CHANGING APPROACHES AND PERCEPTIONS: BIOSTATISTICS AND ITS ROLE IN TEACHING THE STELLENBOSCH DOCTOR

Paul Mostert
Stellenbosch University, South Africa
pjmos@sun.ac.za

The role of Biostatistics in Medicine and Health Care is sometimes only fully understood and appreciated by medical practitioners once they are fully qualified. Biostatistics and Epidemiology are also subjects in the medical curriculum that are disliked by most undergraduate students. The 'Profile of the Stellenbosch doctor' is a set of professional characteristics that every successful graduate possess, once they have qualified. The curriculum is therefore designed around this profile, where the Department of Statistics and Actuarial Science at Stellenbosch University in South Africa was one of the role players, especially with the introduction of a 'golden thread.' The negativity of our medical students towards Biostatistics is real and a lot of that blame can be assigned to the way we teach and what biostatisticians think should be covered in a syllabus. Students normally struggle with probability and its applications and by introducing Bayesian approaches to the analyses will further complicate their understanding thereof.

INTRODUCTION AND BACKGROUND

The medical practitioner in the 21st century will need a far greater ability to evaluate new information and technologies than in the past. A good understanding of Biostatistics and Epidemiology can improve clinical decision-making, programme evaluation and medical research with regard to both individuals and groups of people. The teaching and inclusion of Biostatistics in the medical curriculum at Stellenbosch University is regarded by the Medical Faculty and the South African Medical Council as fundamentally important. It is common practice that Biostatistics is included in the curricula at Medical Schools all over the world.

In the last 15 years at Stellenbosch University, a few changes have been implemented regarding the teaching of Biostatistics. The first big change was made about six years ago, after feedback from students and a proper needs analysis, done by the curriculum development committee. Students fail to understand the link between Biostatistics and Medical Health Care and we as statisticians were solely to blame for that. Previously, Biostatistics was used in the module name and the module content was strictly statistical with here and there medical applications. Feedback from students showed that they were very negative towards Biostatistics, mainly because it was not really integrated into Medicine and Health Care, unlike their other first year modules of Biology, Physics and Chemistry. Students had to calculate by hand all the relevant descriptive statistics, draw by hand the relevant descriptive graphs, did an extensive course in probability theory and conduct by hand hypothesis tests and general statistical inference. Computer literacy formed part of the curriculum, but in a very limited form before 2000, due to the lack of availability of computers in an electronic classroom scenario.

Currently, Biostatistics is covered in the module: Data Management and is a compulsory module in the first year (phase I: Foundation) for the degree MB.ChB. Students have refresher courses on Biostatistics in their third year of the clinical rotation subject of Community Health, where the focus is more on Statistical Epidemiology. In the fifth year, students have their last encounter with certain topics of Biostatistics.

The role of Biostatistics in Medicine and Health Care is sometimes only fully understood and appreciated by medical practitioners and doctors once they are fully qualified, for a number of reasons. Biostatistics and Epidemiology are also one of those subjects in the medical curriculum that is possibly disliked by the majority of undergraduate students. Most students only realise at a tertiary level that there exists a science that involves statistical concepts. The medical curriculum is also revised every couple of years to ensure that it adhere to the basic principles of medical care, national interests of the medical profession and the demands of the 21st century with respect to patient care.

The *Profile of the Stellenbosch doctor* is a set of professional characteristics, which include knowledge, attitudes/views and skills, that the recently graduated Stellenbosch doctor

must possess to optimally utilise the opportunities and be able to function autonomously in the primary health care sector, as well as being equipped with the ability and insight to develop further as practitioner at secondary and tertiary level. The curriculum was therefore designed around this profile, where the concept of a *golden thread* was one of the key aspects. The Department of Statistics and Actuarial Science was also one of the role players in this whole exercise, especially with the introduction of the golden thread.

The subsequent sections discuss the process of designing the new curriculum. It focuses around the role that the Department of Statistics and Actuarial Science played, as well as the identification of the elements of the golden thread in the curriculum. The Profile of the Stellenbosch doctor is briefly explained and the specific outcomes in the revised syllabus are linked with this profile.

Concepts covered in the Biostatistics and Epidemiology modules are sometimes very difficult to understand by students, especially those involving probability and conditional probabilities with their uses in inference and their consequent interpretations in medical journals. Students normally struggle with probability and its applications and by introducing Bayesian approaches to the analyses will further complicate their understanding thereof. The last section discusses some specific difficulties and obstacles that lecturers encounter when teaching Biostatistics to medical undergraduate students.

THE PROFILE OF THE STELLENBOSCH DOCTOR

The recently graduated Stellenbosch doctor must possess the necessary knowledge, skills and attitudes to optimally utilise the opportunities available during the internship, so as to be able to function autonomously in the primary health care sector thereafter, and must also be equipped with the necessary ability and insight to develop further personally and professionally. Students are familiar with this profile from day one of their studies and every module should link up with at least some of these outcomes in the profile. The complete profile can be obtained on the website of Stellenbosch University (http://academic.sun.ac.za/healthsciences/). Only the outcomes relevant to the teaching of Biostatistics and Epidemiology are mentioned here.

To fulfill these requirements, the recent graduate will exhibit the following professional characteristics (the first two are *knowledge* based and the last three are *skills* based):

- Relevant knowledge of the necessary medically applicable scientific and mathematical concepts;
- Relevant knowledge of the principles of research;
- The ability to integrate, interpret and apply knowledge;
- The ability to communicate effectively with patients from different cultural groups in the process of diagnosis and management;
- The ability to effectively utilise relevant technological resources (e.g., computers) in the health environment.

The profile has a number of *attitudes or views*, but none of these are applicable to the teaching of Biostatistics.

THE ROLE OF THE STATISTICS DEPARTMENT AND THE GOLDEN THREAD

More or less every six to seven years the syllabus of the medical curriculum is revised to ensure that the content is still relevant. This is done by a external peer-group audit that evaluates every aspect of the curriculum. The National Department of Health, together with the Department of Education have also decided to cut subsidies to Medical Schools and will in future only subsidise five years of the curriculum instead of the normal six years. Consequently, the original first three years of the curriculum had to be redesigned to fit into a two year study period, without sacrificing important and needed study material. A new foundation phase had to be developed and the Department of Statistics and Actuarial Science, one of the previous role players, together with the Departments of Zoology/Biology and Chemistry, were identified to assist the curriculum development committee in shortening the first year. Various workshops were held over a three year period to decide which modules should be offered, in what depth certain disciplines should

be covered and how the syllabus can be improved to still satisfy the needs of the medical profession.

It was decided that some of the second semester modules (phase II of the curriculum) can successfully be offered in the first semester together with the revised Data Management, Chemistry and Biology modules. Four new modules will replace the existing four modules and they are: *Personal and professional development*; *Life forms and functions of clinical interest*; *Cellular physiology*; and *Health in context*. Biostatistics, after thorough consultation, will mainly be absorbed within the first module. Within this module, a theme called *Information Literacy*, will cover the basic and clinical Biostatistics and Statistical Epidemiology. The Information Literacy theme makes also sense with regard to the inclusion of Biostatistics, since students have to understand and interpret papers published in medical journals and conference proceedings throughout their undergraduate studies and beyond. Papers published in medical journals make extensive use of Biostatistics and Statistical Epidemiology.

Since the first year is basically a foundational year, only the basic statistical concepts should be covered and as students progress to the later years of their studies, they will get refresher courses in Biostatistics, and hence a golden thread is created. Biostatistics is often seen by medical students as one of those subjects that they have to enroll for in the first year and perceive this as a very isolated subject that is not at all connected to their medical training (Sahai and Ojeda, 1999). This opinion was evident on a number of occasions where senior students have to give feedback and evaluate their learning experience. The introduction of the golden thread will ensure that Biostatistics is integrated into their studies by applying the theory to every possible clinical subject of the senior years. The introduction of the golden thread will ensure that problem-based learning (PBL) take place. PBL has been introduced successfully by a number of Medical Schools in the UK and in Australia (Bland, 2004). Caution has to be taken that Statistics, as a golden thread, should not be seen as a burden when included in the advance clinical modules and should still be lectured by Biostatisticians. This principle has been agreed upon by the curriculum development committee and the Statistics Department will be responsible for the teaching part. Some of the ideas in the golden thread are:

- Biostatistics, Epidemiology and Evidence Based Medicine (including Clinical Epidemiology) should be taught as a continuum, with its relevance to thinking about health and disease, prevention and treatment clearly stressed.
- As far as possible, in each medical discipline, a doctor should be identified who can champion thinking in terms of Biostatistics, Epidemiology and Evidenced Based Medicine.
- The same standard, up to date, almost self explanatory textbook be used by all facilitators/lecturers and students in order to ensure consistency and avoid conflicting terminology (e.g., Dawson and Trapp, 2004).
- There should be at least, a one day seminar for all those involved in teaching/facilitating learning, including the Biostatisticians towards the end of each year to:
    - Review the curriculum content and implementation, the learning materials and teaching techniques and outcomes.
    - Brush up on key concepts and co-ordinate the learning/teaching process.
    - Develop opportunities for continuing professional development for colleagues.
    - Continue to offer the Journal Reading Course for health science graduates (including a few refresher theoretical introduction sessions on Biostatistics and Epidemiology).

Every exam paper in the advance modules should therefore also include some statistical questions to illustrate the importance of Biostatistics in Medicine. Mini research projects that students have to conduct in their final year when placed in the various hospitals and clinics will be evidence that they master the important concepts of Biostatistics and Research methodology.

LINKING THE PROFILE WITH BIOSTATISTICS

The contents of the new theme of Information Literacy (i.e., Biostatistical concepts), had to be linked with the relevant outcomes of the Profile. The aim of this theme is that students will utilise the available information resources and infra-structure, such that they can apply with

confidence the acquired skills in their studies, research and ultimately the profession. Specifically, students have to: (i) find published papers in medical journals on the internet or library to determine the type of study design used with the necessary justifications; (ii) summarise and describe a clinical data set with the aid of computer software, with an emphasis on interpretation; (iii) interpret probabilities as it occurs in various formats in the literature and to illustrate the consequences of the probability concepts; (iv) quantify and interpret the accuracy of diagnostic procedures in Evidence Based Medicine. These four specific outcomes of the theme link directly to the mentioned items in the Profile.

The specific outcomes will be achieved by using a comprehensive textbook (e.g., Dawson and Trapp, 2004), which have to include the following topics (with limited spreadsheet exposure):

- Types of study designs in research, risk factors versus outcomes, advantages and disadvantages of the different type of study designs;
- Measures of centrality and spread, scales of measurement, descriptive graphs and tables, risk calculations, growth charts, incidence and prevalence and comparison and standardisation of rates;
- Probability rules and definitions, sampling techniques, probability distributions, sampling distributions, confidence intervals and p-values in inference and Kaplan-Meier estimates;
- Agreement measures like kappa, sensitivity and specificity and prediction of negative and positive tests.

Better understanding of these concepts is achieved by analysing small and simple clinical data sets with spreadsheet software. Emphasis will be on the interpretation of results where a foundation is established for the golden thread of Biostatistics. This approach will hopefully change the existing negative perception of the students with respect to Biostatistics and will from an early stage in their studies realise the importance of Statistics in Medicine.

DIFFICULTIES AND OBSTACLES IN TEACHING BIOSTATISTICS

Student feedback forms have been used in the past as a tool to assess various aspects of the curriculum. Valuable information have been obtained in the way we lecture, what we lecture, what we expect from students and what students expect from lecturers.

These student feedback and discussions with colleagues yielded that most students and doctors seem to view Biostatistics, Epidemiology and Evidenced Based Medicine as more of an exam hurdle, a tool for researchers or a subject of academic discussion, than a discipline that can help them to understand disease and health, improve their clinical and community health practice and enable them to undertake worthwhile projects. This may be because the subject is usually presented and examined as chunks of new information often with neither much time for reflection nor repetition of the concepts in clinical teaching. Students tend to learn the concepts more from memory than from exercise and understanding.

The negativity of our medical students towards Biostatistics is real and a lot of that blame can be assigned to the way lecturers teach and what biostatisticians think should be covered in a syllabus. One should remember that they want to become doctors and not statisticians. Listening to their needs helped to change their attitudes. Statistics should be relevant and directly linked to their studies. Calculating statistics by hand and conduct hypothesis tests with critical values in statistical tables are not the things medical students should do. These were the main contributors to their negativity and students could not see the link between Biostatistics and Health Care. They were overwhelmed with formulae, methods and calculations in such a way that they actually were unable to see the bigger picture.

As mentioned in the introduction, students struggle a lot with the concept of probability and where it occurs in medical journals. The correct interpretation is not always straightforward to a student or the medical practitioner. On the other hand, statisticians find it much easier to interpret and to understand, since a lot of time is spent on research and statistical applications. The following few paragraphs illustrate some of these difficulties, with possible solutions in an effort to explain some of the reasons why students cannot master the most important concepts.

Students normally do not experience any problems with Descriptive Statistics. Any statistical textbook has a chapter on Probability Theory and this is the single concept that medical students fail to master. They cannot distinguish between rules of probability and how to test for certain properties in probabilities like statistical independency. The concept of a conditional probability is the single most difficult concept that medical students fail to understand. Numerous examples are given in textbooks and most medical students have no idea what the term *conditional* mean. The majority of students interpret the conditionality of an event as a joint event.

The concepts of sensitivity and specificity are some of the better ways to explain the conditionality principle. An example in the textbook of Dawson and Trapp (2004, p. 303-306), is where a 57-year-old man presents with a history of lower back pain. A decision has to be made to order erythrocyte sedimentation rate (ESR) or directly order a lumbar MRI based on a diagnostic procedural test for ESR>20mm/h. The reported sensitivity and specificity for an EST>20mm/h were 78% and 67% respectively. A 20% prior probability is used to indicate the chance of a spinal malignancy.

The following table is used in this example:

	Disease +	Disease -
Test +	156	264
Test -	44	536
	200	800

Following the notation of the textbook, Pr(D+|T+)=0,78 and Pr(D-|T-)=0,67. Students realise very quickly that there is a meaningful difference between this on the one hand and Pr(T+|D+), the predictive value of a positive test and Pr(T-|D-), the predictive value of a negative test on the other hand. Students have observed, without any guidance, that if the prior probability changes, these predictive values also change when applying Bayes' theorem.

Another challenging concept, related to the conditionality principle, that students have difficulty with, is the *p-value* in a hypothesis test. Over 90% of the medical journals make extensively use of p-values. Students tend to memorise the criteria for rejecting a null hypothesis based on p-values, but fail to remember how they were derived or calculated. By definition, a p-value is the probability to get a more extreme test statistic than the one obtained in an analysis, given the null hypothesis is true. Students know this definition, but fail to really understand it. They understand that a null hypothesis was set up about some parameter of interest. A test statistic, say T, was selected and then $Pr(T|H_0)$ is calculated. The null hypothesis is rejected if T is extreme, relative off course to $Pr(T|H_0)$. Calculating confidence intervals for the same parameter yielded values for the parameter for which the null hypothesis cannot be rejected. Students quickly realise that you observe only one sample and based on that sample you obtain the boundaries for the interval. Taking more samples using a simple data spreadsheet, they can observe that these boundaries differ drastically from sample to sample. A relative frequency of all these possible intervals is calculated for a parameter, to show students the real interpretation of a confidence interval. The Bayesian approach to inference is subsequently shown to students, where the arguments in the conditional probabilities are reversed, i.e., instead of calculating a p-value, a (posterior) probability is calculated as $Pr(H_0|T)$. Students observe the link between what has been covered in sensitivity and specificity versus the positive and negative predictive values of a test.

South Africa introduced a national racial quota system for all Medical Schools in 1998. According to this system, 50% of all students admitted to study Medicine at South African Universities, should come from the historically disadvantaged communities. Hence, a diverse group of students, having different academical, cultural, language and ethnic backgrounds, find themselves in the same class. According to Smith (1989): "The persistence of students in higher education is determined by a complex interplay between the background characteristics and institutional characteristics. These characteristics have been found to vary in different ethnic groups in the United States. Background characteristics of students are some of the most reliable predictors of success" (p. 30). Students from the historically disadvantaged communities are all

second language speakers, with respect to the language of instruction or teaching. The 2005 student-intake showed that this group has a very large variance when observing their grade 12 Mathematics and Science marks. In the United States it was found that High School grade point has often been found to be largest predictor of persistence in higher education, but only accounts for 12% of the variance in explaining persistence (Smith, 1989). Another study by Hassan (1997) suggests that students' verbal skills, mathematics skills and problem-solving skills are fundamental to success in Medical School. These students can also effectively analyse and synthesise large amounts of theoretical and research-related information.

It is clear that, apart from the obvious difficulties in teaching Biostatistics to medical students, the diverse composition of a class further hampers effective teaching. One of the options that the University is currently considering is to let unprepared students follow an extended curriculum, whereby the first year is spread over a two year period. In this period, these students would follow, amongst others, extensive courses in numeracy skills development and problem-solving.

CONCLUSION

A lot have changed over the years in the way Biostatistics has been taught at Medical Schools all over the world and will probably still change in future. The role that Statisticians play in the development of the academic curriculum is very important, especially if PBL is introduced as a teaching method in Medical Schools. It will remain however a challenge to successfully include it in this type of teaching, if not, it can be detrimental to Biostatistics and Research Methodology, particularly in clinical Epidemiology and Evidence Based Medicine.

REFERENCES

Bland, J. M. (2004). Teaching statistics to medical students using problem-based learning: The Australian experience. *BMC Medical Education*, 4(31), www.biomedcentral.com/1472-6920/4/31/.

Dawson, B. and Trapp, R. G. (2004). *Basic and Clinical Biostatistics* (4th edition). New York: McGraw-Hill.

Hassan, A. S. (1997). *A Complete Preparation for the MCAT*. Maryland: Williams and Wilkens.

Moussa, M. A. A. (2002). Developments in the instruction of Biostatistics at the Kuwait University Health Science Centre in a decade. *Teaching and Learning in Medicine*, 14(3), 194-198.

Sahai, H. and Ojeda, M. M. (1999). Teaching biostatistics to medical students and professionals: Problems and solutions. *International Journal of Mathematical Education in Science and Technology*, 30(2), 187-196.

Smith, D. G. (1989). *The Challenge of Diversity.* ASHE-ERIC Higher Education Report 5.