UNIVERSITY OF PAVIA

Research Doctorate in Psychology, Neuroscience and Medical Statistics

Department of Brain and Behavioral Sciences

Clinical trial: a crossover design application

University supervisor: Prof. Luisa Bernardinelli

Company supervisor: Dr. Laura Menta

PhD Thesis by

Alice Baccino

XXX Cycle (2014-2017)

Contents

Introduction

Overview

A new drug or medical device for being marketed, need to follow a number of complexes and long stages but mandatory for the approval of the investigational product. The practical method of performing these stages is called Clinical Trial. Clinical trials are the way of proving or deny the effectiveness and the safety of a treatment or a medical device on humans. In clinical trials subjects are randomized to receive different treatments or different sequences of treatment throughout the duration of the study. The randomization is the instrument used to create comparable groups of subjects. At the end of the study the hypothesis of the trial is tested and the efficacy/safety of the new drug is claimed or rejected.
The developing process of a new drug is long and expensive. It starts from pre-clinical studies where the experimentation is usually done in vitro and the objective is to find the safe dose for first-in-man study. The process follows with the experimentation on human and is divided in three phases. Moving from phase 1 to phase 3 the number of subjects included in the study increase, the duration of treatments is extended, the centres are added and the objective move from a safety to an efficacy evaluation. In each phase more than one clinical trial with the same objective may be performed. If the treatment pass the phase 3 and it is approved by a regulatory agency, such as EMA[1] for European Union or FDA[2] for United States, the drug is marketed. Every clinical trial needs to be planned and accurately detailed. All the information related to the trial are collected in a document called Protocol. The protocol should be reviewed and approved by the agency before the start of the study. Given the importance of a well written and detailed document the FDA and NIH[3] released a protocol template for phase 2 and 3 clinical trials. The protocol needs to include the rationale of the study,

[1] European Medicines Agency
[2] Food and Drug Administration
[3] National Institute of Health

the objectives, the methods and every other information needed to conduct the trial. During the planning of a clinical trial there are various aspects to consider, such as the design of the study, the randomization method or statistical techniques adopted, but it is important to describe every choice taken and the reason why has been taken. In addition to the Protocol the practical analysis is performed using the Statistical Analysis Plan (SAP), this is a document were technical details about the derivation of the variables, the statistical analysis method and the list of analysis needed is provided. Writing a Protocol and a SAP is a process that requires time and where experts of the different fields such as clinicians, statisticians, programmers should be involved.

Thesis contribution

In this thesis a review of an early stage clinical trial based on real data was performed. The study is a non-inferiority crossover design study, comparing a new drug versus a standard approved treatment. An evaluation of the techniques adopted and an overview on different methodologies was performed and described. Especially, the sample size estimation and the analysis of the primary endpoint based on parametric and non-paramedic techniques were proposed. The review was done in accordance to FDA guidelines and Good Clinical Practice (GCP) standard. The aim of this thesis is not only to present a specific statistical analysis for a single study but to provide an overview of the steps and the mechanics involved when planning a clinical trial form a practical point of view.

Structure of the thesis

In Chapter 1 an introduction on the types of study in medical research is provided. The focus will be on Clinical trials. Study Protocol of phase 2 and 3 Clinical trials and Statistical Analysis Plan will be presented in detail in Chapter 2. In Chapter 3, crossover design studies are introduced and finally in Chapter 4 the review and statistical considerations are presented for a real study. Conclusions and Appendix end this thesis. In the Appendix the SAS code written for implementing the methodology proposed in this thesis is presented.

Chapter 1

Types of study in medical research

The term medical research includes a wide range of different types of researches aimed to analyse the cause, the progress, the effect of diseases and to find the answer to the question "Does this intervention cure a disease, or relieve the symptoms?"
After a general overview on medical research we concentrate on clinical trials and the key concepts for planning a trial that is the subject of this thesis.

1.1 Classification of research

General classification of medical research may be done as follow:

- Experimental studies
- Quasi-Experimental studies
- Observational studies

1.1.0.1 Experimental studies

In experimental studies the experimental unit is manipulated by the investigator, randomization is one of the methods used to control data. They are divided in:

- Laboratory studies: are studies conducted in laboratory not done on human but usually used as prerequisite for human experiments. The experimental unit is exposed to a chemical, physical, biological or psychological stimulus in order to verify the variation of a parameter of interest.

- Clinical trials: are studies on human conducted to assess the efficacy and safety of a new drug or device. Clinical trials are the objective of this thesis and will be described in detail later in this chapter.

- Community intervention: are similar to clinical trials but the experimental unit is not a single human but the community.

1.1.0.2 Quasi-Experimental studies

In quasi-experimental studies the experimental unit is manipulated by the investigator, for example giving a treatment, but the randomization process is not used.

1.1.0.3 Observational studies

In observational studies there is no manipulation by the investigator. The aim of the research is to study the effect of a certain substance or condition on subjects exposed. The most famous observational studies are the cohort studies and case-control. The cohort studies are investigations done on a group of individuals experiencing a same event or having in common certain characteristic, where the researcher is interested on finding the relation between the common characteristic and the incidence of the disease. In case-control study the individual with the disease (case) is compared to an individual without disease (control) but similar for other characteristic such as age or sex.

1.2 Clinical trials

Clinical trials are medical research aimed to verify the tolerability and efficacy of a new treatment (drug, medical device) or to answer a question related to specific intervention. The subjects of the clinical trials are humans, either healthy volunteers, subjects affected by the disease or seriously ill patients depending on the phase of the experimentation.

1.2.1 Phases

A new drug, to be approved and marketed, needs to follow a four steps process plus pre-clinical studies. Each phase corresponds to one or more clinical trials, the approval on a specific stage is the mandatory requirement for entering the following phases.

1.2.1.1 Pre-clinical studies

Pre-clinical studies are the only researches where the experimentation is not done on human but *in vitro* or on animal *in vivo*. The aim is to find the toxicity of a new drug and to estimate a safe starting dose for human. Despite the animal testing in the research have been decreased during the past year for ethical reasons, some studies still involve animals because of similarity in anatomy with human.

1.2.1.2 Phase 1

Phase 1 studies are the first experimentations on human. The number of subjects recruited is usually small (near 20) and they are usually healthy volunteers or final-stage patients whose have had all the treatments available. The aim of the trial is to verify the safety and tolerability on human, recording every side effect. Three different types of study may be identified:

- SAD (Single Ascending Dose study): the goal is to find the maximum tolerated dose (MTD). The subject receives an increasing dose step by step during time, the trial is stopped when a predefined dose is reached or side effects occurred.
- MAD (Multiple Ascending Dose): during these trials the subject receives multiple low doses, the dosage is then increased till a specific predefined level.
- Food Effect: trial evaluating the drug absorption with or without the ingestion of food.

1.2.1.3 Phase 2

In Phase 2 clinical trials the number of subjects involved in the analysis is increased (can be over 200 participants) and the period of examination is extended. The objective is the evaluation of the efficacy of the treatment in addition to confirmation on safety and tolerability already given during phase 1.

1.2.1.4 Phase 3

Phase 3 trials are designed to evaluate the efficacy of the treatment. The number of subjects recruited significantly increases and the duration may be from few months to years, depending on the disease and the availability of patients. The cost of the study is enormous considering the amount of time

and people involved but if a drug succeeds this phase, i.e. the treatment is demonstrated to be effective, it could be submitted to a regulatory agency and then be commercialized. The agency during the study or preliminary phase may request clarification or modification on the methodology of the trial, this leads to an exchange of answers/questions between agency and sponsor that should finally pass the review of the agency. Further, it may request more than one positive clinical trial on a drug to be approved, indeed is not uncommon that for a specific drug/disease two identical clinical trials are conducted at the same time, saving time in case of effectiveness but wasting money in the eventuality that one of the two is negative.
Even if a drug is marketed it may be reported a serious adverse effects, in this case the drug needs to be recalled immediately from the market and further studies need to be planned.

1.2.1.5 Phase 4

Phase 4 studies are post-marketing researches, made on large population or sub-population like pregnant women which could not have been tested before for ethical reasons. These trials evaluate the long term effect of the treatment.

1.2.2 Planning a clinical trial

When planning a clinical trial there are different aspects that have to be taken into account. They depend on the phase of the study, the disease (variable of interest) and the treatment(s) to be tested. In this section the principal components of clinical trial are described.

1. Primary endpoint
   The first important decision is the selection of primary endpoint. The primary endpoint is at the base of the planning of the study, it is the parameter that allows the investigator to claim or deny the effectiveness of the drug, and all the subsequent decisions have to be done in accordance to what is considered to be the primary goal of the trial. The primary endpoint could be a single variable or a combination of different endpoints. The variable of interest may be quantitative, qualitative or also the time until a specified event happens.
   The primary endpoint of the study presented in this thesis will be described in section 4.1.2.

2. Design of the study
   Given the objectives of the study and the interventions to be analysed the first aspect to consider is the design of the study, three big categories are described below:

   - Parallel design : the patient is randomized to one treatment and remains on that treatment throughout the duration of the trial. The number of treatments is arbitrary; it can be placebo or an existing treatment versus a new drug, or more than two different treatments or also the same treatment with different dosage.
   - Crossover design : the patient is randomized to a sequence of treatments, i.e. he/she receives all the treatments in the study in a specific order and the randomization is used to determine the order.
   - Factorial design : the patient is randomized to the different drugs as in parallel study plus to take placebo and the combination of the drugs. The factorial design aims to study the effect of the intervention but also the interactions between different treatments.

   This thesis will focus on crossover design, the theoretical aspect will be described in Chapter 3 and the application in a real study will be presented in Chapter 4.

3. Aim of the study
   If the trial compares a new treatment against an already used drug a second distinction on the design is outlined as follow:

   - Non-inferiority : the aim is to evaluate if the new treatment is not much worse than standard treatment.
   - Equivalence : the aim is to evaluate if the new treatment behave in a similar way of standard treatment.
   - Superiority : the aim is to evaluate if the new treatment is better than standard treatment.

   The word "worse", "similar", "better" in practical are tested by choosing a level of tolerability, if the variable of interest cross this level than the hypothesis of the trial is rejected.
   The study presented in this thesis is a non-inferiority trial, the reason of this choice will be presented in section 4.1.

4. Sample size estimation
   The third question is "How many subjects should be recruited in order

to validate or reject my hypothesis with a certain confidence?" the answer is what is called "sample size estimation". The sample size estimation is the determination of the number of subjects that have to be included in the study for having statistically meaningful results on the primary endpoint analysis. There are different procedures that could be used to estimate the size and they are based both on the nature of the primary endpoint variables and on the assumption that could be done based on a priori consideration or results from previous studies.
The sample size estimation of the study presented is described in section 4.1.8.2.

5. Randomization
   The randomization assigns the subjects to the different arms of treatment randomly. It is a powerful instrument that creates comparable groups of subjects. Different types of randomization could be performed, some of them are:

   - Simple randomization : each patient has equal probability to be assigned to one treatment arm.
   - Block randomization : n patients are assigned to k treatments with block of size m.
   - Stratified randomization : the patients are divided into subgroup based on factors assumed to be prognostic for the endpoint. A simple or block randomization is done for every combination of factors.
   - Minimization: is the only non random technique, the patients are randomized in order to minimize the difference among groups.

   The simple randomization method is used in the study of this thesis as described in section 4.1.8.1.

6. Blindness
   Another aspect is the blindness that is the procedure of hiding the treatment received to the subject and/or to other individuals involved in the trial.
   Open label defines studies where neither the patients nor the investigator are blinded, everybody knows exactly which treatment is administered.
   In *Single blind* studies the treatment is masked only to the patients, the investigator knows the real assignation.

In *Double blind* studies both the investigator and the patients are not aware of the real treatment assignation. The single blind or double blind trial are often used because reduce the risk of bias due to influences from the preferences or expectations of the investigator. Bias in the real effect of treatment may also arise from patient in the so called "placebo-effect". The placebo effect or placebo response is a phenomenon in which a placebo, or in general a inert substance, can sometimes improve a patient's condition simply because the person has the expectation that it will be helpful.
In blinded study, the list of randomization, that contain the name of the real treatment given, is connected to the data only at the end of the study where all subjects have been recruited, all information are available and thus the database is closed. An early loss of information may determine the termination of the study. However there is the possibility of breaking the code for a patient before the end of the study. This is admitted only in case of medical emergency for which knowing the treatment received is fundamental for the patient cure.
The study of this thesis is a double-blind trial.

1.2.2.1 Adaptive design

The adaptive design clinical trials are specific trials in which modification of the design is allowed during the course of the study and it is based on the results obtained from the data itself. The modifications need to be prespecified at the beginning but the general structure is more flexible compared to the common trial.
The development of these methodologies arises from the necessity of creating new instrument for the evaluation of drugs or medical device that makes the study more efficient for example reducing the cost and the duration.
The adaptive designs are very attractive for both safety and efficacy point of view, the advantages, if applied correctly, are remarkable. They may reduce the time and the cost of clinical trials obtaining the same results as common design but with a more efficient method with the advantage of reducing the risk of the subjects due to exposure. Although there are many advantages, the application of adaptive strategies is still a matter of study. In particular the major concerns are related to the choice of the correct statistical model and the control of type I error in presence of modifications of the design. For these reasons in 2010 FDA released draft guidelines for researchers who aim to conduct a clinical trial with adaptive method.
The term "adaptive" includes a wide range of different kind of modification, the most commons are:

- Adaptive randomization design: the adaptive randomization is a procedure in which the probability to be assigned to a treatment is not fixed a priori but it changes during the study, i.e. it is adjusted based on the number of subjects already assigned to a treatment.

- Adaptive dose finding : the adaptive dose finding design is used to find the minimum or maximum tolerated dose. Few subjects are randomized to receive different dosages, the dose-response is evaluated and then the more subjects are allocated to the most-informative doses.

- Sample size re-estimation design : this adaptive method consists on the re-estimation of the sample size based on the data observed during interim analysis. The adjustment is done to preserve the initial type I error and power.

- Group sequential adaptive design : the group sequential design allows the premature stop of the trial if there are strong evidences of efficacy or futility of the treatment. The stopping rule is planned at the beginning of the study.

Chapter 2

Clinical trials: documentation and data

The general introduction presented in Chapter 1 gives an overview of the different types of clinical trials and the different methodologies that could be applied in medical research, but it is important to understand the phases of the experimentation and what is requested for starting a clinical trial. The conduction of a clinical trial can be simplified in the following steps: recruiting subjects, drug administration, evaluation of the safety and efficacy. Every single step needs to be planned and detailed before the start of recruiting, in order to minimize the probability of error and thus undermining the validity of the study. For these reasons, a series of documents containing all information related to trial need to be produced and detailed. The first and most important document is called *Protocol* this is the key document and every action is performed based on what is written in it. The protocol provides also a general explanation of the statistical analysis to be produced, but the detailed description in terms of derivation of the variables and programming actions is described in the *Statistical Analysis Plan*; the SAP could be considered as the technical instruction to produce the analysis of the protocol from the data collected. The document for understanding how data are collected is the *Case Report Form* (CRF). In this chapter we describe what protocol, SAP, CRF are, how they are structured and their usage.

2.1 Protocol

The protocol is the key document of experimentation, it contains the reason at the basis of the trial, the aim of the study, the planning of the research

and all information required to conduct the study or the reference documents needed. Given the importance of a complete and well written protocol the FDA[1] and NIH[2] in May, 2017 released a protocol template for phase 2 and 3 clinical trials that require Investigational New Drug applications (IND) or Investigational Device Exemption (IDE) applications, an earlier version opened to comments was released in 2016, some of them were incorporated in the latest version. Following the template structure and the guideline "Good Clinical Practice (GCP)" released during the International Conference on Harmonisation (ICH) the content and the order of the items are described in the section below.

2.1.1 Structure of the protocol

The first page should include the protocol number, the title (easy to remember), the investigator, the sponsor and the version number. All versions should be recorded with the date and a rationale on the reason of the updates. The following chapters are structured as shown below:

1. Protocol summary: this includes the synopsis, where there is a short study description with the objectives, the primary endpoint, the participants, the duration and the phase. A graphical schema of the study and the scheduled activities (study visit and study endpoints) is presented.

2. Introduction: name and description of the product under study, a summary of the previous clinical trials or pre-clinical studies to justify the starting of the trial, description of the population and the risks or benefits.

3. Objectives and endpoints: description of the objectives and endpoints of the study in addition to the reason why the endpoints were chosen. Includes study visit or time point at which data will be collected for the analysis.

4. Study design: description of the design of the study (e.g. double-blinded, open-label, crossover, dose-finding, superiority, non-inferiority), the number of treatments, methodology for avoiding bias (e.g. randomization, blindness), single centre or multi-centre, end of study definition.

[1]Food and Drug Administration
[2]National Institute of Health

5. Study population: description of the study population with the criterion of exclusion and inclusion and the recruitment strategies. Definition of screening failure subject.

6. Study intervention: description of the interventions to be tested (more than one intervention could be tested), description of the preparation of the intervention (storage, packaging and labeling), administration and schedule, admitted and not admitted medications during and before the trial, rescue therapies.

7. Study intervention discontinuation and participant discontinuation/withdrawal: description of reasons and criteria for study intervention discontinuation or study discontinuation.

8. Study assessments and procedures: description of all assessments or evaluations used to calculate the efficacy of the intervention, how they are collected and when, criterion for safety evaluation's including adverse event.

9. Statistical consideration: description of the statistical methodologies applied. Sample size determination, selection of subjects for the efficacy analysis, methodology for missing data handling, statistical consideration and analysis of primary/secondary endpoint, analysis of safety, planning and methodology of subgroup or interim analysis and exploratory analyses

10. Supporting documentation and operational considerations: in this section operational considerations are described such as the approval of the protocol by the institutional review board (IRB) or independent ethics committees (IECs), the protocol amendment approval process, the obtaining of the informed consent and the final reporting of study results and publication.

The references to the articles or books used in the protocol ends the document.

2.2 CRF

A Case Report Form (CRF) is a printed or electronic (eCRF) document used by the sponsor to collect data required by the protocol from each trial subject. Usually it is designed after the protocol is finalised but it can also be prepared during the development of the protocol, thus a control version

is adopted. The sponsor is in charge of designing the CRF, it need to be accurate and detailed in order to represent the request of the protocol. The electronic CRF is preferred to paper CRF because it is designed is such a way to reduces the risk of duplicated pages or data entry errors and it facilitates the collection of data when multiple centres are involved.

2.3 From Protocol to SAP

After the protocol has been reviewed and approved, all the documents are ready (e.g. informed consent, CRF) and all the people involved are well informed on the procedure to follow, the trial may start with the recruiting of subjects.
During the conduction of the study interim analyses on the blind or unblind data are performed. The analyses to be performed are described on a document called Statistical Analysis Plan (SAP), different SAP for the different purposes should be created:

- CCM (Central Clinical Monitoring): are periodical blind analyses usually produced every month (or two) where mostly safety evaluations are requested. They are performed internally and are used to review the data, check if information are well collected or to evaluate possible variables of interest.

- DSUR (Development Safety Update Report): are blind analyses requested every six months or year from regulatory agencies for monitoring the safety of the drug. They contain few outputs, indicatively less than 10 tables, on demographic, exposure, death and adverse events.

- DSRC (Drug Safety Review Committee): are blind analyses performed internally with a frequency that may vary from 3 months to one year. The DSRC SAP is an extended version of DSUR SAP, the number of outputs is greater and in addition to demographic, exposure and adverse events are also included laboratory, ECG and vital signs assessments and abnormalities.

- DMC (Data Monitoring Committee): this is the only unblinded analysis and is produced by independent experts. The analysis includes a large number of outputs related to both safety and efficacy evaluations. The role of the independent group is to evaluate the safety of the patients in relation to the treatment received. The members of DMC group have a high responsibility role since they can decide for

the termination of the study if they found significant risk to the safety of the subjects.

- CSR (Clinical Study Report): this is the final analysis used to produce the clinical study report document. Includes all analyses needed in the study and they are performed few time with blinded data in order to verify that everything is computed in the correct way and that possible error in the data have been solved. The final unblinded CSR analysis is performed after the database lock. The unblinded results will be used for evaluation of the effectiveness of the drug.

Given the importance of the CSR SAP a presentation of the content of CSR SAP is provided in the following section.

2.4 SAP

The Statistical Analysis Plan is a document written after the protocol is finalized, where more technical details about the inferential analyses and statistical method outlined in protocol are given. The SAP includes explicit guidance for the programmers on how to perform the analysis and the list of tables, listings and figures needed for the final report. The document is reviewed by the programmers, biostatisticians and clinicians during the course of the study and can be updated after a blind review of the data, in any case it should be finalized before the unblinding of the data. If the blind review of the data results in a need of updating the protocol then a protocol amendment has to be done and approved.
In the ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) guidelines [9], especially E9 (Statistical Principles for Clinical Trials), is outlined the importance of writing a detailed SAP, but contrary to protocol there isn't approved version of SAP template. Mimicking the structure of the protocol a list of information that needs to be included in SAP is given. The list of content follows the ICH guidelines E9 and E3 (Structure and Content of Clinical Study Reports), and the CONSORT guidelines (Consolidated Standards of Reporting Trials).
The first page contains the purpose of the study (CCM, DMC, CSR etc), the title as on the protocol, the investigational drug, the sponsor, the author of the document and the reviewers with their role, the version of the document and the status (Draft, Final).

1. Study design and flow: description of the study design, the randomization method and the sample size calculation. Graphical flow of the

conduction of the study and visit assessments.

2. Objectives: definition of primary, secondary and other efficacy endpoints as outlined in the protocol.

3. Change or clarifications to analyses planned in study protocol: in this section minor deviation from the protocol are described. If major change are needed, they has to be done in a protocol amendment and not here.

4. Definition of variables: detailed description of variables displayed in the outputs and their derivation. It includes safety and efficacy assessments.

5. Protocol deviation: definition of the protocol deviation. The protocol deviations should be grouped also in category for example "PV at study entry" "PV during study".

6. Analysis set: definition of analysis sets and reason for exclusion. In this section it is also provided their usage in the analysis for example the safety evaluations are done on the treated set instead the efficacy on the mITT.

7. Subgroup: definition of subgroup, if applicable.

8. Statistical Analyses: description of statistical methodology and implementation for the analysis of the endpoints of the study, it includes imputation of missing data.

9. General definition: list of definitions or derivations used such as baseline, treatment start and end date, end of study.

10. Handling of missing/incomplete date and time field: description of the imputation rules used for date and time variable

11. List of summary tables, listings and figures: this includes the list of the outputs to be produced indicating the name, the title, the analysis set and the reference to the shell document.

12. References and Appendices: the SAP ends with the list of references used for the analysis and in the appendices further detail on statistical consideration may be described.

Together with SAP the layouts shell documents is provided. It contains the exact layout for every tables, listings and figures included in the SAP.

2.5 Data collection

The data collected from the medical records needs to be pulled together and prepared for the statistical analysis written in SAP. A common standard for data submission for clinical studies, required by some regulatory agencies such as FDA, PMDA and used by EMA, is the CDISC standard [1]: SDTM and ADaM.
SDTM stands for Study Data Tabulation Model and corresponds to the translation of the data from the medical record to dataset. The SDTM is composed by a group of datasets and a metadata to describe their content. Each dataset corresponds to a specific domain and its name is a clue of the content; for example the DM dataset contains demographic characteristics, the AE dataset is related to adverse event, EX to exposure etc. The creation of the dataset needs to follow specific rules regarding the variables to be included their name and the length. The structure is subject oriented, this means that usually there is one raw per subject, per visit and per assessments. The number of record is not fixed except for DM domain where a single raw for each subject is required.
ADaM (Analysis Data Model) are the final datasets used for the creation of outputs. They are created from SDTM according to what is required in SAP and like SDTM they need to follow specific standard. The name of the dataset is 4/5 letters long, the first two letters correspond to prefix AD for Analysis Dataset the other letters are related to the contents. ADSL is the subject-level dataset it is created from DM domain adding information related to treatment exposure, baseline characteristics, population flags and all other subject specific information. The other ADaM datasets may follow two general structures, BDS and OCCDS.
BDS, basic data structure, it is used for many analysis datasets and the structure is suitable when there is one or more records for analysis parameters for timepoint, an example of the use is for laboratory or vital sign assessment. The variables included in BDS, in addition to subject identifier or variables coming from ADSL such as subject specific variables (i.e. treatment, population flags etc), are:

- Analysis Parameter: includes all variables used to identify the parameter, for example PARAM (parameter name), PARAMCD (parameter short name), PARAMN (unique numeric identifier of the parameter), PARCAT(category of the parameter).

- Analysis Timepoint: includes all variables for describing the timepoint, for example ADT/ATM/ADTM (date, time, datetime of the assessment), VISIT (name of the visit), VISITNUM (numeric identifier of

the visit).

- Analysis Value: includes the variables for the results of the assessment, for example AVAL (numeric value of assessment), AVALC (character value of AVAL), DTYPE (derivation type).

Some of the variables such as PARAM and PARAMCD are required some are permitted but not explicitly required.
In table 2.1 is presented a simple example of BDS structure applied to laboratory assessment. The USUBJID variable is the unique subject identifier, each row represent the information collected during a specific visit; the ABLFL flag is the baseline flag that is the value of the parameter before the start of treatment. The BDS structure could be used also for efficacy endpoint, in this case the name of the dataset is indicated with ADEFF or also for time to event analysis and the common name used is ADTTE.
The other standard structure is OCCDS used for Occurrence Dataset. The OCCDS is used when one or more "events" may happen to a subject and the interest is to analyse how many subjects have this given term. For OCCDS datasets as for BDS the variables included are the subject identifier or subject specific variables (from ADSL) and in addition:

- Event term: includes all variables used to identify the event, for example the term reported in the CRF (–TERM) and the term coded with a specific dictionary (–DECOD).

- Event Timepoint: includes all variables for describing when the event occurred, for example the event start date (ASTDT) and end date (AENDT).

One example of occurrence dataset is ADAE. ADAE is the adverse event dataset, it is created from AE and merged with ADSL. In Table 2.2 is presented a simple example of Adverse Event dataset, AETERM is the term of the adverse reaction as reported in CRF instead AEDECOD is the coded term using the MedDRA (Medical Dictionary for Regulatory Activities) dictionary, AEBODSYS is the body system organ class, ASTDT and AENDT are the start and end date of the adverse reaction, and TRTEMFL is the flag indicating whether the event occurred during the drug administration or not.

USUBJID	AETERM	AEDECOD	AEBODSYS	ASTDT	AENDT	TRTEMFL
idsubj1	BRONCHITIS	Bronchitis	Infections and infestations	01JAN17	16JAN17	
idsubj1	RUNNING NOSE	Rhinorrhoea	Respiratory, thoracic and mediastinal disorders	03FEB17	07FEB17	Y
idsubjn	COUGH	Cough	Respiratory, thoracic and mediastinal disorders	13FEB17		Y

Table 2.2: Example of ADAE, Analysis Dataset for Adverse Event

USUBJID	VISIT	VISITNUM	ADT	ATM	ADTM	PARAMCD	PARAM	ABLFL	AVAL	BASE	CHG
idsubj1	1	Screening	01JAN17	13:15	01JAN17T1315	ALB	Albumin	Y	120		
idsubj1	2	Week 2	16JAN17	17:00	16JAN17T17:00	ALB	Albumin		130	120	10
idsubjn	3	Week 4	3FEB17	16:50	03FEB17T16:50	ALB	Albumin		130	120	10
.....											
idsubjn	3	Week 20	10MAY17	13:00	10MAY17T13:00	HGB	Hemoglobin		131	132	-1
idsubjn	8	Week 24 /EOT	16JUN17	11:30	16JUN17T11:30	HGB	Hemoglobin		105	132	-27

Table 2.1: Example of ADLB, Analysis Dataset for Laboratory assessment

The conversion from SDTM to ADaM is performed for every domain needed in the analysis.
The description of CDISC standard presented in this section is just a general overview of the method, for understanding the potentiality or for learning the rules for creating datasets it is possible to download the ADaM implementation guide from the CDISC web site [1]. The importance of having a worldwide standard for data is enormous, it facilitates the exchange and the submission of data, it enables the interoperability of clinical research, and it reduces the time for setup and conducts a study and is easier to understand. Even if SDTM and ADaM seem quite simple, creating a correct structure and metadata requires a considerable amount of planning and effort.

2.6 From data to CSR analysis

The description given in the previous sections provides the ingredients needed in a clinical trial, the interesting part is also understanding how exactly they are combined together and the timing of the process. Given a specific question: "Does this medication cure the disease?" the first thing to do is writing the Protocol, this may takes a lot time because as shown in Section 2.1 this is the base document of the trial, many experts are involved in this process from programmers, statisticians, pharmacologists for review pre-clinical studies or clinicians in general. In addition to the protocol is created the informed consent document to be signed by the subject before entering into the study. The protocol and the informed consent should be submitted for review to the regulatory agencies and to institutional review board (IRB)/ ethics committees (EC) in order to ensure that the study is planned following the ICH guideline and subjects are exposed to minimal risks compared to any benefits that might result from the research. If the review teams approve the documents the study may start. Together with the protocol the CRF is created this should be well-designed in order to allow the easy collection of data relevant for the analysis. When all the documents have been reviewed and approved and every people well instructed and informed on the procedures to adopt, the enrolment of the subject may start. The starting of the enrollment means that the first data based on the designed CRF are collected. The data management team collects all the data coming from different centres and merges them together for creating datasets with SDTM structure as described in previous section. The statisticians start developing the CCM SAP, this SAP is used for checking data if there are unexpected situation and for preparing the final analysis, when the SAP is finalized the statistical programmers start to develop the analysis. The statistical pro-

grammer called "main" converts the SDTM datasets in ADaM structure and from them produces all the analysis requested in the SAP, in a double blind study this analysis are blinded. Another programmer called "validator" reproduces independently the same datasets and analysis as requested in SAP and check if the main programmer results are correct, this process is called validation. When the validation process is terminated the programmer informs the statisticians. The statisticians review the analysis and may decide to modify the SAP or request changes to the program based on the results of the data. This process is repeated during the course of the study and it is used for preparing the final analysis. Close to the end of the study the CSR SAP for the final analysis is prepared. The final CSR SAP should be approved before the end of the study, when the last data are collected. In a double-blind study when the last data are collected is called "database lock" this means that data can't be further modified, after the database lock the programmer receive the real list of randomization and attach it to the data, at this point the final analysis is produced. The ADaM and the outputs are created and validated, the statisticians start the review of the final results and when approved they could be released to the external. From the database lock to the release of the first outputs usually takes 3 days, all the outputs (over than 100) should be then produced in one week.
The process described above is really short and simplifying summary of the actions and stages for producing a clinical trial analysis. Many people are involved and the timing is really demanding, thus the high preparation and the experience are fundamental in this process.

Chapter 3

Crossover design

The parallel clinical trial are commonly used to design a clinical research but the biggest issue related to this design is the requirement of a great number of subjects and consequently needs more time and resources. Sometimes, due to the kind of disease, is not easy to find a sufficient number of recruitable subjects and thus parallel design is not always feasible, a solution is the crossover design. In this chapter a general description of crossover design and its characteristics is provided, some of the procedures presented will be applied in Chapter 4 in a real study.

3.1 Crossover design: overview

Crossover trials are studies where the experimental unit (patient) receives different treatments in the different periods of study, crossing over from one treatment to another. The simplest crossover design is the AB/BA design where two treatments are given in two periods, subject are randomized to take first A and then B or the opposite as shown below.

Design	Period 1	Period 2
sequence	A	B
sequence	B	A

Adding treatments or number of periods is possible to obtain more complex designs for example the 3-treatments 3-periods may correspond to the sequences:

Design	Period 1	Period 2	Period 3
sequence	A	B	C
sequence	B	A	C
sequence	C	A	B

or also it is possible to incorporate non-crossover sequence like in the Balaam's design as shown below.

Design	Period 1	Period 2
sequence	A	A
sequence	A	B
sequence	B	A
sequence	B	B

It is logical that the more treatments/periods are incorporated the more the design is complicated and difficult to analyse.

Advantages

The first advantage of crossover design compared to parallel study is the fact that the patient serves as her/his own control. This is useful because removes the between subjects variability and in this way the influence of confounding factors is reduced. The second advantage is that requires fewer patients to obtain the same power compared to parallel study and as a consequence lower resources are needed.

Disadvantages

On the other side there are limitations of choosing this kind of design. The first is related to the type of disease, it has to be a long term disease with the primary endpoint of reducing symptom instead of cure the disease. Indeed, for example, assuming that during the first period treatment A cured the disease, in the second period there is no possibility to demonstrate the efficacy of treatment B or also if the treatment in first period lasts longer than the disease, the effect on response in the subsequent period can't be evaluated. The second limitation is the carryover effect; it is defined as the effect of the previous treatment period on the response of the current period therefore it may affect the final results.
The drop-out subjects are another aspect to be considered. A drop-out, i.e. subject who prematurely discontinued the study, is an inconvenient also in parallel study even if it could give information on the reason of the discontinuation, in crossover design is extremely difficult to gain advantages from subject who discontinued because information are not directly linked to the treatment or sequence if for example the subject discontinued during the first period.

Carryover

As mentioned before one of the biggest disadvantages of crossover design is the carryover effect, but what can be done to limit it? A common solution is the introduction of a wash out period that is a time interval between the treatment periods in order to "prepare" the subject for the second administration. The issue is that is not always clear how much longer the wash out should be in order to eliminate the effect of the previous treatment. The second issue is that the carry-over effect of the treatment could be different, making the interpretation of the results. The suggestion is to apply quite long wash out period in order to not apply statistical correction at the end. The very last chance if a strong carry-over is showed is to perform the analysis only taking the data of the first period, in this case the results and the power of the analysis are questionable.

3.2 Sample size estimation for AB/BA design

Sample size estimation is the first step in clinical research. The sample size should be computed based on the parameter of interest and the hypothesis tested in order to obtain statistical inferences with a certain confidence. The sample size estimation should be divided for qualitative and quantitative outcome and also for testing the non-inferiority, superiority or equality of a drug A compared to B. The sample size computation is presented for a generic significance level α and power $1 - \beta$.

3.2.1 Continuos Outcome

Assuming that the variable of interest is a continuous outcome, we should distinguish between parametric and non-parametric techniques.

3.2.1.1 Parametric

Using a parametric method the mean response can be used as indicator of worsening or improvements. The hypothesis can be written as follow:

Equality:

$$\begin{aligned} H_0 &: \mu_A - \mu_B = 0 \\ H_1 &: \mu_A - \mu_B \neq 0 \end{aligned} \tag{3.1}$$

Non-inferiority/Superiority:

$$\begin{aligned} H_0 &: \mu_A - \mu_B \leq (\geq)\delta \\ H_1 &: \mu_A - \mu_B > (<)\delta \end{aligned} \tag{3.2}$$

where
μ_A = *mean of treatment A*
μ_B = *mean of treatment B*
δ = *non-inferiority/superiority margin*
If we assume a normal distribution of the response, then the difference in mean with known variance can be estimated by a t-student [2]. The formula is then approximated by:
Equality:

$$N = \frac{\sigma^2(z_\alpha + z_\beta)^2}{2(\epsilon)^2} \tag{3.3}$$

where
z_α = *value of Normal distribution at* α *level*
σ = *is the standard deviation*
ϵ = *is the estimated treatment effect under H1*

Non-inferiority/Superiority:

$$N = \frac{\sigma^2(z_\alpha + z_\beta)^2}{2(\mu_A - \mu_B - \delta)^2} \tag{3.4}$$

Has been demonstrated [9] that the use of a normal approximation to the t-student may result in an overestimating of the power or alternatively underestimating the sample size. The solution is given by the use of a non-central t-distribution, the formula for non-inferiority(superiority) is given by:

$$T_{2n-2}\left(t_{\alpha,2n-2}|\frac{2n(\mu_a - \mu_B - \delta)}{\sigma}\right) = \beta \tag{3.5}$$

where
α = *significance level*
β = *1- power*
$t_{\alpha,2n-2}$ = *value of t distribution at* α *level with 2n-2 degree of freedom*
μ_A = *mean of trt A*
μ_B = *mean of trt B*
δ = *non-inferiority margin*
σ = *is the standard deviation*

This equation is solved using the distribution values in tab. 3.1.

	α=2.5%		α=5%	
	$1-\beta=$		$1-\beta=$	
θ	80%	90%	80%	90%
0.30	176	235	139	191
0.32	155	207	122	168
0.34	137	183	108	149
0.36	123	164	97	133
0.38	110	147	87	120
0.40	100	133	78	108
0.42	90	121	71	98
0.44	83	110	65	90
0.46	76	101	60	82
0.48	70	93	55	76
0.50	64	86	51	70
0.52	60	79	47	65
0.54	55	74	44	60
0.56	52	68	41	56
0.58	48	64	38	52
0.60	45	60	36	49
0.65	39	51	30	42
0.70	34	44	26	36
0.75	29	39	23	32
0.80	26	34	21	28
0.85	23	31	18	25
0.90	21	27	16	22
0.95	19	25	15	20
1.00	17	23	14	18
1.05	16	21	12	17
1.10	15	19	11	15
1.12	14	18	11	15
1.15	13	17	11	14
1.20	12	16	10	13
1.25	12	15	9	12
1.30	11	14	9	11
1.35	10	13	8	11
1.40	10	12	8	10
1.45	9	12	7	9
1.50	9	11	7	9

Table 3.1: Values of $T_{2n-2}(t_{\alpha,2n-2}|\sqrt{n}\theta/\sqrt{2}) <= \beta$

3.2.1.2 Non-Parametric

When information on data suggest that a normal distribution can't be assumed then a non-parametric techniques can be used. Noether in 1987 [13] and subsequently Rahardja in 2009 [15] provided an estimation of the sample size based on Wilcoxon–Mann–Whitney test, traduced in the following formula:

$$N = \frac{(z_{\frac{\alpha}{2}} + z_{\beta})^2}{12t(1-t)(\pi_1 - 0.5)^2} \tag{3.6}$$

where
$\alpha =$ *significance level*
$\beta =$ *1- power*
$t =$ *fraction of subject treated* n_A/N
$\pi_1 = Pr(A > B) + 0.5Pr(A = B)$, *where A and B are random variables with cumulative density function* F_A *and* F_B *, respectively, referring to trt A and B.*
Formal estimator of π_1 was given in [15].

3.3 Statistical analysis for AB/BA design

The statistical analysis of crossover design depends on the nature of the data, i.e. qualitative or quantitative variable. Consider a crossover AB/BA design where the aim is to evaluate if A is equal to B, in the following section an overview of the methodologies will be presented

3.3.1 Binary Outcome

Suppose that the variable of interest is a binary outcome, where the results is given by 0 for failure and 1 for success. In Table 3.2 the possible results are presented. Our hypothesis is to test whether the probability of preferring A

Group	(0,0)	(0,1)	(1,0)	(1,1)	Total
1(AB)	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1.}$
2(BA)	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	n

Table 3.2: AB/BA binary crossover design

(i.e. response 1) is equal to preferring B, in mathematical annotation could

be written as:

$$H_0 : \pi_A = \pi_B \tag{3.7}$$

A statistical test that can be used is the McNemar's test. This test considers the number of patients who prefer (i.e. response 1) each treatment regardless of the order the treatments were received in and ignoring those patients who do not express a preference. If we define $n_p = n_{.2} + n_{.3}$ the total number of subject that show a preference and $n_A = n_{13} + n_{22}$ the subject who have a preference for A under the null hypothesis, n_A has a binomial distribution with parameters 1/2 and n_P. The p-values is then given using the equation of the binomial distribution under the alternative hypothesis that $n_A/n_p > 1/2$ (one-tailed test) the equation is given by:

$$P = \sum_{r=n_A}^{n_p} \binom{n_p}{r} \frac{1}{2}^r \frac{1}{2}^{n_p - r} \tag{3.8}$$

For n large the binomial could be approximated with the normal distribution.

3.3.2 Continuous Outcome

Suppose now that the variable of interest is a continuous outcome, our hypothesis in mathematical annotation could be translated in this way:

$$H_0 : \mu_A - \mu_B = 0 \tag{3.9}$$

T-test

The simplest way of analysing a crossover design is using a t-test. Let Y_{ij} be the response of the i^{th} subjects during period ($j = 1, 2$), n_{AB} number of subjects in sequence AB, n_{BA} number of subjects in sequence BA and define $d_i = y_{i2} - y_{i1}$ the difference between period 2 and period 1 irrespective of the sequence, then the treatment effect can be estimated by:

$$T = \frac{\bar{d}_{AB} - \bar{d}_{BA}}{\hat{SE}(\bar{d}_{AB} - \bar{d_{BA}})} \tag{3.10}$$

where
$\bar{d}_{AB} = \sum_{i \in AB} d_i$
$\bar{d}_{BA} = \sum_{i \in BA} d_i$
$\hat{SE}(\bar{d}_{AB} - \bar{d_{BA}}) = s_d \sqrt{\frac{1}{n_{AB}} + \frac{1}{n_{BA}}}$ and $s_d = \sqrt{\frac{(n_{AB}-1)s^2_{d_{AB}} + (n_{BA}-1)s^2_{d_{BA}}}{n_{AB}+n_{BA}-2}}$ with

$s^2_{d_{AB}}$, $s^2_{d_{BA}}$ the estimated variance of the differences of the sequences.
Assuming the normality distribution of the y and the equal variance the test statistics will have a t-distribution with $n_{AB} + n_{BA} - 2$ degrees of freedom. From the table of t-distribution is then computed the p-value and the hypothesis of equality is then validated or rejected.
The α confidence intervals for the mean difference are computed as follow:

$$\frac{1}{2}(\bar{d}_{AB} - \bar{d}_{BA}) \pm \frac{1}{2} t_{\alpha/2}(n_{AB} + n_{BA} - 2)\hat{SE}[\bar{d}_{AB} - \bar{d_{BA}}] \tag{3.11}$$

T-test is easy to compute but needs normally distributed data to performed, assumption that is not always easy to do.
In the non-inferiority scenario that is the design used in the study presented in this thesis, the hypothesis is adjusted in the following way:

$$\begin{aligned} H_0 &: \mu_A - \mu_B \leq \delta \\ H_1 &: \mu_A - \mu_B > \delta \end{aligned} \tag{3.12}$$

δ is considered as the inferiority margin. Simirarly, as above the treatment effect is estimated by:

$$T = \frac{\bar{d}_{AB} - \bar{d}_{BA} - \delta}{\hat{SE}(\bar{d}_{AB} - \bar{d_{BA}})} \tag{3.13}$$

In the non-inferiority test the result is evaluated using the confidence interval, indeed for claiming the non-inferiority the lower bound of the confidence limit should be greater than the margin. This means that with a certain confidence (alpha level) the new treatment is not much worse than the standard.

Linear mixed model

The linear mixed model is an extension of the linear model where the relation between a response and covariate is still linear but the assumption of a single slope coefficient is relaxed giving the possibility to vary across the individuals and be predicted by the covariate. The model is composed from a fixed part that are the observed covariate (qualitative or quantitative) and a random part that usually is the individual experimental units.
The mixed model can be written as:

$$Y = X\beta + Z\gamma + \epsilon \tag{3.14}$$

where
Y is the vector of responses
X is the fixed-effects design matrix

β are the fixed effects
Z is the random-effects design matrix
γ are the random effects
ϵ is the vector of errors
The assumption of the model is that γ and ϵ are normally distributed with mean 0 and variance: $\begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$
The estimation of the covariance can be done by using a likelihood-based method, two methodologies can be used: the maximum likelihood (ML) or the restricted/residual maximum likelihood (REML). ML method underestimates the covariance parameters because it assumes that the fixed parameters are known without uncertainty. The REML method updates the likelihood function in order to estimate the covariance parameters independently of the estimates for the fixed effects.
The fixed effect is then computed by minimizing the mixed model equations obtaining: $\beta = (X^t\hat{V}^{-1}X)X^t\hat{V}^{-1}y$ where V is the variance of y, this is called generalized least squares.
In crossover design Y is the variable of interest, X is the matrix of the observed effects that is the sequence (i.e. AB or BA), the period (1 or 2) and other possible covariates of interest for example the value of the parameter of interest at study entry and finally Z are the subjects.
The advantages of using a model is the possibility to include possible explanatory covariates.
In the hypothesis of the non-inferiority, the confidence limit of the coefficient related to treatment effect will be evaluated. As for the t-test if the lower bound of the CL is above the non-inferiority margin then the null hypotheis will be accepted.

Wilcoxon–Mann–Whitney test

The Wilcoxon-Mann-Whitney test is a non parametric approach for analysis of continuous data. This test is useful when parametric assumption on the distribution of the data can't be derived. Using the annotation described above in t-test section for Y_{ij} and $d_i = y_{i2} - y_{i1}$, then equation 4.1 in a non-parametric view can be translated as:

$$H_0 : F_{d_{AB}}(.) = F_{d_{BA}}(.) \tag{3.15}$$

where F is the cumulative distribution function of d_i.
The statistics is then computed pooling the estimated differences for both sequences and assigning a rank to each observation, then equation is then

written as:

$$W = \frac{\sqrt{12}(R_{AB} - n_{AB}(n_{AB} + n_{BA} + 1)/2}{\sqrt{n_{AB}n_{BA}(n_{AB} + n_{BA} + 1)}} \tag{3.16}$$

where R_{AB} is the sum of the rank in sequence AB.
Under the null hypothesis 3.15 W is normally distributed with mean 0 and variance 1, p-value is computed using the standard normal tables. The confidence interval for the treatment effect are estimated using the Hodges-Lehmann estimator [14]. The non-inferiority could be evaluated by adding the non-inferiority margin to the traetment cosidered to be non-inferior and then the statistics will be computed as shown above calculating the one-sided p-value.

Chapter 4

Crossover design: a case study

In this chapter we present the application of the crossover design in a real study. First a presentation of the disease and the primary objective is provided, then we will try to go through the point of the protocol as described in Chapter 3, the analysis of secondary endpoint will not be presented.

4.1 Case study

The case study of this thesis is non-inferiority, early stage, multicentre, randomized, double-blind, active comparator-controlled and crossover trial. For confidentiality reasons, due to undergoing process of submission, actual information on drug and disease cannot be disclosed. Thus, only methodological and statistical review will be presented. The hypothesis of the trial and the data will be slightly modified in order to not connect to the specific trial, but maintaining the overall statistical and methodological consideration. Considering that data and parameters have been modified, this work should be considered as an example of application of a crossover design and not a presentation of the real study results.
The objective is to evaluate if a new drug A is non-inferior to standard drug B in improving the disability in subjects affected by a chronic worsening disease.
The standard drug B is the only approved treatment in commerce for the disease under study, the non-inferiority design is the best configuration in order to find an alternative treatment, because contrary to superiority the requirement are lower.

4.1.1 Introduction

The disease of this study is a chronic progressively worsening disease, the prevalence is estimated to be 3 per 200,000 and incidence 10 per million years. Usually affects subjects from 20 to 70 years and the median onset time is around 50 years. The standard and first line treatment used is drug B, the placebo-controlled trials have demonstrated an improvement on disability and the safety of that treatment.
Drug A is in the same therapeutic class of drug B, pre-clinical studies have demonstrated the safety profile of the drug and clinical studies have confirmed the safety and demonstrated the efficacy on other related diseases. Given the positive results obtained on the safety of drug A, the objective is to provide data on efficacy in treatment of the disease.

4.1.2 Objectives and endpoints

The primary objective of this study is to evaluate the efficacy of drug A compared to drug B for treatment of patients with this worsening disease. The variable used for this evaluation is the sum of 5 scores representing the level of disability on a specific parameter measured in the body. Each score scale ranges from 0 to 10 in 1 unit increments that represent lower levels of disability. The hypothesis that will be tested is the non-inferiority, i.e. drug A is not much worse than drug B in improving the disability. The primary endpoint chosen is commonly used to evaluate the improvement or worsening of the disease on subjects.
The secondary objective is to have confirmation on safety. The safety evaluation is assessed recording the adverse effects happened during treatment period and evaluating the change from baseline to post-baseline assessment in vital sign or laboratory parameters.
Exploratory objective is to evaluate if the speed of worsening in drug A is lower than drug B. The evaluation is done using the same variable used in the primary endpoint evaluation.
Other objectives included in the protocol will not be presented in this thesis.

4.1.3 Study design

The study is a non-inferiority, early stage, multicentre, randomized, double-blind, active comparator-controlled and crossover trial. The diagram is showed in Fig. 4.1.
Considering the low prevalence of the disease, the issue was to find adequate number of evaluable subjects to be included in the analysis, this is the major

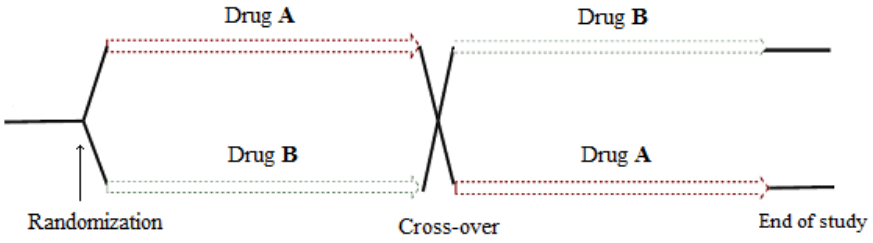

Figure 4.1: Study design

reason why the crossover design has been chosen. Indeed the crossover requests a lower number of subjects compared to parallel study to obtain the same power and is particularly suitable when the endpoint is the change in some parameters and not the cure from the disease.
The study includes 2 consecutive periods:
The Screening period will be completed in a maximum of 30 days prior to Randomization. During this period patients meeting the eligibility criteria, if they agreed to participate, should sign the informed consent. Baseline disease characteristics and demographics information will be collected in this phase.
The treatment period will last from randomization up to Week 32. Subjects are randomized to follow drug A for 16 weeks and subsequently drug B for other 16 weeks or vice versa.
The end of study visit will take place from 4 to 8 weeks after the last administration of drug.
Carryover effect bias is not expected in this trial.
Even if the drug is considered to be safe an independent DMC will review blinded and/or unblinded data on a regular basis during the trial.

4.1.4 Study population

Twenty subjects will be randomized in order to have at least 18 subjects who have an evaluable efficacy assessment. The patients included are all subjects affected by the disease aged from 18 to 70 that did not receive the standard treatment or other investigation product in the last month. Other

exclusion criteria were considered and there were related to potentially risk of the subject.

4.1.5 Study intervention

Active drug and controlled have been provided as identical film-coated tablets to be taken once a day.
The drug will be supplied in a blid-labeled bottle containing 28 tablet for 4 weeks of treatment.

4.1.6 Study intervention discontinuation and participant discontinuation/withdrawal

All patients were free to withdraw from participation in this study at any time, for any reason, and without prejudice. Patients who withdraw their consent will be asked to undergo the study assessments scheduled at the end of study visit. The participation to the end of study visit is voluntary.

4.1.7 Study assessments and procedures

4.1.8 Statistical analysis

Recalling previous sections the disease of this study is chronic and progressively worsening disease, the aim of the trial is to evaluate the non-inferiority of a new drug A, compared to the standard drug B. Patients were randomized in a ratio 1:1 to sequence AB or BA and then were followed for about 16 weeks in each period. The primary endpoint is a continuous outcome indicating the level of disability. Therefore the hypothesis to be tested can be written as follow:

$$\begin{aligned} H_0 &: \mu_A - \mu_B <= \delta \\ H_1 &: \mu_A - \mu_B > \delta \end{aligned} \tag{4.1}$$

where μ_A is the mean of the response variable in drug A, μ_B is the mean of the response variable in drug B and δ is a fixed level of non-inferiority. The chosen level was δ = -1, the level was considered to be clinically significant.

4.1.8.1 Randomization

Patients meeting the eligibility criteria were randomized in a 1:1 ratio in the two groups corresponding to the sequence AB or BA. The simple randomization method was used.

Comments: in parallel study the decision of simple randomization method compared to other more complex methodology is evidence that no strong relation between the primary endpoint and a possible covariate is supposed. If otherwise, previous studies or clinicians showed a prognostic factor, this should be included in the randomization process avoiding creation of incomparable groups.

4.1.8.2 Sample size estimation

The sample size was estimated to be of 18 subjects, accounting for a drop-out rate of 10% a total of 20 subjects have been recruited. The sample size estimation was done considering a non-inferiority margin of 1 point on the mean difference between the two treatment groups. The trial was planned in order to have a power of 90% with a level of significance $\alpha = 2.5\%$ and based on the assumption of no difference between the two treatments mean. The last assumptions, computed from previous studies was the within subject variance to be 3.2.
Considering the information provided from clinicians and from previous studies related to the mean and variance, the non-parametric approach is not a useful technique because we don't have the information required in equation 3.6. The assumption of a normal distribution is not denied from previous studies, so replacing in 3.3 the values given above we obtain:

$$N = \frac{3.2(1.96 + 1.29)^2}{2(0 - (-1))^2} \approx 17 \tag{4.2}$$

if we apply the exact equation defined in 3.4 and using Table 3.1 with our value on the non-central t-distribution with $k = 1$ and $\theta = \frac{2(\mu_a - \mu_B - \delta)}{\sigma}$ we obtain 18 subjects. In crossover design with a low prevalence disease, i.e. expecting a lower number of subjects, is more advisable to use the non approximated equation.

4.1.8.3 Analysis set

Four different analysis sets were defined and each analysis set was used for a specific analysis:
Randomized set : The randomized set consists on all subjects who underwent randomization.
All-Treated Set : This analysis set includes all randomized subjects, who took at least one dose of study drug.
Modified Intention-To-Treat set: all treated subjects, with the baseline and at least one post treatment baseline efficacy assessment available.

Per Protocol Set: This analysis set comprises all subjects included in the Modified Intention-To-Treat set who did not violate the protocol in a way that might affect the evaluation of the effect of the study drug on the efficacy endpoint. The list of protocol deviation were given in the study.

The Randomized set was used for subject's listings, in this way for all subjects included in the study is possible to know the data collected. Safety evaluations were done on All-Treated set because only subjects who received a treatment are potentially at risk and the safety profile need to be explored, the Modified Intention-To-Treat set was used for the analysis of the primary endpoint because include all subject treated with at least one efficacy parameter available for the evaluation and finally the Per Protocol set was used on secondary endpoints where more strict condition have been applied. Subjects included in the different population were summarized in a table as shown in Fig. 4.2.

Study: <Study Name>
<Title>
Analysis set: <analysis set>

	Drug A N = XX n (%)	Drug B N = XX n (%)	Total N = XX n (%)
Randomized Set	xx (yy.y)	xx (yy.y)	xx (yy.y)
All-Treated Set	xx (yy.y)	xx (yy.y)	xx (yy.y)
Modified Intention to Treat	xx (yy.y)	xx (yy.y)	xx (yy.y)
Per Protocol Set	xx (yy.y)	xx (yy.y)	xx (yy.y)

Figure 4.2: Summary of analysis set

Twenty subjects were randomized, one subject randomized to sequence BA discontinued the study after the start of treatment and so he was not included in the Modified Intention-To-Treat set and Per Protocol Set.

4.1.8.4 Demographics, baseline and disposition

The first analysis performed in clinical trial is the description of the population. Demographics such as age, sex, race, height, weight, BMI at screening were presented in a table as shown in Fig. 4.3.
The same was done for baseline disease characteristics, for example the disability score or other disease specific parameters evaluated at screening visit have been showed.

Study: <Study Name>
Demographic
Analysis set: <analysis set>

	Sequence AB N = XX	Sequence BA N = XX	Total N = XX
Age at screening (years)			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Q1, Q3	xx, zz	xx, zz	xx, zz
Min, Max	xx, zz	xx, zz	xx, zz
Sex [n (%)]			
Male	xx (yy.y)	xx (yy.y)	xx (yy.y)
Female	xx (yy.y)	xx (yy.y)	xx (yy.y)
<Parameter>			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Q1, Q3	xx, zz	xx, zz	xx, zz
Min, Max	xx, zz	xx, zz	xx, zz

Figure 4.3: Summary of demographics

No substantial difference where highlighted during the subject characteristic analysis.
The baseline information are presented only in a descriptive way.

4.1.8.5 Exposure

The exposure of the subject to treatment is evaluated in term of duration of study treatment. Reasons for discontinuation were collected and displayed in a summary table. The exposure was evaluated per drug considering the sum of the two periods and displayed as in Fig. 4.4.
All subjects, except for the patient who previously discontinued, have been treated as planned.

4.1.8.6 Safety evaluation

The onset of adverse events during the treatment period was evaluated as safety parameter. The treatment emergent adverse event are those adverse event started during the treatment period, these were summarized by period, drug and overall as shown in Fig. 4.5.
The number of AE recorded for drug A was quite similar to drug B, especially nine during administration of drug A and ten during drug B. The same summary was done on serious and related treatment emergent adverse event, no serious adverse event was registered. Vital sign parameters including systolic and diastolic blood pressures (mmHg), heart rate (beats/minute) and body temperature (C) and laboratory parameters including hemoglobin

Study: <Study name>
<Title>
Analysis set: <analysis set>

	Drug A N = XX	Drug B N = XX	Total N = XX
Duration of study participation			
n			xx
Mean			xx
SD			xx
Median			xx
Min, Max			xx, zz
Duration of exposure			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Min, Max	xx, zz	xx, zz	xx, zz

Figure 4.4: Summary of exposure

(g/L), platelets (109/L), AST (IU/L), ALT (IU/L), ALP (IU/L), Creatinine (umol/L), total and direct bilirubin (umol/L) were also considered for safety analysis. The absolute value at each post-baseline visit as shown in Fig. 4.6 and the change from baseline to end of study (Fig. 4.7) were evaluated. No statistical analysis was performed to compare the value, the manual evaluation didn't highlight substantial differences in the different phase of treatment.

4.1.8.7 Statistical analysis for the primary endpoint

The primary endpoint is to evaluate if treatment A is not inferior to treatment B, using as level of non-inferiority -1 and comparing the mean difference for A minus B. If the lower limit of the confidence limit for the mean difference is bigger than -1 then the non-inferiority is demonstrated.
The first consideration to be done in crossover design is the evaluation of carryover effect, i.e. the assessments during the second period may be influenced by the first period. A solution is to plan a long wash-out interval between the two phases of administration, in this way the data of the second period are "cleaned" from the influence of the first. The wash-out should be planned at the beginning of the study but it was not done in this trial because no carryover effect is expected, in any case a statistical evaluation of the carryover effect was performed. This is done in combination with the

Study: <Study Name>
<Title>
Analysis set: <analysis set>

System Organ Class Preferred Term	Number of subjects n (%) e Period 1 Drug A N = XX	Drug B N = XX	Period 2 Drug A N = XX	Drug B N = XX	Total Drug A N = XX	Drug B N = XX
Subjects with at least one <event>	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<System Organ> Class 1	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> 1	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> 2	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> n	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<System Organ> Class 2	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> 1	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> 2	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> n	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<System Organ> Class n	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> 1	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> 2	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx
<Preferred term> n	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx	xx (yy.y) xx

Figure 4.5:

analysis on primary endpoint; in the following section we will see different methodology.
The analysis was performed using SAS 9.3 version, the code is available in Appendix.

Two sample t-test:

The first approach that can be used is a two sample t-test for the difference in mean as described in Chapter 3. The variable of interest is the mean of the 5 sum score during the evaluation period of each treatment. In SAS 9.3 we can already compute the t-test for crossover design. First the carryover effect could be tested, a t-test is applied to the sum of the two periods for each subject the p-value is given by:

	p-value
Difference seq BA/AB	0.1564

The non significant p-value for the sequence implies that a carryover effect has not been highlighted. Then a two sample t-test for comparing the treatment is applied:

Mean	95% CL
-0.2060	(-0.6811, 0.2690)

as we can see the lower limit of the confidence limit is bigger than -1 showing the non-inferiority of treatment A. Introducing the option `ho=-1` the non inferiority test is performed giving the result:

Study: <Study name>
<Title>
Analysis set: <analysis set>

	Drug A N = XX	Drug B N = XX	Total N = XX
Period 1 Visit 1			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Q1, Q3	xx, zz	xx, zz	xx, zz
Min, Max	xx, zz	xx, zz	xx, zz
Period 1 Visit 2			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Q1, Q3	xx, zz	xx, zz	xx, zz
Min, Max	xx, zz	xx, zz	xx, zz
Period x Visit n			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Q1, Q3	xx, zz	xx, zz	xx, zz
Min, Max	xx, zz	xx, zz	xx, zz

Figure 4.6: Summary at each post-baseline visits in safety parameters

	tValue	p-value
Non-inferiority	3.50	0.0012

The rejection of the primary hypothesis is the demonstration of the non-inferiority of treatment A compared to B. In case of evidenced treatment sequence effect the analysis could have been produced by replacing the mean of the 5 sum score with the last available assessment per period.
A negative aspect of the t-test is the impossibility of taking into account the baseline value or other baseline characteristic that may be of interest in the analysis.

Linear mixed model:

The second solution consists of applying a statistical model to the response variable. As anticipated in previous section, the model is a more advisable choice because introduce the possibility of adding covariates of interest.
The model chose in this trial is the linear mixed model described in Chapter 3 and includes the treatment, the period and the sequence as fixed factors

Study: <Study name>
<Title>
Analysis set: <analysis set>

	Drug A N = XX	Drug B N = XX	Total N = XX
Baseline			
n			xx
Mean			xx
SD			xx
Median			xx
Q1, Q3			xx, zz
Min, Max			xx, zz
End of Study			
n	xx	xx	xx
Mean	xx	xx	xx
SD	xx	xx	xx
Median	xx	xx	xx
Q1, Q3	xx, zz	xx, zz	xx, zz
Min, Max	xx, zz	xx, zz	xx, zz

Figure 4.7: Summary of change from baseline to end of study in safety parameters

then the baseline 5 sum score is added as continues covariate, patients are considered as random factor. The baseline 5 sum score is the value obtained before the first drug intake of period 1. The results of the fitted model are shown in Table 4.1. The lower limit of the confidence interval is bigger than

Covariate	Least Squares Means Estimate (95% CI)	Differences of Least Squares Means Estimate 95% CI	p value
Treatment			
A	44.5 (43.48, 45.59)	-0.01 (-0.51, 0.48)	0.9640
B	44.5 (43.61, 45.49)		
Treatment period			
Period 1	44.4 (43.36, 45.49)	-0.24 (-0.73, 0.25)	0.3224
Period 2	44.7 (43.74, 45.59)		
Treatment sequence			
Sequence AB	45.0 (43.94, 46.05)	0.90 (-0.85, 2.65)	0.2959
Sequence BA	44.1 (42.58, 45.61)		

Table 4.1: Linear mixed model results

-1, confirming the non-inferiority of A versus B. The p-value of the sequence is not significant meaning that there isn't an appreciable difference of taking

A before B or vice versa. The least square means gives an estimate of the mean of the response variable for a specific setting (example: mean in drug A or mean in period 1) controlling for the other covariates. As we can see from the table the means of drug A and drug B is quite similar with a non significant p-value. The p-value of non inferiority was computed adding 1 to the mean of the 5 sum score at the end of the treatment period only for treatment A, the mean of treatment B was left unchanged. The resulting p-value is shown in Table 4.2, since it is below the significance level the null hypothesis of inferiority is rejected thus A is considered non-inferior to B.

	p value
Non-inferiority A versus B	0.0002

Table 4.2: Non-inferiority tests

Non-parametric :

During the planning of the study a strong deviation from the Normal assumption may be observed from previous studies or supposed from the nature of the data, in this case a non-parametric approach could be used. The Wilcoxon test described in Chapter 3 is applied as for the t-test to the difference period 1 - period 2. First a comparison between the periods is done by applying the test to the sum of period 2 and period 1, the p-value obtained us given by:

	p-value
Difference seq BA/AB	0.2906

The Wilcoxon test is not a test based on the mean, thus is not possible to obtain a confidence interval for the mean treatment difference. The non-inferiority could be tested with a one-tail test adding the estimated level for non-inferiority to treatment A, the results is given by:

	p-value
Non-inferiority A versus B	0.0027

This presentation is only for describing different methodologies and their application, in a clinical trial the statistical method should be planned at the beginning.

4.1.8.8 Exploratory endpoint

An interesting aspect that could be evaluated in presence of a variable that may increase or decrease is the relationship with time, in this case how fast the new drug compared to standard treatment is beneficial to the subject. For doing this evaluation a random coefficient model is applied. The random coefficient model is an extension of the linear mixed model described in Chapter 3 and applied previously where the time is incorporated as random effect. The time is not considered as a fixed covariate with a corresponding slope but the slope may vary with the subject. The model is therefore
Fixed effect: time, treatment, treatment*time
Random effect: subject, subject*time
In cross-over trial, the subject is allocated to different treatment over-time thus it is not meaningful to include all data in the model, an evaluation of the two periods is performed separately.
Period 1, Reference Drug A

Effect	Estimate	SD	p-value
time	0.02934	0.05665	0.6102

Period 1, Reference Drug B

Effect	Estimate	SD	p-value
time	0.07509	0.06347	0.2502

The time was evaluated in weeks, the rate of increase of the 5-score in mean per weeks in treatment B is 0.03 instead in treatment A 0.06, and the values are both non-significant.
Period 2, Reference Drug A

Effect	Estimate	SD	p-value
time	0.01836	0.01703	0.2940

Period 2, Reference Drug B

Effect	Estimate	SD	p-value
time	-0.01438	0.01548	0.3636

In period 2, as for period 1 there is no significant difference in the rate of change, thus the only conclusion is that Drug A is not much different from drug B in improving the disability.

4.2 Conclusion

In this study more than 80 tables and 30 figures were programmed and validated, as discussed above they include subject characteristic at study entry, safety evaluation (adverse events, vital signs and laboratory parameters), exposure and study treatment discontinuation/ interruption, analysis of primary endpoint and even if they were not mentioned secondary and exploratory endpoints. This study example showed a non-inferiority of drug A and the safety profile of the drug was confirmed.

Discussion

In this thesis we have presented a randomized controlled trial application in a real life context.
Clinical trials are the techniques used to demonstrate the safety and efficacy of a new drug or a medical device. The difficulty of planning a medical research is to find the correct way to set up the study and the appropriate statistical method for the analysis. Furthermore the request of regulatory agencies for the submission of a clinical trial are becoming increasingly demanding and the direction is to have a worldwide standard method for the report of the trials.
The study of this thesis is a non-inferiority, early stage, multicentre, randomized, double-blind, active comparator-controlled and crossover trial. The choice of the design depends on different factors and is not automatically easy to find the best configuration. The disease under study is a chronic progressively worsening disease, with a really low prevalence and incidence. The issue was to have an adequate sample size to obtain a statistical power of 90%. The crossover design seemed to be the appropriate choice in this situation because compared to parallel study needs to recruit less subjects and is suited for researches where the aim is not to cure the subject but to relieve the symptoms. The non-inferiority to the standard drug was assessed comparing the mean of the sum of 5-score representing the level of disability on a specific parameter measured in the body, higher values of the score represent a lower disability. The second issue raised from the application of crossover design was the carryover effect that is the effect of the first treatment period on the second period. Usually this is reduced incorporating a wash out interval between the two periods, but this was not done in this study assuming from previous study improbability of carryover effect, anyway during the analysis of the primary endpoint the test for estimating the possible carryover has been described. For the analysis of the primary endpoint parametric and non-parametric techniques were presented to give a general overview of the different methodologies, but the model chosen was a linear mixed model because it allows the introduction of covariate and considering the subjects

as a random effect allow a more flexible structure. The non-inferiority of drug A was claimed in this example analysis. An exploratory endpoint was then proposed, when we are in presence of a variable that may increase or decrease over time, it may be of interest to evaluate the relationship with the time. The evaluation was performed using a random coefficient model that is an extension of the linear mixed model where the time is included in the analysis not as covariate but as a random effect. The variation over time including all data is not meaningful in crossover trial due to the sequence of treatment taken, the evaluation was then performed dividing the analysis in the two periods and comparing the results obtained. The results showed no significant difference on the rate of change in the two treatment, two period, it may be of interest of future research to find how to link the different results in presence of contrasting value and if it can be evaluated in term of a carryover effect. To complete the presentation of the study the description of the population, the subject characteristic and the safety parameters were presented, displaying how the results are summarized for external review. No serious adverse events or abnormal vital sign or laboratory parameter was collected during the administration of drug A.
The aim of this work was not only to present a clinical trial with its statistical consideration but to gives an overview of the difficulty that may arise in planning a clinical trial, and the steps and documents that has to be done.
The randomized clinical trials are the most common method used in clinical research, but is not always easy to find the perfect design that allow a powerful analysis of the data reducing the exposure of the subjects to potentially ineffective drug. Furthermore planning a clinical trial is expensive in term of cost and duration, and the results are not always positive even with a rigorous planning. Future works are aimed to apply more advanced techniques that may increase the efficiency of the analysis. This area of study is called adaptive design and it means the possibility to vary some parameters of the study, during the course of the trial, given the results obtained from the data itself. Given the importance of this techniques FDA in 2010 released a draft guidelines on adaptive design encouraging this new approach to drug trials. In the example of the crossover design, one "adaptive" application is based on the allocation of the subject to the most powerful treatment sequence, considering two-treatment two-period design the possible treatment combination are AB, BA, AA, BB, the adaptive cross-over design may assign the new subject to the sequence more safe or effective. This choice is ethical because more subjects are allocated to better treatment limiting the drop-out and avoiding the missing data issue. In a crossover trial a useful adaption could be the sample size re-estimation, considering the application in presence of rare disease, the treatment is usually long and as a consequences it

could be frequent the withdrawal of the subject, a simple size re-estimation could be the method to increase the number of subjects and the power of the analysis in situation where there is an inadeguate sample size. The concern of the adaptive design is the statistical procedure to be used and its validity changing the initial hypothesis. Although the results of adaptive design are still a matter of study because of the doubt on the statistical procedure and the interpretability of the results, the application in clinical trials is increasing over time.

Appendix A

SAS-code

A.1 sas

```
*dat dataset have one row for each subject and the variables are:
  subj = number of subject;
  seq = sequence of treatment (AB or BA)
  trt1 = treatment for period 1
  trt2 = treatment for period 2
  mp1= mean of endpoint of period 1
  mp2= mean of endpoint of period 2
  sump =  mp1 +  mp2
  diff=mp1-mp2
  diif2=mp2-mp1 (increasing the A values of 1 point);
* t-test carryover;
proc ttest data=dat;
 class seq;
 var sump;
run;

* t-test treatments difference;
  proc ttest data=dat ;
   var mp1 mp2 / crossover= (trt1 trt2) ;
   run;

* t-test non-inferiority;
  proc ttest data=dat1 ho=-1 alpha=0.025 side=U;
   var mp1 mp2 / crossover= (trt1 trt2) ;
   run;
```

```
*mixed model;
proc mixed data=dat2 method=reml empirical;
    class trt period seq subj;
    model  val = trt period seq base / s residual;
    random int / type=UN subject=subj;
    lsmeans trt period seq / pdiff diff alpha=0.05 cl;
run;


*non-parametric carryover;
proc npar1way data=datt wilcoxon;
class seq;
var sump;
run;

*non-parametric non-inferiority;
proc npar1way data=datt wilcoxon;
class seq;
var diff2;
run;

*random coefficien model;
proc mixed data=dat1;
 class trt subj;
model val= trt time trt*time/ solution ddfm=kenwardroger;
random int time/subject=subj type=un solution;
run;
```

Bibliography

[1] CDISC. Clinical Data Interchange Standards Consortium. `https://www.cdisc.org/`.

[2] S.C. Chow, H. Wang, and J. Shao. *Sample Size Calculations in Clinical Research, Second Edition*. Chapman & Hall/CRC Biostatistics Series. Taylor & Francis, 2007.

[3] Shein-Chung Chow and Mark Chang. Adaptive design methods in clinical trials – a review. *Orphanet Journal of Rare Diseases*, 3(1):11, May 2008.

[4] Andrea Cipriani and Corrado Barbui. What is a clinical trial protocol? *Epidemiologia e Psichiatria Sociale*, 19(2):116–117, 2010.

[5] FDA and NIH. Protocol Template. `https://osp.od.nih.gov/clinical-research/clinical-trials/`, 2017.

[6] PR Freeman. The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in medicine*, 8(12):1421—1432, December 1989.

[7] Joachim Werner Otto Gerß and Wolfgang Köpcke. *Clinical Trials and Rare Diseases*, pages 173–190. Springer Netherlands, Dordrecht, 2010.

[8] CONSORT Group. CONSORT checklist. `http://www.consort-statement.org/consort-statement/checklist`, 2010.

[9] David Harrison and Anthony R. Brady. Sample size and power calculations using the noncentral t-distribution. 4:142–153, 01 2004.

[10] HO Hartley and R Hocking. The analysis of incomplete data. *Biometrics*, 27:7783–808, 1971.

[11] M. Hills and P. Armitage. The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology*, 8(1):7–20, 1979.

[12] ICH. ICH Guideline. `http://www.ich.org/products/guidelines.html`.

[13] Gottfried E. Noether. Sample size determination for some common nonparametric tests. *Journal of the American Statistical Association*, 82(398):645–647, 1987.

[14] Mary E. Putt and Vernon M. Chinchilli. Nonparametric approaches to the analysis of crossover studies. *Statist. Sci.*, 19(4):712–719, 11 2004.

[15] Dewi Rahardja, Yan D. Zhao, and Yongming Qu. Sample size determinations for the wilcoxon–mann–whitney test: A comprehensive review. *Statistics in Biopharmaceutical Research*, 1(3):317–322, 2009.

[16] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[17] Stephen Senn. *Cross-over Trials in Clinical Research, Second Edition*. John Wiley & Sons, Ltd, 2003.

[18] Gail Tudor and Gary G Koch. Review of nonparametric methods for the analysis of crossover studies. *Statistical Methods in Medical Research*, 3(4):345–381, 1994. PMID: 7889227.

[19] Stefan Wellek and Maria Blettner. On the proper use of the crossover design in clinical trials. *Deutsches Arzteblatt international*, 109(15):276—281, April 2012.

[20] J. Woodcock. "fda introduction comments: clinical studies design and evaluation issues". *Clinical Trials*, 2, 2005.

[21] Fanghong Zhang, Etsuo Miyaoka, Fuping Huang, and Yutaka Tanaka. Test statistics and confidence intervals to establish noninferiority between treatments with ordinal categorical data. *Journal of Biopharmaceutical Statistics*, 25(5):921–938, 2015. PMID: 24919157.