HOW TO PREPARE STUDENTS FOR WORK IN THE PHARMACEUTICAL INDUSTRY?

Ziad Taib
Early Clinical Development, Astrazeneca RD, Gothenburg, Sweden
ziad.taib@astrazeneca.com

In this article, we consider the skills and competencies needed for a statistician to make a successful career in the pharmaceutical industry. We argue that there is gap between a theory-centered academia and a problem solving and data driven industry. As a result, companies face a constant shortage of statisticians with the right skills and capabilities. However, this problem is not specific to the pharmaceutical industry and eventually necessary reforms will be made to remedy this situation. In the meanwhile, there is a need to design courses that build on existing curricula. We discuss some recent experience of such courses.

INTRODUCTION

According to a web page of the Royal Statistical Society (2017), "*Statisticians in the pharmaceutical industry are key players in all areas of drug research and development from the initial identification of a chemical right through to the manufacturing and commercialization of pharmaceutical products*." The Pharmaceutical Industry (PI) hungers for statisticians at various levels of knowledge and experience but there seems to be a chronic global lack of this competence. Large numbers of statistical graduates leave the universities every year but only few of these seem to qualify for work in the PI.

We believe that the main reason for this lack is that there is a gap between the theoretical knowledge that the Academia equips the students with and the focus on problem solving and data by the Pharmaceutical Industry. Because of that, the students are ill equipped to deal with live data from real life projects. This gap is perfectly reflected by the following statement of the International Conference on Harmonization (1998) that *"...the statistician should have a combination of education/training and experience sufficient to implement the principles articulated in this guidance*". In reality, this means that the gap can be painfully bridged by years of hands on experience of working as a clinical statistician after obtaining a degree in statistics, which is problematic for fresh graduates, to say the least. However, we argue that this gap can be narrowed and sometimes even bridged with modest efforts.

What makes this problem even more serious is that it is not restricted to a single country or group of countries but is a truly global phenomenon. To understand why, one should note that the pharmaceutical industry has, over the years, attained a high degree of globalization, harmonization and standardization in practically all parts of the world. One key driver of this harmonization is the wish that proof of efficacy of a drug, based on clinical trials results performed in one region, can lead to approval of the product also in other regions. The focus on the US and European markets has over the last decade been replaced by a global focus partly due to emerging new markets in Asia and Africa. Another driver is the competition for patients, which led to clinical trials being conducted in practically all countries of the world.

Because of this, new global data standards such as Clinical Data Interchange Standards Consortium (CDISC), Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) (2) have been created and endorsed by health authorities in many countries. Therefore, the job market for statisticians in the pharmaceutical industry has become truly global, with not only a common theoretical knowledge basis, but also common data standards. Unfortunately, in contrast with a harmonized pharmaceutical industry, different countries have different educational systems and various traditions of teaching statistics at various faculties. Different companies in the PI have their own ways of dealing with this issue. We illustrate by citing a recent ad for graduate opportunities (Astrazeneca, 2017) where graduates are invited to apply for graduate opportunities to qualify for a pharmaceutical industry career: "*We have (...) post-graduate opportunities across our late stage development statistics activities! You will spend the first two years (...) working in a structured graduate program, which will provide at least three different experiential learning opportunities - working on real projects across the function. (...) Capitalizing on your successful*

postgraduate statistics degree now is an exciting time to take the next step (...) to work on a wide range of global projects, and to apply statistical methods to a variety of problems".

The above-mentioned common ground created by the PI makes it relevant to think about the skills and competencies needed to enable the students to make successful careers in this domain. We will discussed this in the next section, while our own experience of dealing with this issue is in section 3. Section 4 contains some future perspectives.

WHAT SKILLS AND COMPETENCIES ARE NEEDED?

Much has been written about teaching statistics and about the skills and competencies that a statistician needs to have in general. (Gelman & Nolan, 2002; Tishkovskaya & Lancaster, 2012). However, there is very little written specifically about the PI. Citing one of the main recruiters of statisticians in PI: *going from a non-pharmaceutical statistics job straight into a bio-statistician role might prove tricky.*

In this section, we describe what we believe to be the key skills and competencies that a statistics graduate needs to develop in order to make a successful career in the PI. These are based on our own experience but most of them have also been suggested in various forms by different statistical professional organizations such as the American Statistical Association (ASA) and PSI. Based on own experience and review of existing literature (Justina, 2016; AFSPI Working Group, 1999) on this subject, one can identify the following key categories of skills and competencies:

1.   *Mathematics:* Statistical methodology relies heavily on mathematical concepts. Therefore, it is of uttermost importance to have a certain level of understanding of such concepts in order to apply them correctly and to be able to learn more as needed. The most important element in this area is modern statistical theory. These skills are the easiest to define and acquire.

2.   *Statistical modelling:* Applied Statistics is a craft on top of being a science. One needs to be able to use fit for purpose empirical models and understand the consequences of using these. More importantly, it is important to be able to understand what the purpose of a work assignment is and translate that into certain elements that can be formulated in terms of the models used. The statistician needs thus to have good problem solving skills on top of good grasp of statistical thinking.

3.   *Computational skills:* Knowing which analysis one needs to perform to answer a certain question using data is one thing, and actually performing the analysis is another. For the latter, one needs to be able to clean the data and transform it so it becomes fit for purpose. Then one needs to either write new code in some suitable environment or use some existing statistical software. Additional skills should include computer intensive statistical techniques.

4.   *Special area of application knowledge:* To earn the trust and respect of other scientists, one needs to be able to communicate with them. Therefore, it is important to develop a minimal level of basic understanding of the area of application, e.g. disease area knowledge.

5.   *Other skills:* Under this heading, we can collect many other skills that are almost as important as the technical ones:

1)   *Writing:* A statistician in the pharmaceutical industry is supposed to write e.g. statistical analysis plans, statistical study reports, and to review many other documents. The ability to write clearly is, therefore, essential.

2)   *Communication:* One needs to make presentations using appropriate media to various types of audiences.

3)   *Teamwork:* Clinical research is a collective effort where the role of collaboration with scientists from other disciplines cannot be exaggerated

4)   *Organization:* Within the clinical projects, the statisticians are accountable for the whole process of data handling, starting by data management and ending by an analysis report passing by many operational and programming steps performed by programmers and data managers.

5)   *Consulting skills:* Non-statisticians can only try to explain what they need from their own perspective. It is up to the statistician to understand and uncover what the issue is and how best to deal with it based on discussions and questions asked.

6)   *Ethics:* It is not a coincidence that there exist many ethical guidelines for statistical practice (American Statistical Association, 1999). Statistical integrity is what makes other scientists respect statisticians and what drives regulatory authorities to require involvement of statisticians in clinical trials.

7)   *Influencing:* statistical work is interdisciplinary by definition where statisticians almost never are the owners of the data or even of the scientific problem under investigation. Therefore, the need to influence other scientists regarding how to deal with certain issues is crucial both within an individual trial and at a strategic level.

8)   *Leadership:* This area has been recognized by groups such as the ASA as a privileged area. Statisticians need to demonstrate leadership not only within their own area but also strategically within their larger organization.

WHAT CAN ACADEMIA DO?

As mentioned above, the issue discussed in this article is not restricted to the PI and is, in fact, related to the issues that led many universities to the creation of new programs often with names like "Data Science". We think that, ultimately, there will be reforms that will solve some of the problems. In the meanwhile, it is more meaningful to think about what can be done to make a difference on top of existing programs. In this section, we account for the experience gained from years of working for a large PI and in the same time being an adjunct professor in bio-statistics at the university. More specifically, we will describe our experience from two university courses aiming at preparing the students for work in the PI.

As discussed above, the technical mathematical skills cited in the first item in section 2 are relatively easy to acquire. This is because there are many theoretical statistical courses among which the students can choose. The gap in the curriculum often consists of topics that correspond to the remaining items in section 2. To prepare the students for a career in bio-statistics without designing new pharmaceutical statistics programs from scratch, it is possible to add some key courses on top of the usual curriculum. Many types of courses can be suitable but two in particular are crucial. The first one is a course on mixed effects models and the second one is a course on statistical design and analysis of clinical trials. Below is a brief description of the author's own experience from developing and teaching such courses. Other courses such as survival analyses are usually readily included in many statistics programs albeit without the necessary focus on modelling and data.

A. The course Linear mixed models (LMM) for longitudinal data was intended to fill a major gap in the curriculum. LMM are quite common in statistical practice, while very few academic departments offer regular courses on this type of models. In the same time LMM are closely related to PROC MIXED in the software SAS, which is the tool of choice in the clinical trial world, partly because it is supposed to be validated and acceptable to the authorities. Therefore, it made sense to develop a course that teaches LMM while introducing PROC MIXED and SAS to the students. Such a course is still enough mathematically challenging to appeal to mathematically minded students. The topic also presents an opportunity to expose the students to realistic problems based on complex clinical trial data such as repeated measures and hierarchical data. An-other advantage of such a course is that LMM provide an excellent method for handling missing data, which is often a problem in clinical trials. It also exposes the students to many practical issues such as such as model checking, imputation of data, sample size planning, report writing etc. requiring skills of the type described under item 2 of section 2. The main bulk of such a course can be based on one of many excellent books on this topic, such as the book by Verbecke and Molengerghs (2009), in addition to some lecture notes to cover some additional topics such as on non-linear models.

B. The second course "design and analysis of clinical trials" is supposed to give an overview of various types of statistical models with the corresponding types of data that can occur in the PI with emphasis on the scientific (clinical) research questions addressed. On top of such specific topics, the course also gives the students knowledge and hands on experience related to various processes and documents (e.g. protocols and statistical analysis plans) and data analysis using validated

software such as SAS. Moreover statistical areas of special interest such that multiple inference (gate keeping, alpha spending etc.) and adaptive designs are also included.

Author Experience

The author's own experience of this course is related to a collaboration between AstraZeneca and the Department of Mathematical Sciences at Chalmers University of Technology in Gothenburg, Sweden. Having one foot on each side, and in view of the constant need for biostatisticians in the PI, it was natural to organize such a course. On top of a number of regular statistics students, the course could also be followed by new recruits at the company as a professional development activity. In addition, the course, not having a heavy focus on mathematics, turned out to satisfy an even greater need for those studying various biomedical topics such as bioinformatics, systems biology, pharmacology, medicine etc.

The course has so far been delivered five times, always with a mix of students from all of the above mentioned groups. Usually, the instruction is given by a number of professionals from the PI including the current author who is the main organizer of the course. Currently this course uses the book "Design and Analysis of Clinical Trials: Concepts and Methodologies" by Chow & Liu (2014) supplemented with other material handed out at the lectures. The main ingredients in the course are different types of clinical trials as part of a development program for new drugs (Phase I-IV), the principles of controlled, randomized, blinded trials different types of designs for comparative clinical trials, and the broad principles of statistical inference and statistical methods applicable to clinical trials. Besides the lectures, the course has an extensive computer project component, whereas the students have to actually perform all the steps of a clinical trial from writing a protocol and a statistical analysis plan and all the way to analysis and reporting of the data in form of a report.

GOING FORWARD

Everything points at a bright future of statistics in all areas of applications and especially within the pharmaceutical indistry. This has many different coinciding reasons of which we mention a few. The accumulation of historical data from decades of clinical research made available through various data transparency initiatives. New detailed knowledge about the patho-physiology of diseases and increasing computer power. The desire to use statistical methods not only on the level of an individual clinical trial but also on the level of a whole clinical programs, including decision making between phases based on quantitative criteria and all available data including information related to disease modelling and biomarker data. In the same time, it will be even more crucial for statisticians to have a data driven focus, whereas the need for reforming the teaching of statistics.

Moreover, the author believes that the issues discussed above are not restricted to the PI. Similar observations have been made in many other areas where there is a crying need for this type of data driven statistical skills. A recent study by the McKinsey Global Institute (Manyika et al., 2011) concludes that there will be four to five million jobs in the U.S. requiring data analysis skills by 2018, and that large numbers of positions will only be filled through training or retraining. The McKinsey authors also project a need for 1.5 million more managers and analysts with deep analytical and technical skills "who can ask the right questions and consume the results of analysis of big data effectively".

As a reaction to this, attempts are being made to create special programs in "Data Science" whereas Harvard Business Review elected data scientist to be "The Sexiest Job of the 21st Century", a term which soon became a buzzword (Davenport & Patil, 2011). However, while many university programs now offer a data science degree, there exists no consensus on a definition or curriculum contents. In the question-and-answer section of his keynote address at the Joint Statistical Meetings of American Statistical Association, applied statistician Nate Silver said, I think data-scientist is a sexed up term for a statistician" (Statistics Views, 2013). The current author also believes the solution is not to create an entirely new breed of super data scientists but to

improve the education of the usual "boring" statisticians so they are equipped with knowledge and tools to deal with real world data.

REFERENCES

AFSPI Working Group. (1999). Qualified statisticians in the European pharmaceutical industry. *Drug Information Journal, 33*, 407–415.

American Statistical Association. (1999). ASA Issues Ethical Guidelines. *Amstat News*, 269, 9–15.

Astrazeneca. (2017). Graduate opportunities. Retrieved May 26, 2017, from https://job-search.astrazeneca.com/job/cambridge/graduate-statistician-opportunities-biometrics-and-information-sciences/7684/4079124

Chow, S. C., & Liu, J. P. (2014). *Design and Analysis of Clinical Trials*. Hoboken, NJ: Wiley.

Davenport T. H., & Patil, D. J. (2011). Data Scientist: the sexiest job of the 21st century. *Harvard Business Review,* October 2011, 70–76.

Gelman, A. and Nolan, D. (2002). *Teaching Statistics: A Bag of Tricks*. Oxford, England: Oxford University Press.

International Conference on Harmonization. (1998). Guideline E9. Retrieved May 26, 2017, from *http://www.ich.org*.

Justina M. F. (2016). Careers in Biostatistics and Clinical SAS® Programming: An Overview for the Uninitiated. *SAS Global Form 2016*. Paper 7940-2016.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Buyer, A. (2011). Big Data: The next frontier for innovation, competition, and productivity. *Mackenzie Global Institute*. May 2011. Retrieved May 25, 2017, from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Royal Statistical Society (2017). Pharmaceutical Statistician. Retrieved May 26, 2017, from Statslife website: https://www.statslife.org.uk/careers/types-of-job/1128-pharmaceutical-statistician

Statistics Views (2013). Nate Silver: What I need from statisticians. Retrieved May 26, 2017, http://www.statisticsviews.com/details/feature/5133141/Nate-Silver-What-I-need-from-statisticians.html

Tishkovskaya, S., and Lancaster, G.A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations, and strategies for reform. *Journal of Statistics Education, 20*(2).

Verbecke, G., &Molenberghs, G. (2009). *Linear Mixed models for Longitudinal Data*. Berlin, Germany: Springer Verlag.