Paper PD04

Medical Research:
What should Programmers prepare for?

Emilie Brennion, F.Hoffmann - La Roche Ltd., Basel, Switzerland

ABSTRACT

Since a couple of years medical research has taken a new orientation, switching progressively from a medication "for the largest number" to a more targeted "personalized medicine". What do we programmers do in this story? Why should we bother? How can we adapt our work while we're developing more and more standard procedures and applications? Scientific improvement and technical changes on the medical research side mirror changes in our way of working. We will face large varieties of data and many new types of differing request. In order to continue to provide our customers with clear and concise information derived from raw data contain we'll have to balance innovative and validated techniques.

INTRODUCTION

For three years now I have been working for F.Hoffmann-La Roche Ltd in a team which was created to support a new kind of studies where statistical programming support was required. These studies did not come from a single department; instead they were instigated by groups from all around the company including Pharma Development, Research and Diagnostics departments. Most of the studies were in early phase but some were already part of late stage drug development projects (phase 3). The number of these very heterogenic studies has been continually increasing. I was involved in these studies because of my knowledge of biosciences, bioinformatics and large range of programming languages.

This development and increase in numbers of this category of study was due to a new strategy which pharmaceutical companies have been adopting over the past few years. We have been moving from a kind of healthcare to treat the highest number of patients to the idea of only treating the patients that would benefit the most from the therapy. This idea might not be understandable to some people. Why would a company want to exclude possible "customers" from being treated? But if you think about it, there are advantages to the pharmaceutical companies targeting the best responder. A better knowledge of the disease population would save expensive research time required to develop new drugs, which would allow the pharmaceutical companies to be able to deliver a new drug for a specific disease with more efficiency. In parallel another big objective of these studies was to better understand the patient's response to specific drugs and how to measure the best response. Reducing costs and time for drug development means more projects, more development, and more the possibility to increase the number of different diseases investigated which can be studied and eventually more patients treated with the best therapy.

1. BACKGROUND

1.1 PERSONALISED HEALTHCARE

Despite the huge advance made in medical care over the past years there are still many improvements which are required for disease diagnosis and in the effectiveness of the healthcare. For a long time the principal target of the pharmaceutical companies was to find drugs to treat the highest number of patients in as much of the population as possible. It's obvious if you consider the huge amount of studies that we handle in our companies where the principal aim is to be able to propose a marketed drug for a new population or a new application.

This way of thinking has been changing in the last few years. Healthcare companies are becoming more and more focused on an individual type of care. For a long time now we're used to treat the patient after he becomes symptomatic, at this point the disease is already at quite an advanced stage. Treating the patient even with the best therapy available is not without risk. Too many prescriptions can result in adverse events threatening the patient quality of life or life itself. At the same time it's difficult to know how the patient will react to a treatment before starting it. Thus for one given disease and one given treatment there are many different patients who react quite differently. There is no controversy in the idea that patient treatment would benefit a lot from better diagnosis and more targeted therapies.

This idea is what personalised healthcare is based on: Binding the two pillars of the drug discovery process; Pharmaceutical R&D and Diagnostics to target a specific part of the population which will benefit the most from the therapy with the least number of adverse reactions. It's the end of the 'one-fits-all' model. More and more pharma companies recognize the benefits of this new type of healthcare and it stands to be a key element in the drug development strategy.

1.2 WHAT DOES IT MEAN IN PRACTICE?

This change in the strategy of our companies means a great deal of change in our daily work. Why? Because scientific improvement and technical changes on the medical research side directly impact on our way of working. Most of us, programmers and developers, are placed in the pharmaceutical development at the end of a complex chain of collaborators. If science changes, we must adapt.

This is today a big challenge as many of our daily tasks are standardized. We have macros developed and validated or even complete tools. These applications and macros developed in-house are, most of the time, used for standardized reporting of clinical trial safety data. They aim to save our time and also save costs for the companies when we have to generate validated outputs for clinical analyses. This is useful as long as the data received can fit into a standard generic model and the outputs are also standard.

With new type of studies, which include a personalised healthcare component arriving, standard doesn't mean anything anymore. We have to adapt to each study. I will now present the new types of studies I have had to deal with.

2. NEW TYPE OF STUDIES

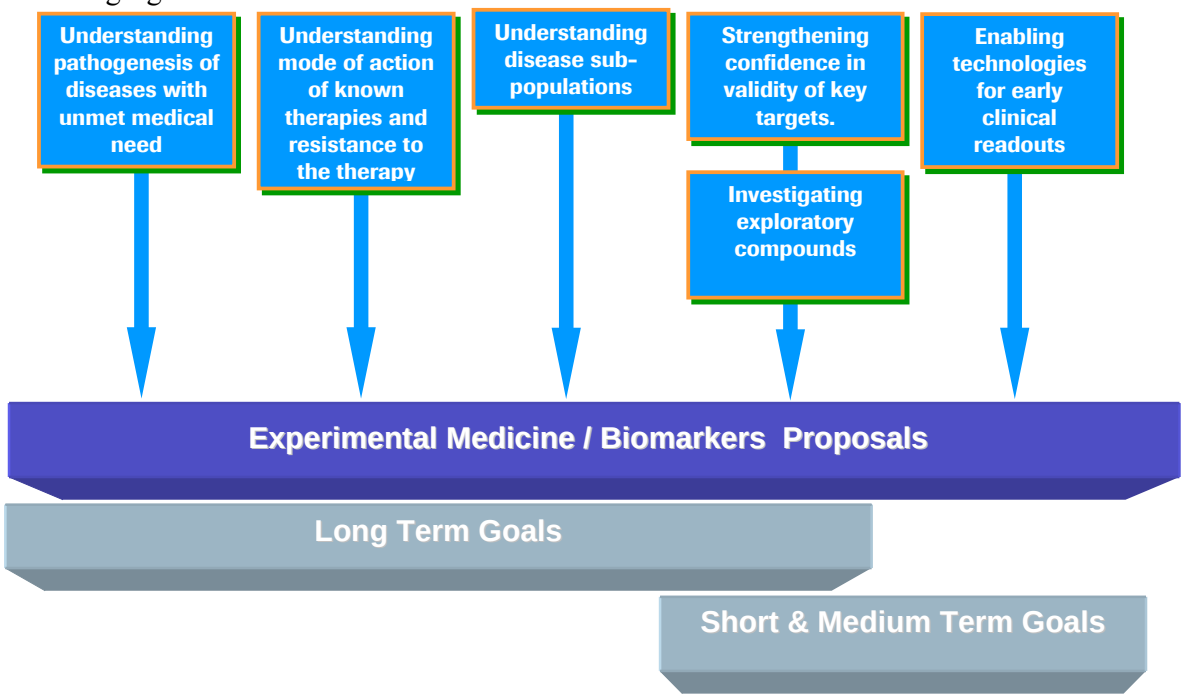
2.1 WHAT KIND OF STUDIES?

In my department I used to work on a specific type of studies called experimental medicine and biomarkers studies. Their principal characteristic is their heterogeneity. All of them are very different in terms of purpose, protocol, scientists involved etc… Many partners are involved with different experiences in clinical (non-clinical) research. These partners possibly include clinical scientists from very early to late phases, pharmacologists, Pharma Research, bioinformatics, Diagnostics and biomarkers experts. In addition, a good network of statisticians (clinical and non-clinical), bioinformaticians, geneticists, pharmacology statisticians and programmers are needed for the analysis of these studies.

From a data point of view, studies can be either managed in-house or outsourced (fully or partly) to a CRO, they can come from an academic lab or a clinical trial center. The way we have to analyze them can match the analysis of very standard clinical data or be totally different. The amount of work can be fixed and easy to plan or totally dependent on the outcome of interim analyses. In other words there is no standard. So why should we put all this studies under the cover of one single type? The aim of all these studies is to help the clinical scientists head towards a personalized treatment strategy.

2.2 EXPERIMENTAL MEDICINE STUDIES: MAIN OBJECTIVES

The main objectives of the studies I had to deal with in my team could be summarized in the following figure:



2.3 EXPERIMENTAL MEDICINE STUDIES: A TENTATIVE OF DEFINITION

Having a simple definition is not easy. I can only talk about this from a Roche point of view. A simplified definition would be to group all the studies whose main objective is to create a competitive advantage in clinical target assessment and strengthening the confidence in key efficacy and safety targets and/or to enable the availability of applicable methodologies for drug evaluation in Early Development.

Behind these main objectives can be placed:

- All clinical and non-clinical studies in Early Clinical Development that are performed under Roche responsibility without any Roche compound involved
- Interventional clinical studies with marketed non-Roche drugs (see example in section 3)
- Interventional clinical studies in healthy volunteers not involving any drug dosing
- Non-interventional studies (for example, retrospective assessments of tumor samples etc.)
- Biomarker studies

As I said before, this definition groups many heterogenic types of study. The challenges are big as there are no best or standard data handling or analysis practices available yet. Tasks have to be planned and approached very flexibly and adapted depending on the study design, study aim and study partners.

I will now describe an example study to introduce you the work that I had to undertake as a statistical programmer for one of these studies.

3. THE PROGRAMMER'S PART OF THE JOB

AN EXAMPLE: CROSSOVER TRIAL DESIGN, NON-ROCHE DRUG, OUTSOURCED

3.1 SUMMARY

This study is an HDL functionality study. The primary objective was to assess the effect of two lipid modifying drugs, Fenofibrate (a PPAR-α agonist) and Niaspan (slow release Niacin), on HDL-C plasma levels and HDL functionality in patients with low (<40 mg/dL) and normal (between 40 and 59 mg/dL) HDL-C levels at baseline.

The protocol was similar to that of a 'classical' randomized crossover clinical study with marketed non-Roche drugs, as such it is very open to further analysis. The science specialist was based in Basel. The study was conducted in one central lab with additional small laboratories in France and Italy involved for specific lipid analyses. Study conduct, management and safety analysis of the main study were fully outsourced to the CRO ICTA in France.

3.2 INTERIM ANALYSIS AND DATA DELIVERY

The work on this project should have been simple if you consider that the CRO was responsible for the main part of the safety analysis. But in these kinds of experimental medicine studies the SMT (study management team) usually decides to undertake an interim analysis on the initial data. The aim of this interim analysis done after the first period was to obtain a confirmation that :

- the monitoring of the study was undertaken correctly
- the patient selection was undertaken in accordance with the inclusion/exclusion criteria
- the adverse events were correctly registered and monitored
- the study conduct was done according to the protocol
- all the data on the case report form were entered
- the first results were consistent, with what we already know about the marketed drugs involved

Here is it when it starts to get complicated: you don't have a single coordinator. To ensure that there is consistency in any data received it is very important to define data delivery specifications. This is a very important part of the programmer's job when preparing a new study. However, for some experimental medicine studies, statistical programmers are not usually involved with the study during the planning stages, so at the beginning it is only the science department who has contact with the CRO and the labs.

In the example which I will describe the statistics department became involved in the study after the protocol was written and the patient selection started. This meant that we were not able influence the way the data was collected, stored or delivered. We decided that the way forward was to work out a good compromise and define priorities. To be able to use the data in a safe way, you should make sure that:

- data are clean, if the data management is fully outsourced, this should be the responsibility of the CRO but it's always better to do a short check on your side for missing and improbable data
- all the data from all the providers are sent to the CRO, this ensures that you only get data from a single provider
- the data manager is identified and you can communicate directly with him/her
- data are delivered in a correct format consistent from one update to another. To ensure this happens it is recommended that a comparison program is written, which allows checking that the formats are exactly the same, no variables are missing or added, the number of lines are as expected etc..
- data are delivered in a 'safe' file format, for example, a SAS file is preferable to an excel spreadsheet etc.
- data are delivered in a secure way, for example via a secure ftp site; avoid emails as much as possible

Communication is very important and it is preferable that the programmer is an active member of the study management team.

3.3 NON-"STANDARD" OUTPUTS

The next step is to produce the outputs. In my study, the outputs are not standard because the data are not standard which means that everything has to be fully programmed. We use SAS in our department as a recommended language but S-plus outputs are often produced.
For this part of the work there are no specific rules, just the well known "best practice": always comment your code and make sure all programs are well documented!

The outputs delivery always depends of your customer, again communication is really important but most of the time a Microsoft Word format is preferred with tables and graphs embedded. The SAS output delivery system (ODS) is really a great help in this case.

I won't describe in detail the procedure that I used as this is not the purpose of this paper but having in mind what you can do, what you still have to learn and what you're sure you won't be able to achieve in the timelines is really important. You have to base your communication with the customer on what is possible. 99% of the time they have no idea how an output is produced. Again being an active member of the SMT is the key to this customer expectation management.

3.4 AD-HOC REQUESTS & EXPLORATORY ANALYSIS

As I have described, the experimental medicine study protocol usually describes a large range of possible further analyses. Unfortunately, it is not possible to plan this part of your work before it has started. In each company/department some rules should be well defined, but it is important to work with your manager to request more of your time to be assigned to the study. Priorities have to be defined and a close collaboration with the statistician is very important for this. From the experience that I have gained over a number of years, the exploratory part of an analysis is always the one where the workload can vary the most. A recommended way to deal with this part is always to define what the customer requires in an output, this should include title, a description and what the output should show. At all times a justification of why the output is needed will allow your workload to be managed and the priorities will be easier to define.

3.5 USEFUL DOCUMENT AND CODE

A non exhaustive list of useful documents:

- data delivery specification
- data checking macro
- dataset comparing macro
- outputs delivery specification
- documentation of your code

And keep in mind that this is only the minimum. It can sound very usual to many of you but it's always good to be reminded of it.

4. AND SO…

4.1 WHY SHOULD YOU BOTHER?

You're not working in experimental medicine, you might not even work in early phase development but the fact is that you're going to have to deal sooner or later with a study that has an exploratory part to it. Whatever you will call it: experimental medicine, exploratory, biomarker, experimental statistic or modeling study, it's going to come soon on your desk.

When I started to work on these studies three years ago most of them were early phase studies. Today we have studies linked to phase 3. In each project there are some studies whose main objectives are to understand disease sub population or understand the response to a particular drug or just investigate which technique is the best to evaluate the response.

A big category of studies related to personalised healthcare are the studies whose aim is to analyze the biomarkers. Biomarkers are used in PHC to:

- Identify and select drug targets
- Identify patients who are likely to benefit the most or will not benefit from treatment
- Predict and monitor how patients respond to treatment
- Better predict and monitor safety

For each drug in development, there is a biomarker program which aims to discover and validate biomarkers that identify those patients who are most likely to respond to the new medicine.

You will then certainly understand that you should be prepared to face new situations. We, programmers, have to be flexible. We will have to deal with unplanned requests, non-standard analysis, lots of new graphical outputs. Being prepared means being informed and pro-active by undertaking training courses, going to conferences, etc... This is especially true because at the moment sciences specialists in collaboration with data managers, statisticians and programmers must increase their knowledge in this area. They and we have to better understand how to analyze the huge amount of data collected.

4.1 WHAT ABOUT STANDARDIZATION?

Standardization when done in a proper way means a huge saving of time and costs. It is obviously one of the most important challenges which need to be overcome to handle experimental medicine study and biomarker data in the future.

The diversity of biomarker data which we could be asked to handle is very challenging not least because of the different methods and technologies which are used. Data generated can range from single protein or gene assays, for example, from immunohistochemistry staining or taqman experiments, to whole genome gene and protein expression or genotyping data.

A major issue which we all need to consider when working with these data, is a fundamental requirement to store these in a systematic, structured and controlled manner. In order to do this efficiently we need to have a very close working relationship with data management experts in order to ensure that database structures and data formats are always consistent from project to project. In this way the data will be 'future proofed' and will allow efficient

data mining which will be required when, for example, undertaking meta analyses of data from different projects.

This is only the very first step but it is the prerequisite to all the next ones. We are now with these types of study at the beginning of a new cycle of programming and development. The challenge is great and everyone needs to collaborate to reach this new step of the evolution of drug development

CONCLUSION

The fact is that improvements in medical research, in the same ways in any kind of technological advances, are done step by step. You will always have a revolution followed by a time of adaptation. With the new personalized healthcare strategy adopted by many of our pharmaceutical companies, we are now heading to a new revolution.

From a scientific point of view personalised healthcare is a new way of thinking pharmaceutical companies and medical research in general should adopt. Our work is driven by scientific improvement and technical changes. We're a link in the chain of drug development. When standardization tends to reach its limits innovation and flexibility tend to occupy the front of the stage. Our work is now to keep our place in the chain and be as efficient as possible.

It's a big challenge for each of us. Be aware of this new turn, document yourself and wait. You won't have to wait too long; we will all be involved soon or later in this huge step of medical advances.

ACKNOWLEDGMENTS

I would like to thank my company F.Hoffmann – La Roche Ltd. for giving me the opportunity to write this paper and present it at the PhUSE conference.

My colleagues and manager Ben were really supportive and their advices very helpful every day.

A special thank to Neil Jones, a nice colleague and an expert in biomarker, for his advices, tips and correcting, he was really helpful in many circumstances.

Many thanks to Franz for his support when I joined the team of experimental medicine and for his help in defining EM studies and to Jan for answering all my questions about SAS programming and his great support.

And eventually thank you, my reader, for the time you spent and the attention you gave to my paper.

RECOMMENDED READING

For my Roche colleagues: you will find a lot of corporate information about PHC on the group intranet. For all the other: I'm sure you'll find the same on your intranet or just type "personalized health care" in Google, you will find many good articles about it.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Emilie Brennion
Statistical Programmer
Experimental Medicine

F. Hoffmann-La Roche Ltd.
PDDB
Building 670/4
CH-4070 Basel
Phone +41 61 68 70982
Fax +41 61 68 79156
E-Mail: emilie.brennion@roche.com