socially responsible, proactive individuals. It enabled them to plan and realize social improvement at the local and global levels.
- The Application of Technology. The work that they did using MS Excell® and SPSS® showed that statistics can be learned and applied with 'less' mathematics. Grievous math computations were removed, enabling them to focus on the understanding of statistical concepts and the interpretation of the results.

The doctors themselves expressed their trust and gratitude to the students for helping them in the statistical section of their studies. Without such partnership, they expressed concern whether they could have finished their studies on time due to their hospital load as resident doctors. They would recommend to other doctors this collaboration with senior students enrolled in Biostat classes.

Finally, the proceedings of all the three studies were documented for inclusion in the next prints of learning materials in Biostatistics.

REFERENCES

[1] Berhman R., ed., 1992, National Center for Health Statistics (NCHS) percentiles. *Nelson's Textbook of Pediatrics 14th ed.* Philadelphia, W.B. Saunders Company, pp.34-36.

[2] Florentino R., Santos-Ocampo P,, et al. 1992, FNRI-PPS Anthropometric Tables and Charts for Filipino Children, Manila, Philippines.

[3] Harcourt Academic Press Dictionary of Science and Technology, "False Negative", http://www.harcourt.com/dictionary/def/3/8/2/1/3821800.html

[4] Introduction to Clinical Reasoning -Evidence Based Medicine Home Page, Medical University of South Carolina, "Diagnostic Tests", http://www.musc.edu/dc/icrebm/diagnostictests.html

[5] Introduction to Clinical Reasoning -Evidence Based Medicine Home Page, Medical University of South Carolina, "Epidemiology", http://www.musc.edu/dc/icrebm/epidemiology.html

[6] Introduction to Clinical Reasoning -Evidence Based Medicine Home Page, Medical University of South Carolina, "Evidenced Based Medicine Terms", http://www.musc.edu/dc/icrebm/ebmterms.html

[7] Introduction to Clinical Reasoning -Evidence Based Medicine Home Page, Medical University of South Carolina, "Statistical Significance", http://www.musc.edu/dc/icrebm/statisticalsignificance.html

[8] Randy Rejda , "What is a false positive?" http://service2.symantec.com/SARC/sarc.nsf/info/html/what.false.positive.html

[9] Reyes, Floredliza C., 2000, Engineering the Curriculum: A Guidebook for Educators & School Managers, DLSU Press, Inc., Philippines, pp. 94-95.

APPENDIX A

Evidence Based Medicine Glossary

Incidence Rate: Number of new cases of a disease in a specified period / average population during that period. Rate is usually expressed as per 100,000. [5]

Likelihood Ratio: The likelihood that a given test result would be expected in a patient with a disease compared to the likelihood that the same result would be expected in a patient without that disease. [4]

Negative Predictive Value (NPV): The percentage of people with a negative test who do NOT have the disease. [4]

False Negative a test result that wrongly excludes an individual from a diagnostic or other category. [3]

False Positive, also known as a false detection or false alarm, a test result that wrongly detects a disease in an uninfected individual. [8]

Positive Predictive Value (PPV): The percentage of people with a positive test result who actually have the disease. [4]

Prevalence Rate: Number of people with a disease at a given point (period)/ population at risk at a particular point (period). Rate is usually expressed as per 100,000. Prevalence = Incidence X duration [5]

Odds ratio is used in case control trials: Odds of a case patient being exposed divided by odds of a control patient being exposed. [6]

Relative Risk: Event rate in treatment group divided by the event rate in the control group. Also known as risk ratio. RR is used in randomized trials and cohort studies. [6]

Sensitivity: The probability of the test finding disease among those who have the disease or the proportion of people with disease who have a positive test result. [4]

Specificity: The probability of the test finding NO disease among those who do NOT have the disease or the proportion of people free of a disease who have a negative test. [4]

Statistical vs. Clinical Significance: Statistical significance means the likelihood that the difference found between groups could have occurred by chance alone. In most clinical trials, a result is statistically significant if the difference between groups could have occurred by chance alone in less than 1 time in 20. This is expressed as a p value < 0.05. Remember that a trivial difference can have a very low p value if the number of subjects is large enough. Clinical significance has little to do with statistics and is a matter of judgment. It answers the question: "Is the difference between groups large enough to be worth achieving?" Studies can be statistically significant yet clinically insignificant. [7]

APPENDIX B

Diagnostic Tests Formula

Table B1. 2x2 Distribution Table of Test Outcome against Actual Outcome [4]

		Disease		
		Positive	Negative	
Test	Positive	True Positive (TP)	False Positive (FP)	TP + FP
	Negative	False Negative (FN)	True Negative (TN)	FN + TN
		TP + FN	FP + TN	

SENSITIVITY = TP / TP+FN

SPECIFICITY = TN / TN+FP

POSITIVE PREDICTIVE VALUE (PPV) = TP / TP+FP

NEGATIVE PREDICTIVE VALUE (NPV) = TN / FN+TN

FALSE POSITIVE (F+) = FP / TP+FP

FALSE NEGATIVE (F-) = FN / TP+FN

NEGATIVE PREDICTIVE VALUE (NPV) = TN / FN+TN

LR(-) = [FN / (TP + FN)] / [TN / (FP + TN)]

LR(+) = [TP / (TP + FN)] / [FP / (FP + TN)]

Table B2. 2x2 Distribution Table of Outcome of Case-Control Study [6]

		Outcome	
		Event	No Event
Exposure	Case	a	B
	Control	c	D

RELATIVE RISK = a/(a+b) / c/(c+d) = a(c+d) / c(a+b)

ODDS RATIO = a/c / b/d = ad /bc