Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

METODOLOGIA STATISTICA PER LA RICERCA SCIENTIFICA

Ciclo XXVII

Settore Concorsuale di afferenza: 13/D1

Settore Scientifico disciplinare: SECS-S/02

STATISTICAL METHODS FOR DEALING WITH SELECTIVE CROSSOVER IN RANDOMISED CONTROLLED TRIALS

Candidata: Dott.ssa Sara Balduzzi

Coordinatrice Dottorato

Prof.ssa Alessandra Luati

Relatrice

Prof.ssa Rossella Miglio

Correlatore

Prof. Roberto D'Amico

Esame finale anno 2016

A tutti coloro a cui voglio bene

Table of contents

Acknowledgements

First of all, special thanks go to Prof. Rossella Miglio and Prof. Roberto D'Amico, for having given me the opportunity to work on this project: I hope to keep on working with you, your support, guidance, patience and understanding on this and many other projects in the future.

I want to thank Dr. Elisabetta Petracci, who offered me her precious help from the beginning.

Elisabetta, Linda, Giovanni, Elena, Arianna: it was an honour (and pretty fun!) to share this path with you.

My life at work, which I love, would not be the same without my beautiful colleagues: Cinzia, Elena, Nika, Roberto V. Thank you for your support and for listening to me every time I need it.

An important thank goes to my parents – I know you are my biggest fans.

Thanks to Alberto, especially for letting me buy a desk and a shelf: without them it would have been much more difficult to write this thesis.

Last four years have surely been the most intense ones of my life. I would like to thank all the people who walked by my side, at least for a moment, during this period.

I am happy because I can understand how lucky I am.

Introduction

The evidence to support the efficacy and safety of a drug derives from randomised controlled trials (RCTs) [1]. The ethical basis of a RCT is provided by the principle of equipoise, stating that the randomization is allowed when there is genuine uncertainty on the actual benefit of the intervention under study.

An RCT ideally should permit an unbiased estimation of the treatment effect (e.g., overall survival (OS) or disease-free survival (DFS)), through the comparison between an experimental treatment and a comparator, administrated in two separate arms. However, sometimes patients may be offered the possibility to cross over from one arm of the trial to the other, a phenomenon usually called selective crossover (SCO), in order to distinguish it from the studies in which crossover is planned. Other terms to refer to this switch are drop-in or cross-in [2]. It can occur as a consequence of the diffusion of (un)favourable results of one treatment that is being compared, e.g. from interim analysis or from concurrent studies. In both cases the equipoise principle is challenged. If so, the investigators may offer patients the opportunity to cross over to the arm where the more promising treatment is administered.

This thesis considers treatment crossover as the switching of patients randomised to the control group of an RCT on to the experimental treatment, at a certain point after randomisation.

The occurrence of SCO breaks the randomization process and may give rise to problems in the data analysis and interpretation of results. In fact, in presence of SCO, the Intention to Treat (ITT) analysis, that is the analysis considering groups as randomised, will give an unbiased estimate of the experimental treatment assignment effect, but that effect will also

include the one of the experimental treatment administered to patients in the control group who switched. So the results of this type of analysis may not reflect the actual efficacy of the experimental intervention, and adjustments to allow for crossover are needed.

My personal interest in this topic derives from the work that I conducted for my master degree thesis, when I presented the results of a systematic review on the efficacy and safety of a drug, i.e. trastuzumab, for the treatment of early and metastatic breast cancer. The work of my colleagues and me has been published in two articles [3,4]. In both adjuvant/neoadjuvant and metastatic settings, we observed that more than half of the included studies let patients in the control arm cross over to the trastuzumab arm at a certain point after randomisation. While in the metastatic setting crossover was permitted generally after progression, making it difficult to estimate the actual benefit of trastuzumab on OS, in the adjuvant/neoadjuvant setting it was permitted before disease progression, making it difficult to analyse and interpret both OS and DFS results.

The present work aims to:

(i) assess the prevalence of SCO in RCTs assessing the efficacy and safety of biological and hormonal therapies for breast cancer and published in the scientific literature;

(ii) identify the reported statistical methods used to handle crossover, in particular when the effect of the experimental treatment is a time-to-event outcome;

(iii) assess whether different statistical methods provide different results and interpretations.

Overview of the thesis

Chapter 1 presents the review of the medical scientific literature carried out in order to assess the prevalence of the SCO in the field of breast cancer, and the approaches adopted to handle it. Details on the methods adopted to conduct the review are given, along with the presentation of the results, distinguished for adjuvant/neoadjuvant and metastatic setting. The Chapter closes with a discussion regarding the findings of the review and the needs for further research.

Chapter 2 attempts to collect the more relevant approaches that have been considered in literature to address treatment crossover.

Along with ITT analysis, other naïve methods include censoring patients at the time they cross over, or excluding them from the analysis. These methods will lead to biased results if the switching process is not random, because of selection bias. If patients who switch are picked completely at random, their exclusion from the analysis would not result in bias, but in the loss of power and increase of the uncertainty in the results. Instead, a switching process depending on patients' characteristics implies informative censoring, and censoring or excluding from the analysis patients who cross over will lead to biased results.

Another method described in literature is to consider the treatment as a time-varying covariate. This approach may be subject to selection bias, as groups may no longer be balanced after a patient is censored or excluded, and bias is likely if a patient's probability of switching treatment is related to their underlying prognosis.

Morden et al. [5] studied some statistical approaches for dealing with SCO. They considered naïve methods, as well as more complex ones, such as Robins and Tsiatis's Rank Preserving Structural Failure Time Model (RPSFTM) [6]. The key assumption, though not always reasonable, of the RPSFTM method is the so-called "common treatment

effect" assumption: it is assumed that switching patients experience the same treatment effect, from the time they start taking the experimental treatment, as patients randomised to the experimental group from the beginning.

Another approach considered in the paper by Morden et al., and evaluated in this thesis, is the one described in Loeys and Goetghebeur [8], who present a method for calculating the actual treatment effect in situations where all patients take their allocated treatment in one group, and compliance is assumed as "all-or-nothing" in the other. So, if a patient in this arm cross over, the switch is assumed to have happened right after the randomisation, and the patient is assumed to have only received the treatment he/she switched onto and not the treatment he/she was randomised to.

The inverse probability of censoring weights (IPCW) method, introduced by Robins and Finkelstein [9], not evaluated in the paper by Morden et al., is also considered in the present work. The IPCW method do not assume the "common treatment effect", but its fundamental assumption is the "no unmeasured confounders", that is the requirement of data on all covariates that might influence the crossover.

All these methods are presented in Chapter 2 of this thesis.

In order to assess the potential bias associated with the methods described in Chapter 2, the actual effect of an intervention under study needs to be known, so a simulation study was performed. The idea was to reproduce a two-arm RCT, similar to one of the trials emerged from the review of the literature of Chapter 1: this work focuses the attention on the adjuvant/neoadjuvant setting, where the crossover is usually permitted before disease recurrence. Chapter 3 details the simulation study design and Chapter 4 presents the results.

The main issues raised from the present work are discussed in Chapter 5, along with potential future developments of the research in this field.

Chapter 1

Review of the literature

In this chapter it is presented the review of the scientific literature conducted in order to evaluate the prevalence of the SCO among trials evaluating the efficacy and safety of the biological drugs and hormonal therapies for breast cancer, and the statistical methods used to handle it.

1.1.Methods

RCTs (phase III only) published between January 2000 and June 2015 in the Annals of Oncology (AoO), Journal of Clinical Oncology (JCO), Journal of the National Cancer Institute (JNCI), Lancet (L), Lancet Oncology (LO), New England Journal of Medicine (NEJM), assessing the efficacy and safety of biological drugs and hormonal therapy in breast cancer patients, were searched. We excluded chemotherapy agents because, in those years we selected as study time period, there were few new agents granting marketing authorisation: last innovative agents – doxorubicin pegylated and paclitaxel – were approved by the European Medicine Agency in 2000 [9].

The keywords used for the research were "random*", "breast" and "cancer" – with the restriction that all words had to be present in the title or in the abstract.

Trials were distinguished by the therapy setting they considered, i.e. adjuvant/neoadjuvant or metastatic. The prevalence was calculated as the percentage of trials in which crossover has occurred out of the total number of trials reflecting the inclusion criteria.

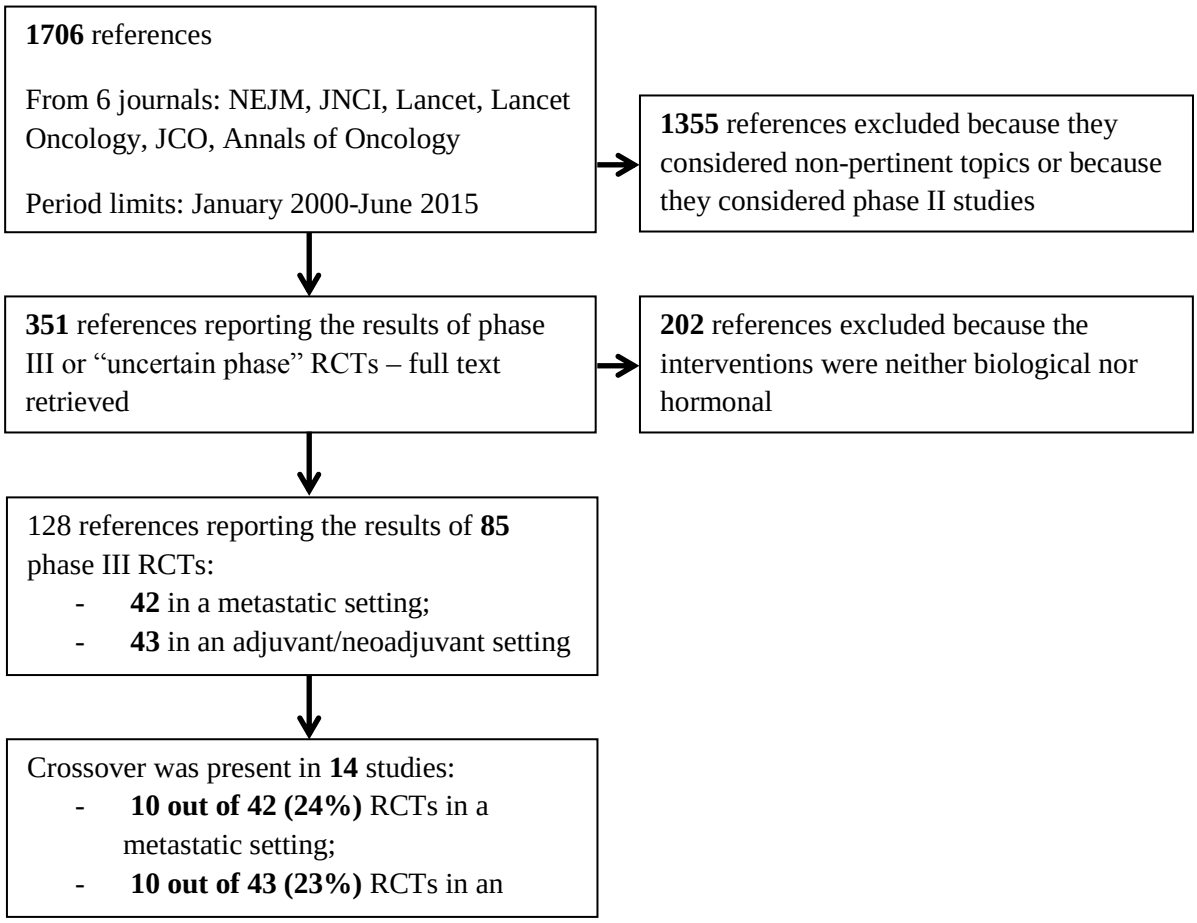
For trials in which crossover has occurred, the following characteristics were recorded: primary end point, reason for crossover (i.e. interim analysis), number of patients totally

randomized, number and type of patients allowed to cross over, statistical methods used to analyse results after the crossover. When reported, characteristics of patients who crossed over in respect to patients who remained in the arm of original allocation was also recorded.

If a trial reported the results for the same outcome (i.e. for overall survival (OS) or disease/progression-free survival (DFS/PFS)) in terms of hazard ratios (HR) from more than one type of analysis, the ratio of HRs (RHR), along with its 95% confidence interval (95%CI) was calculated in order to compare them.

1.2. Results

Figure 1. Flowchart of the bibliographic research



The flowchart of the literature research is presented in Figure 1. Out of 1706 references totally retrieved (AoO 434; JCO 553; JNCI 192; L or LO 104; NEJM 423), 1355 were excluded because they did not consider treatments for breast cancer or because they considered phase II studies. The full text of the remaining 351 references was retrieved: 202 were then excluded because the interventions were therapies neither biological nor hormonal and 21 because the study phase was unclear. The remaining 128 references reported the results of 85 RCTs, 43 of which enrolled women in early breast cancer and 42 metastatic.

Crossover was present in 20 RCTs (23.5%) equally distributed across settings: ten in adjuvant/neoadjuvant settings (23.3%) and ten in metastatic setting (23.8%).

When SCO occurred, the methods used to analyse data were:

1. Intention To Treat (ITT) analysis;
2. Censored analysis;
3. Inverse Probability of Censoring Weighting (IPCW) analysis.

In the ITT analysis, controls that crossed over are analysed as belonging to the control arm, despite the fact that they crossed over to the treatment intervention.

In the censored analysis, follow-ups of controls that switched to the intervention arm are censored at the time when the crossover occurred.

The IPCW analysis allows the estimation of the missing follow-ups of those controls who switched arm by using the information comprised in follow-ups of those controls who instead decided against crossing over and who were similar in terms of prognostic factors to their counterparts [8].

1.2.1. Adjuvant/Neoadjuvant setting

Characteristics of the ten RCTs (IMPACT, HERA, MA17, NSABP-B-33, BIG 1-98, NOAH, BCIRG-006, TEAM, B31, N9831) assessing efficacy and safety of treatments in the adjuvant/neoadjuvant setting in which SCO occurred are shown in Table 1.a. The studies B31 and N9831 were analysed and presented jointly, because they evaluate the same intervention, administered on a similar schedule. Five trials evaluated biological drugs and five hormonal therapies. Nine trials had DFS as primary outcome, one (IMPACT) had clinical tumour overall response. OS was always a secondary outcome. Two trials (IMPACT, NOAH) considered neoadjuvant strategies, and these are the trials with the lowest number of randomized patients (330 and 334 respectively). In five big trials (BIG 1-98, HERA, MA17, B31+N9831, with 8010, 3401, 5170, and 4390 randomized patients respectively), crossover was allowed after positive results obtained at a pre-planned interim analyses. HERA and MA17 were the trials with the highest percentage of patients crossing over, 52% and 61% respectively. In three trials (IMPACT, NSABP-B-33, TEAM) patients were permitted to cross over after results from other similar studies were published leading to a protocol amendment. For two trials (BCIRG-006, NOAH) the motivation for crossover was not reported. These trials had the lowest percentage of patients crossed over, 2.1% and 16% respectively. Other two trials (IMPACT, TEAM) did not clearly report the percentage of patients who crossed over. In two trials (BIG 1-98, HERA) the crossover was allowed only to patients who did not experience recurrence yet (HERA also required an adequate left ventricular ejection fraction); in the other studies, the crossover seemed to be offered to all patients in the control arm.

All the studies presented ITT analyses; two studies (BIG 1-98, HERA) conducted censored analysis; two studies (BIG 1-98, MA17) conducted the IPCW analysis. The only study which conducted all three analyses is the BIG 1-98: in 2009 the results of the ITT and

censored analyses were published and in 2011 another paper reported an update of the ITT analysis and the IPCW analysis.

Only two trials (HERA, MA17) reported the characteristics – in terms of age, previous therapy, menopausal status, hormone-receptor status, and lymph nodal status – of patients of the control arms who crossed over to the treatment arms, along with the characteristics of patients who did not. In both cases, patients in the SCO cohort compared with patients remaining in the control arm were more likely to be younger and have hormone-receptor-positive disease.

In table 2 are reported the results, expressed in HR (95%CI), for the three trials reporting censored or IPCW analysis, for DFS and OS respectively. The ITT analysis always seemed to be the more conservative one, although for the BIG 1-98 trial the RHRs both for OS and DFS were not statistically significant. OS results deriving from censored or IPCW analyses were more distant from the ones obtained with the ITT analysis in respect to DFS results.

1.2.2. Metastatic setting

Characteristics of the ten RCTs (Slamon 2001, Mouridsen 2003, TANDEM, AVADO, EGF104900, RIBBON-1, NCT00435409, NCT00938652, NCT00075764, CONFIRM) assessing efficacy and safety of treatments for metastatic breast cancer in which SCO occurred are shown in Table 1.b. All the trials but three (Mouridsen 2003, NCT00075764, CONFIRM) evaluated biological drugs. All the trials considered PFS as primary outcome and OS as secondary outcome.

All the trials' protocols permitted crossover to the experimental treatment arm for a patient in the control arm who experienced progression. RIBBON-1 was a four-arms trial, two controls and two experimental arms; after progression, patients in both the control arms were permitted to switch to the respective experimental arm. A particular case is represented by the Mouridsen 2003 trial, where two different hormonal therapies were

compared and in which patients at progression were permitted to switch to the other arm, irrespective of the arm in which they were initially allocated.

The percentages of switched patients were over the 40% in all the studies but three (AVADO 36%, NCT00435409 36%, CONFIRM 2%). All the studies reported the ITT analysis and two (EGF104900, Mouridsen 2003) conducted censored analysis.

Only one study (EGF104900) reported the main characteristics – age, performance status, prior therapies, hormone-receptor status – of the control arm by crossover status (crossover *versus* non-crossover), without statistically evaluating the differences between the two groups.

In table 3 are reported the OS results, expressed in HR, for the two trials reporting censored analysis. Mouridsen 2003 calculated the median time to death in each group, from which it was possible to estimate the HR; too less information was provided to calculate the RHR. The HRs from the ITT and the censored analyses reported by the EGF104900 trial do not seem to differ significantly.

Table 1.a – Characteristics of the trials in which SCO occurred – Trials assessing efficacy of treatments in an adjuvant/neoadjuvant setting

	Study	BD/ HT	Journal	Year of publication*	Year of reported SCO	Primary End Point	Motivation for SCO	Pts totally randomized/ to be enrolled	Pts crossed over	ITT	Cens	IPCW
1	IMPACT°	HT	JCO	2005	2005	Clinical tumour OR	After ATAC results	330/330	Not reported	√		
2	HERA	BD	NEJM,L,LO	2005-2011	2007	DFS	Interim analysis	5102/4482	885/1698 (52%)	√	√	
3	MA17	HT	JCO,AoO,NEJM	2003-2012	2008	DFS	Interim analysis, unblinding	5187/4800	1579/2587 (61%)	√		√
4	NSABP-B-33	HT	JCO	2008	2008	DFS	After MA17 (interim analysis) results	1598/3000	344/779 (44%)	√		
5	BIG 1-98	HT	NEJM,JCO,LO	2005-2011	2009	DFS	Planned interim analysis	8010/8028	619/2459 (25%)	√	√	√
6	NOAH°	BD	L	2010	2010	DFS (EFS)	Not reported	235/232	19/118 (16%)	√		
7	BCIRG-006	BD	NEJM	2011	2011	DFS	Not reported	3222/3150	23/1073 (2.1%)	√		
8	TEAM	HT	L	2011	2011	DFS	After IES results	9779/9300	Not reported	√		
9+ 10	B31+N9831	BD	NEJM,JCO	2005-2014	2014	DFS	Planned interim analysis	4390/4130	413/2018 (20%)	√		

* If more than one publication refers to the same study, the year of first publication and the year of last publication are reported
Pts=patients
NEJM=New England Journal of Medicine; JNCI=Journal of the National Cancer Institute; L=Lancet; LO=Lancet Oncology; JCO=Journal of Clinical Oncology; AoO=Annals of Oncology
DFS=Disease Free Survival; OR=Objective Response
HT=Hormonal therapy; BD=Biological Drug
° Neoadjuvant trials

Table 1.b – Characteristics of the trials in which SCO occurred – Trials assessing efficacy of treatments in a metastatic setting

	Study	BD/ HT	Journal	Year of publication*	Year of reported SCO	Primary End Point	Motivation for SCO	Pts totally randomized/ to be enrolled	Pts crossed over	ITT	Cens	IPCW
1	Slamon 2001	BD	NEJM	2001	2001	PFS	After progression	469/450	154/234 (66%)	√		
2	Mouridsen 2003	HT	JCO	2003	2003	PFS	After progression	907/907	233/458 (51%) + 226/458 (49%)	√	√	
3	TANDEM	BD	JCO	2009	2009	PFS	After progression	208/208	73/104 (70%)	√		
4	AVADO	BD	JCO	2010	2010	PFS	After progression	736/705	83/231 (36%)	√		
5	EGF104900	BD	JCO	2010-2012	2010	PFS	After progression	296/296	77/145 (53.1%)	√	√	
6	RIBBON-1	BD	JCO	2011	2011	PFS	After progression	1237/1200	112/206 (54.4%) + 105/207 (50.7)	√		
7	NCT00435409	BD	JCO	2013	2013	PFS	After progression	432/430	77/215 (36%)	√		
8	NCT00938652	BD	JCO	2014	2014	PFS	After progression	519/420	161/258 (62%)	√		
9	NCT00075764	HT	NEJM	2012	2012	PFS	After progression	695/690	143/345 (41%)	√		
10	CONFIRM	HT	JNCI	2014	2014	PFS	After progression	736/834	8/374 (2%)	√		

* If more than one publication refers to the same study, the year of first publication and the year of last publication are reported

Pts=patients

NEJM=New England Journal of Medicine; JNCI=Journal of the National Cancer Institute; L=Lancet; LO=Lancet Oncology; JCO=Journal of Clinical Oncology; AoO=Annals of Oncology

PFS=Progression Free Survival; EFS=Event Free Survival

HT=Hormonal therapy; BD=Biological Drug

° Neoadjuvant trials

Table 2 – DFS and OS results, expressed in HR (95%CI), for the three trials assessing efficacy and safety of treatments, in an adjuvant/neoadjuvant setting, reporting censored or IPCW analysis – results from the most recent publication for each trial are reported. RHRs (95%CI) were calculated in order to compare the results from different type of analyses (ITT as reference analysis).

Study	Type of analysis			RHR
	ITT	Censored	IPCW	
DFS				
HERA (2011)	0.76 (0.66;0.87)	0.69 (0.59;0.79)		0.91 (0.66; 0.87)
BIG 1-98 (2011)	0.86 (0.78;0.96)		0.82 (0.74;0.92)	0.95 (0.82; 1.11)
MA17 (2012)	0.68 (0.56;0.83)		0.52 (0.45;0.61)	0.76 (0.60; 0.98)
OS				
HERA (2011)	0.85 (0.70;1.04)	0.53 (0.44;0.65)		0.62 (0.47; 0.82)
BIG 1-98 (2011)	0.87 (0.77;0.99)		0.79 (0.69; 0.90)	0.91 (0.76; 1.09)
MA17 (2012)	0.99 (0.79;1.24)		0.61 (0.52;0.71)	0.62 (0.47; 0.81)

Table 3 – OS results, expressed in HR (95%CI), for the two trials assessing efficacy and safety of treatments, in a metastatic setting, reporting censored or IPCW analysis – results from the most recent publication for each trial are reported. RHRs (95%CI) were calculated in order to compare the results from different type of analyses (ITT as reference analysis).

Study	Type of analysis			RHR
	ITT	Censored	IPCW	
EGF104900 (2012)	0.74 (0.57;0.96)	0.80 (0.56;1.12)		1.08 (0.70; 1.67)
Mouridsen 2003 (2003)	0.88 (p=0.53)	0.71 (nr)		-

nr = not reported

1.3. Review findings

Between January 2000 and June 2015, one out of five RCTs assessing efficacy of innovative biological drugs and hormonal therapy for breast cancer permitted patients to cross over at a certain point during the course of the study. From a clinical standpoint, early discontinuation of RCT due to unequivocal observed benefit, harm or futility are always justified by ethical issues. An early interruption for benefit leads to stopping further recruitment in a potential inferior arm, and patients randomized to the control arm can opt to cross over to receive the experimental treatment. However, from a methodological standpoint, this approach leads to uncertainties surrounding the true magnitude of the actual effect.

The scenario might be different when considering the early or the advanced disease setting. Indeed, it is well accepted that patients with metastatic disease, at the time of disease progression, are given the opportunity to switch to the arm with the more promising therapy. This approach has no effect on the earlier measures of treatment effect such as PFS, which represents the primary outcome in all the found studies. On the other hand, the crossover can definitively preclude the possibility to demonstrate an OS benefit, which is often considered the ultimate test of efficacy.

The situation is even more critical in the adjuvant setting, where SCO is generally offered to patients still free of disease recurrence, thus affecting the clear interpretation of both DFS and OS. The scientific community has to deal with the ethical imperative to offer the best treatment to those patients who decide to enter a clinical trial, and with the need of obtaining the most clean evidence to be applied in the whole population. Indeed, all the interventions which raise uncertainties in data interpretations can delay the full acceptance of a clinically relevant intervention.

All the studies included in the present analysis were always analyzed following the ITT approach, which is the traditional analysis. According to the ITT principle, patients are analysed in their assigned treatment arm regardless of the actual treatment received. Therefore, when a substantial fraction of the patients from the less active treatment cross to the more effective treatment, the net benefit of the latter tends to be reduced. Censored analysis can be used to account for disruptions in treatment allocation. This approach censures patients after crossover, and can be more informative on the real performance of the experimental arm. Only 2/10 adjuvant and 2/10 metastatic studies included in the present analysis reported the censored analysis. The censored analysis was associated with an increased benefit for the experimental arm as compared to the ITT analysis. However, censored analysis can introduce bias itself, in particular when censored patients are more likely or less likely to experience the event than uncensored patients (informative censoring). In both HERA and MA17 trials, patients in the SCO cohort were more likely to be younger and have hormone-receptor-positive disease, as compared to patients remaining in the control arm.

One of the most recent type of analysis, IPCW [8], which accounts for prognostic factors, was rarely used (2/10 adjuvant trials). Similarly to the censored analysis, IPCW analysis led to results which favour the experimental arm in respect to the ITT analysis, which instead tends to dilute the treatment effect. By the way, the adjustment made by the IPCW analysis is valid only if the variables which determine crossover are known and measurable, as pointed out by Rimawi et al. [10]. This is not always possible, leaving the choice to imply this type of analysis doubtful.

The main limitation of the research presented in this chapter is the fact that only articles that appeared within a 15-year period were searched in six selected medical journals, reasoning that these ones published most of the RCTs in breast cancer. It would be helpful to further investigate the phenomenon with a more comprehensive literature research. The

prevalence and impact of SCO in fields other than breast cancer should also be studied. The magnitude and direction of the potential bias introduced by the SCO needs to be clearly evaluated, as well as the impact on the results for different effect sizes when results concern safety and when the reason for switching depends on a combination of prognostic factors.

However, the lack of an appropriate reporting is not a trivial concern. In 2005, Montori et al. published a systematic review of RCTs stopped early for benefit [11], which might lead to SCO, in which they highlighted the lack of adequately reported such an important information as the motivation to stop the trial. More attention should be paid by the authors also in the reporting of the characteristics of patients to whom the crossover is offered.

Considering the increasing frequency of the phenomenon, the Consolidated Standards of Reporting Trials (CONSORT) statement [1] could be modified in order to recommend the specification of that information, fundamental for a better understanding and interpretation of the results of the trial.

This chapter clearly points out that the treatment crossover phenomenon is quite common in breast cancer trials. Different methods may lead to different results and interpretations, but there is still no consensus on the appropriate approach to deal with SCO.

Chapter 2

Methods for dealing with selective cross over

Chapter 1 highlighted the need to find appropriate strategies to deal with treatment crossover.

The present chapter describes the most relevant existing statistical approaches to handle SCO. These methods will be then assessed through simulation, as explained in Chapter 3, and the results of the simulations will be presented in Chapter 4.

2.1.Naïve methods

2.1.1. Intention-to-treat (ITT) analysis

As emerged from the medical literature review presented in Chapter 1, all authors use an ITT analysis. According to this approach, patients are analysed depending on which treatment group they were initially allocated to, and data from all randomised patients is used.

ITT analysis results should always be reported, as this method reflects the design of the study. By the way, if the experimental treatment is actually superior to the control, and a fraction of patients have crossed over from the control to the experimental arm, the ITT analysis will tend to dilute the magnitude of the experimental treatment effect estimate, making it appear more similar to the control effect.

A Cox proportional hazard model is usually fitted in order to estimate the treatment effect.

2.1.2. Censored analysis

A possible approach is to censor patients at the time of crossover; this method is used in two out of ten breast cancer trials where SCO has occurred, as seen in Chapter 1.

Groups may actually no longer be balanced after patients are censored, so this type of analysis may be exposed to selection bias, and this is particularly true if patients' probability of crossing over is related to their underlying prognosis.

2.1.3. Treatment as a time-varying covariate

Although no study from the review in Chapter 1 has reported it, an approach to deal with SCO is considering the treatment as a time-varying covariate.

It is an extension of the Cox proportional hazards model:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta X_i(t))$$

where $\lambda_0(t)$ is the baseline hazard function and $X_i(t)$ takes a value of zero when a patient is on control and 1 when on experimental treatment.

This approach also may be subject to selection bias if SCO is related to prognosis.

2.2. More complex methods

2.2.1. Inverse Probability of Censoring Weighting (IPCW) analysis

This method is rarely used in RCTs assessing the efficacy of biological drugs and hormonal therapies for breast cancer, and in particular only in the adjuvant setting, as described in Chapter 1.

Although in another clinical field, Robins and Finkelstein [8] firstly used the IPCW approach to try to overcome noncompliance issue. It can be thought as a method to improve the censoring strategy described in paragraph 2.2. Instead of simply censoring

patients, the covariates of censored patients are considered in order to try to remove selection bias.

IPCW method creates a scenario of missing follow-up data by censoring the follow-up of each subject at the time of crossover. The weight in the analysis for time periods after crossover is then equal to 0.

For subjects in the control with similar characteristics that do not cross over, IPCW method assigns bigger weights to "re-create" the population that would have been observed without crossover. So, a patient in the control arm who remains in the control arm will be assigned a weight > 1 if other patients with similar characteristics crossed over. Weights are based on factors affecting a patient's decision to cross over.

The method relies on the assumption usually called "no unmeasured confounders" in the censoring process. Conditional on the treatment arm *R* and on the recorded history $\bar{V}(t)$ of the time-dependent covariates $V(t)$, the cause-specific hazard of censoring *C* at time t does not further depend on the possibly unobserved failure time T:

$$\lambda_C\,(t\,|\bar{V}(t), R, T, t < T) = \lambda_C\,(t\mid \bar{V}(t), R, t < T)$$

where

λ_C = cause-specific hazard of censoring *C*

R = treatment arm

T = possibly unobserved failure time

$\bar{V}(t)$ = $\{\, V\,(x);\ 0 \le x < t \,\}$ is the recorded history up to time t

V(x) is a vector of all measured time-dependent factors for failure time recorded at time x.

This assumption specifies that, within a treatment arm, patients censored at time t have the same distribution of failure time as those uncensored at time t with the same recorded history.

Given the "no unmeasured confounders" assumption, the IPCW estimators based on the time-dependent prognostic factors can be constructed as follow.

Time-dependent Cox proportional hazards models are used to estimate the treatment-specific hazards of censoring conditional on time-dependent prognostic factors (reason for censoring can differ between arms).

$$\lambda_C(t \mid \bar{V}(t), R, t < T) = \lambda_{0R}(t) \exp(\alpha_R' \bar{V}(t))$$

IPCW Kaplan-Meier (KM) estimator for failure differs from the ordinary KM estimator for failure by weighting the contribution of a subject at risk at time t by the inverse of an estimate of the conditional probability of having remained uncensored until time t, based on the fit of this model.

By denoting

$\hat{\alpha}_R$	as Cox partial likelihood estimate of α_R in treatment arm R
$X = min(T, C)$	
$Y(u) = I(x \geq u)$	as the "at risk" indicator
$\tau = I(T = x)$	as the failure indicator (1 = failure; 0 = censored)

an estimate of the conditional probability of patient i having remained uncensored until time t is provided by the time-dependent extension of KM product limit estimator of censoring

$$\hat{K}^{V}{}_{i}(t) = \prod_{\{j; X_j < t, \tau_j = 0, R_j = R_i\}} [1 - \hat{\lambda}_{R_i}(X_j) \exp\{\hat{\alpha}'_{R_i} V_i(X_j)\}]$$

where

$$\hat{\lambda}_{R_i}(X_j) = (1 - \tau_j) / [\sum_{i=1}^{n} \exp\{\hat{\alpha}'_{R_i} V_i(X_j)\} Y_i(X_j) I(R_i = R)]$$

is the Cox estimator of the baseline hazard function for censoring λ_{0R} in arm *R*.

The subject-specific weight can be defined

$$\hat{W}_i(t) = \hat{K}^{0}{}_{i}(t) / \hat{K}^{V}{}_{i}(t)$$

Where $\hat{K}^{0}{}_{i}(t)$ is the usual treatment arm specific KM estimator of the probability of being uncensored by the time t in treatment R_i.

So the IPCW KM estimator for failure in treatment arm *r*, $r \in \{0, 1\}$, differs from the ordinary KM estimator for failure only in that the contribution of a subject at risk at any time X_i is weighted by the subject-specific weight $\hat{W}_i(X_i)$.

The IPCW KM estimate of the treatment arm specific marginal probability of remaining alive through time t is

$$\hat{S}_T(t \mid r) = \prod_{\{i; X_i < t\}} \frac{1 - \tau_i W_i(X_i) I(R_i = r)}{\left\{\sum_{k=1}^{n} Y_k(X_i) W_k(X_i) I(R_k = r)\right\}}$$

where

$\tau_i W_i(X_i) I(R_i = r)$ is the estimate of the number of subjects in arm *r* who would have been observed to fail at time X_i in the absence of any censoring and

$\sum_{k=1}^{n} Y_k(X_i) W_k(X_i) I(R_k = r)$ is the estimate of the number of subjects in arm z who would have been at risk at time X_i in the absence of any censoring.

$\hat{S}_T(t \mid r)$ estimates the probability $S_T(t \mid r)$ of surviving without failure until time t in the absence of censoring.

It is possible to compare the marginal survival in the two arms by using the Cox proportional hazards model

$$\lambda_T\,(t \mid R) \;=\; \lambda_0\,(t) e^{\beta R}$$

The IPCW Cox partial likelihood score for β is

$$\mathrm{U}(\beta) = \sum_{\mathrm{i}} \tau_{\mathrm{i}} \hat{W}_i(X_i) \times [R_i - \frac{\sum_{j=1}^{n} Y_j(X_i) \hat{W}_j(Xi) Z_j e^{\beta R_j}}{\left\{ \sum_{j=1}^{n} Y_j(X_i) \hat{W}_j(Xi) e^{\beta R_j} \right\}}]$$

The estimating equation $\mathrm{U}(\beta) = 0$ gives a consistent and asymptotically normal estimator of parameter β.

The major concerns on this method are the "no unmeasured confounders" assumption, which is untestable, and the fact that the IPCW approach cannot work if there are any covariates which ensure (that is, the probability equals 1) treatment crossover will or will not occur.

2.2.2. Loeys and Goetghebeur estimator

An approach to estimate the real treatment efficacy in situations where all patients take their allocated treatment in one arm of the trial and compliance is "all-or-nothing" in the other arm is reported by Loeys and Goetghebeur [7]. "All-or-nothing" means that, if a patient allocated to that arm crosses over to the other one, the crossover is assumed to have

happened at the very beginning, right after the randomisation, and the patient is assumed to have only received the treatment he/she switched onto.

The authors present the method by considering that all patients in the control arm comply fully, and patients in the experimental arm may either comply fully (complier) or not at all (non-complier). Individuals in the control arm are also classified as compliers and noncompliers according to how they would have behaved if they had been randomized to the experimental arm. This thesis considers the opposite case in which all patients in the experimental arm comply fully, and patients in the control arm may either comply fully or not at all.

The proportion of noncompliers, α, is assumed to be the same in both arms due to randomisation; this assumption is often called "exclusion restriction assumption".

The probability of survival to time t is denoted by $S_{n0}(t)$ and $S_{c0}(t)$ for noncompliers and compliers randomized to control, and $S_{n1}(t)$ and $S_{c1}(t)$ for noncompliers and compliers randomized to the experimental arm. For each arm $j = 0, 1$:

$$S_j(t) = \alpha * S_{nj}(t) + (1 - \alpha) * S_{cj}(t)$$

Let us assume that allocation to intervention has no effect on noncompliers and has hazard ratio ψ for compliers:

$$S_{n0}(t) = S_{n1}(t)$$

$$S_{c0}(t) = S_{c1}(t)^{1/\psi} \leftrightarrow S_{c0}(t)^{\psi} = S_{c1}(t)$$

The compliance-adjusted intervention effect estimate, $\hat{\psi}$, is obtained by using Kaplan–Meier estimates of $Sn_0(t)$ and $Sc_0(t)$ to estimate the survivor function in the experimental arm:

$$\widehat{S_1^*}(t|\,\psi) = \hat{\alpha} * \hat{S}_{n0}(t) + (1 - \hat{\alpha}) * \hat{S}_{c0}(t)^{1/\psi}$$

A value of ψ is found at which this quantity matches the observed survival in the experimental arm.

Defining

$$\widehat{\Lambda_1^*}(t|\psi) = -log\widehat{S_1^*}(t|\psi),$$

$$G_1^*(\psi) = \sum_j \left[\widehat{\Lambda_1^*}\left(T_j|\psi\right) - \delta_j\right]$$

where the sum is over all individuals in the experimental group, T_j is the censoring/event time for the *j*th individual, and δ_j is the failure indicator for the *j*th individual.

$G_1^*(\psi)$ can be thought of as the difference between observed and expected events in the experimental arm, based on predictions from the control arm if the hypothesized ψ is correct.

The value of ψ that represents the final estimate of the compliance-adjusted intervention effect is found by solving $G_1^*(\psi) = 0$.

The authors demonstrate that

$$G_1^*(\psi) \sim N\{0, s(\psi)^2\}$$

where $s(\psi)^2 = 2\sum_j \widehat{\Lambda_1^*}\left(T_j|\psi\right)$.

Confidence limits for ψ are found, as described in Kim and White [12], by solving

$$G_1^*(\psi) = \pm z_{crit}\, s(\psi),$$

where z_{crit} is the critical value for the appropriate significance level.

The estimation of the point estimate and confidence limits can be obtained using a loop employing interval bisection. The target value is firstly set as either 0, $-z_{crit}\, s$ or $+z_{crit}\, s$, depending whether the point estimate, lower confidence limit, or upper confidence limit is being estimated. Then, minimum and maximum values of ψ are initialized.

At the start of each run of the loop, ψ is defined as the midpoint of the current minimum and maximum values, and $G_1^*(\psi)$ is calculated. If $G_1^*(\psi)$ is greater than the target value, the minimum is reset to the value of ψ used in this run, or if it is less than the target, the maximum is reset to ψ. The loop is then run again, applying these new minimum and

maximum values, unless the difference between them is less than a user-defined value (e.g. 0.01).

An important limitation of this method is clearly the all-or nothing compliance assumption, as this type of compliance is only likely to occur in very specific situations. By the way it remains interesting to evaluate the method in a simulation. The exclusion restriction assumption, though untestable, is likely to hold in the majority of the cases, due to randomisation.

2.2.3. Rank Preserving Structural Failure Time Model (RPSFTM)

Robins and Tsiatis [6] developed the RPSFTM in order to estimate causal effects in the presence of non-compliance in an RCT. This method identifies the treatment effect by using the randomisation of the trial, observed survival and observed treatment history.

An assumption of this method is that, given two patients *i* and *j*, if *i* failed before *j* when on one treatment, then *i* would also fail before *j* if both patients took the same alternative treatment. That is way this approach is called "rank preserving".

Another important assumption is the so called "common treatment effect": the treatment effect is assumed to be the same for patients crossing over to a treatment as for those initially allocated to receive it.

The observed event time T is related to an underlying event time U that would have been observed in the absence of treatment, through an accelerated life model. The parameter ψ of the model represents the factor by which life is accelerated by treatment and is estimated as the value at which U is balanced between the treatment groups (on a user-specified test). The method splits the observed event time for each patient (T_i) into two:

$$T_i = T_{0i} + T_{1i}$$

where T_{0i} and T_{1i} are the lengths of time that the patient spent on control and on experimental treatment before the event, respectively.

For patients randomised to the experimental treatment, who do not cross over onto the control treatment, T_{0i} is equal to 0.

For patients randomised to the control group who do not switch onto the experimental treatment, T_{1i} is equal to 0, while for patients who cross over both T_{0i} and T_{1i} will be greater than 0.

The observed event time T_i is related to counterfactual treatment-free event time U_i by a causal model

$$U_i = T_{0i} + e^{\psi_0}\, T_{1i}$$

where ψ_0 is the true causal parameter.

Assuming that $U \perp\!\!\!\perp R$, where R_i = 0/1 indicates the randomized treatment arm, for any given value of ψ, the hypothesis $\psi = \psi_0$is tested by computing

$$U_i\,(\psi) = T_{0i} + e^{\psi}\, T_{1i}$$

and calculating $Z(\psi)$ as the test statistic for the hypothesis $U(\psi) \perp\!\!\!\perp R$.

Basically U_i is estimated using the causal model for each value of ψ, and the true value of ψ is that for which U(ψ) is independent of randomised groups.

A log-rank test can be used for testing the hypothesis that the baseline survival curves are identical in the two treatment groups. $Z(\psi)$ is a step function, and the point estimate is the value of ψ for which $Z(\psi)$ crosses 0.

The RPSFTM makes different important assumption:

- if a patient fails before another individual on one treatment arm, he/she will also fail before that other individual on all other treatment regimens;
- the time at which a patient would experience the outcome if never treated is not related to the allocation arm (randomisation assumption);

- the treatment effect does not change in relation to the time in which a patient starts receiving the treatment (common treatment effect assumption).

The randomisation assumption should be reasonable in the context of an RCT. The common treatment effect, instead, could be a concern because it is assumed that there is not a difference in the treatment effect in patients initially randomised to the intervention compared to patients in the control group who cross over.

Chapter 3

Simulation study design

The effort to estimate the real effect of an intervention is of crucial importance, and SCO makes it undoubtedly more difficult.

Chapter 1 has described SCO in the field of therapies for breast cancer, highlighting the non-negligible spread of this phenomenon. The most relevant methods for dealing with SCO were presented in Chapter 2.

In order to assess the bias that the different methods may lead to, the real effect of an intervention under study needs to be known. This is possible through a simulation study.

The attention is focused on the adjuvant/neoadjuvant setting, where the crossover is offered before disease recurrence. So a two-arm RCT similar to one emerged from the medical literature search was simulated.

In the first paragraph of this chapter, the hypothesized scenarios are explained, while in the second one the description of the performance measures is given.

3.1. Scenarios' description

A sample size of 3000 was chosen, with 1500 patients allocated each to receive the experimental treatment or the control.

As mentioned in previous chapters, bias is likely to occur when patients with different underlying prognoses have different probabilities of crossing over between treatment arms. In order to explore this, patients were divided into two groups, those called "with a good prognosis", or "at low risk", and those "with a poor prognosis", or "at high risk".

The probability of being “at low risk” was set at either 30% or 70%.

Survival times were generated from an exponential distribution.

The rate chosen for patients “at low risk” was 0.05, while the one for patients “at high risk” was 0.25. Randomisation guarantees that the proportion of patients “at low risk” and “at high risk” is balanced between treatment arms.

Three different scenarios for the actual treatment effect were hypothesized: the hazard ratio (HR) was chosen to be 0.55 ($\beta = -0.598$), to represent a highly effective treatment, or 0.80 ($\beta = -0.223$), a less effective treatment, or 1 ($\beta = 0$) to represent a treatment with no effect.

All patients were assumed to have entered the trial at the same time, and an administrative censoring at 3 years was considered, to represent the end of follow-up.

The crossover was assumed to be unidirectional, from the control to the treatment arm.

Two assumptions on the crossover probabilities were made: in one case, the probability did not change between the two prognostic groups and was chosen to be 0.50; in the other case, patients “at high risk” were considered more likely to crossover, with a probability of 0.80, while for patients “at low risk” the probability was set to be 0.20.

These probabilities were then used to generate a binary variable representing the presence or the absence of crossover for each patient. If present, crossover was assumed to have occurred after 1 year from randomisation.

The summary of all the simulation scenarios is given in Table 3.1.

For each scenario, 1000 different datasets were generated as described in this paragraph, and the various methods applied to each dataset.

Simulations were made using the R statistical software, and the STATA packages *stcomply* [12] and *strbee* [13].

Table 3.1 – Summary of the simulation scenarios

# Scenario	Treatment effect (ln(HR))	% good prognosis = % at low risk	Crossover probabilities*	
			% good prognosis	% poor prognosis
1	- 0.598	30	50	50
2	- 0.598	30	20	80
3	- 0.598	70	50	50
4	- 0.598	70	20	80
5	-0.223	30	50	50
6	-0.223	30	20	80
7	-0.223	70	50	50
8	-0.223	70	20	80
9	0	30	50	50
10	0	30	20	80
11	0	70	50	50
12	0	70	20	80

* Computed on patients who reach 1 year after randomization

3.2. Performance measures

Criteria for evaluating the performance of the obtained results from the different scenarios and statistical approaches are summarised in this paragraph. Performance measures as described in Burton et al [14] were used in order to compare the simulated results with the true values used to generate the data. They include an assessment of bias, accuracy and coverage.

The bias of each method was calculated as

$$\delta = \bar{\hat{\beta}} - \beta,$$

where β is the true initial treatment effect for the scenario under study, and $\bar{\hat{\beta}} = \sum_{i=1}^{B} \hat{\beta}_i / B$, β_i is the estimate of interest within each of the $i = 1,\ldots, B$ simulations.

The mean square error (MSE) provides a useful measure of the overall accuracy, as it incorporates both measures of bias and variability.

Coverage is defined as the proportion of times the 100(1 - α)% confidence interval (i.e. 95% confidence interval) for a particular method contains the true treatment effect, β. The coverage should be approximately equal to the nominal coverage rate, e.g. 95 per cent of samples for 95% confidence intervals, to appropriately control the type I error rate for testing a null hypothesis of no effect. Over-coverage, where the coverage rates are above 95%, suggests that the results are too conservative: more simulations will not find a significant result when there is a true effect, leading to a loss of statistical power with too many type II errors. Conversely, under-coverage, where the coverage rates are lower than 95%, indicates over-confidence in the estimates: more simulations will incorrectly detect a significant result, which leads to higher than expected type I errors.

In Table 3.2. the considered performance measures and formulas are given.

Table 3.2 – Performance measures

Evaluation criteria	Formula
BIAS	
Bias	$\delta = \bar{\hat{\beta}} - \beta$
Percentage bias	$\left(\frac{\bar{\hat{\beta}} - \beta}{\beta}\right) * 100$
Standardised bias	$\left(\frac{\bar{\hat{\beta}} - \beta}{SE(\hat{\beta})}\right) * 100$
ACCURACY	
Mean square error	$\left(\bar{\hat{\beta}} - \beta\right)^2 + \left(SE(\hat{\beta})\right)^2$
COVERAGE	Proportion of times the 100(1 - α)% confidence interval $\widehat{\beta_i} \pm Z_{1-\frac{\alpha}{2}} SE(\widehat{\beta_i})$ include β, for $i = 1,\ldots, B$

$SE(\hat{\beta}) = \sqrt{\frac{1}{(B-1)} \sum_{i=1}^{B} \left(\widehat{\beta_i} - \bar{\hat{\beta}}\right)^2}$ is the empirical standard error, and it represents an assessment of the uncertainty in the estimate of interest between simulations.

An alternative measure of uncertainty is the average of the estimated within simulation SE for the estimate of interest $\sum_{i=1}^{B} SE(\widehat{\beta_i}) / B$.

The empirical SE should be close to the average of the estimated within simulation SE if the estimates are unbiased, so it may be appropriate to consider both estimates of uncertainty.

Chapter 4

Results of the simulation study

The results of the simulations for the different scenarios are reported in Tables 4.1-4.12.

Scenarios from 1 to 4, reproducing a trial in which the experimental treatment is highly effective ($\beta = -0.598$), are reported in Tables 4.1-4.4.

In the first scenario (Table 4.1), the patients are mostly at high risk (only 30% are at good prognosis) and the crossover does not depend on prognosis, but the probability to switch is the same, i.e. 50%, for patients at low and high risk. It is clear from the simulation that the ITT analysis gives the most biased results. The other methods, both naïve and non-naïve, perform better; the method by Loeys and Goetghebeur gives a low biased estimate, but with a wider confidence interval, while the RPSFTM performs particularly well, with a very low biased and accurate estimate. When adjusted by prognosis, the IPCW method gives an unbiased and very accurate estimate.

The second scenario (Table 4.2) reproduces a trial similar to the one in the first scenario, but considers different probabilities to cross over for the two prognosis groups, with a higher probability for patients with a poor prognosis (20% among patients with a good prognosis, 80% for patients with a poor prognosis). ITT analysis still gives biased results, but the other naïve methods also do not perform well. The Loeys and Goetghebeur and the RPSFTM methods give low biased results. Also in this case, when adjusted by prognosis, the IPCW method gives an estimate which is the most similar to the true value of the effect.

In scenarios 3 and 4 (Tables 4.3 and 4.4), trials similar to the ones of the first and second scenario, respectively, are reproduced, with the difference that the patients are mostly at low risk (70% are at good prognosis). Results are similar to the ones observed for the first two scenarios: when the crossover probability does not depend on prognosis, all the considered methods perform well, except the ITT which underestimates the true effect; when the crossover probability depends on prognosis, all the naïve methods and the unadjusted IPCW method give biased results, while the Loeys and Goetghebeur and the RPSFTM approaches perform well. Again, when adjusted by prognosis, the IPCW method provides an unbiased and accurate estimate.

Scenarios from 5 to 8, reproducing a trial in which the experimental treatment is less effective than in previous scenarios ($\beta = -0.223$), are reported in Tables 4.5-4.8.

In scenario 5 (Table 4.5), patients are mostly at high risk (30% are at good prognosis) and the crossover probability is the same, i.e. 50%, for patients at low and high risk. The ITT is still the method that gives the most biased results, although the bias is lower if compared to the one observed in the first scenario (that is identical to scenario 5, except for the value of the true treatment effect). The other methods, both naïve and non-naïve perform well, with the Loeys and Goetghebeur and the RPSFTM approaches giving wider confidence intervals if compared to the other methods, with the RPSFTM tending to slightly overestimate the treatment effect.

In scenario 6 (Table 4.6), a trial similar to the one in scenario 5 is reproduced, but the probability to cross over is higher for patients with a poor prognosis (20% among patients with a good prognosis, 80% for patients with a poor prognosis). All the naïve methods do not perform well, but in particular the model in which the treatment is considered as a time-dependent covariate gives results with the higher bias. Also the unadjusted IPCW analysis does not perform well. The method by Loeys and Goetghebeur and the RPSFTM

give low biased results, again with the RPSFTM tending to slightly overestimate the treatment effect.

Similar results to the ones observed in scenarios 5 and 6 are obtained in scenarios 7 and 8, respectively (Table 4.7 and 4.8), which reproduce trials similar to the previous two scenarios, with the difference that the patients are mostly at low risk (70% are at good prognosis).

Basically across all these scenarios, it is easy to see that, when adjusted by prognosis, the IPCW method has always a good performance.

Scenarios from 9 to 12, reproducing a trial in which the experimental treatment has no effect ($\beta = 0$), are reported in Tables 4.9-4.12. It was not possible to determine the percentage bias for these scenarios, because of the nullity of the parameter posed at the denominator of the performance measure.

Scenario 9 (Table 4.9) considers a trial in which patients are mostly at high risk (30% are at good prognosis) and the crossover probability is the same, i.e. 50%, for patients at low and high risk. In this case, the ITT is the method which performs better. The other approaches perform well, too, with the RPSFTM and the unadjusted IPCW analysis presenting a slightly higher bias.

The following scenario (Table 4.10) reproduces a trial similar to the one in scenario 9, but with different probabilities to cross over for the two prognosis groups (20% among patients with a good prognosis, 80% for patients with a poor prognosis). ITT analysis still gives low biased results, along with the Loeys and Goetghebeur method and the adjusted IPCW analysis. The censored analysis and the analysis which considers the treatment as a time dependent variable lead to biased results, along with the unadjusted IPCW analysis.

In scenarios 11 and 12 (Tables 4.11 and 4.12), trials similar to the ones in scenarios 9 and 10, respectively, are reproduced, but with patients mostly at low risk (70% are at good

prognosis). Results are similar to the ones observed for the previous two scenarios. When the crossover probability does not depend on prognosis, all the methods perform well, the unadjusted IPCW being the one which leads to a higher bias. When the crossover probability depends on prognosis, the ITT still gives low biased results, along with the Loeys and Goetghebeur; while the other naïve methods and the unadjusted IPCW lead to high biased estimates.

SCO = selective crossover; ITT = Intention to treat; TD-COV = time dependent covariate; RPSFTM = Rank Preserving Structural Failure Time Model; IPCW = inverse probability of censoring weighting
CI = confidence interval; STD = standardised; MSE = mean square error; SE = standard error

Table 4.1. Results from scenario 1 (% at good prognosis = 30; crossover probabilities among patients at: good prognosis = 50, poor prognosis = 50)

True effect		95% CI								
β = -0.598	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.567	-0.571	-0.563	0.031	-5.190	48.890	0.005	0.063	0.065	0.936
ITT	-0.419	-0.423	-0.415	0.179	-29.955	273.810	0.036	0.065	0.066	0.233
CENSORED	-0.571	-0.576	-0.567	0.027	-4.441	38.028	0.006	0.070	0.071	0.939
TD - COV	-0.575	-0.579	-0.571	0.023	-3.855	34.554	0.005	0.067	0.068	0.941
Loeys & Goetghebeur	-0.560	-0.808	-0.327	0.038	-6.282	32.979	0.014	0.114	0.123	0.948
RPSFTM	-0.591	-0.785	-0.384	0.007	-1.120	8.348	0.006	0.080	0.102	0.949
IPCW (unadjusted)	-0.563	-0.568	-0.559	0.035	-5.772	49.578	0.006	0.070	0.071	0.922
IPCW (adjusted by prognosis)	-0.599	-0.604	-0.595	-0.001	0.259	-2.143	0.005	0.072	0.071	0.949

Table 4.2. Results from scenario 2 (% at good prognosis = 30; crossover probabilities among patients at: good prognosis = 20, poor prognosis = 80)

True effect		95% CI								
β = -0.598	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.567	-0.571	-0.563	0.031	-5.190	48.890	0.005	0.063	0.065	0.936
ITT	-0.347	-0.351	-0.343	0.251	-42.020	377.501	0.068	0.067	0.067	0.034
CENSORED	-0.364	-0.368	-0.359	0.234	-39.149	324.441	0.060	0.072	0.076	0.124
TD - COV	-0.306	-0.310	-0.302	0.292	-48.827	426.496	0.090	0.068	0.072	0.011
Loeys & Goetghebeur	-0.582	-0.784	-0.384	0.016	-2.750	17.356	0.009	0.095	0.102	0.960
RPSFTM	-0.593	-0.772	-0.416	0.005	-0.786	7.028	0.004	0.067	0.091	0.951
IPCW (unadjusted)	-0.409	-0.414	-0.405	0.189	-31.557	260.893	0.041	0.072	0.075	0.285
IPCW (adjusted by prognosis)	-0.600	-0.605	-0.595	-0.002	0.346	-2.531	0.007	0.082	0.079	0.955

Table 4.3. Results from scenario 3 (% at good prognosis = 70; crossover probabilities among patients at: good prognosis = 50, poor prognosis = 50)

True effect		95% CI								
β = -0.598	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.552	-0.557	-0.547	0.045	-7.605	57.079	0.008	0.080	0.083	0.923
ITT	-0.406	-0.411	-0.401	0.192	-32.047	230.299	0.044	0.083	0.085	0.383
CENSORED	-0.556	-0.562	-0.551	0.042	-6.949	47.466	0.009	0.088	0.091	0.934
TD - COV	-0.560	-0.565	-0.555	0.038	-6.288	45.588	0.008	0.082	0.087	0.934
Loeys & Goetghebeur	-0.548	-0.827	-0.289	0.050	-8.318	39.788	0.018	0.125	0.137	0.948
RPSFTM	-0.592	-0.831	-0.352	0.006	-0.953	4.586	0.015	0.124	0.122	0.966
IPCW (unadjusted)	-0.546	-0.551	-0.540	0.052	-8.748	59.869	0.010	0.087	0.091	0.916
IPCW (adjusted by prognosis)	-0.600	-0.605	-0.594	-0.002	0.279	-1.778	0.009	0.094	0.091	0.946

Table 4.4. Results from scenario 4 (% at good prognosis = 70; crossover probabilities among patients at: good prognosis = 20, poor prognosis = 80)

True effect		95% CI								
β = -0.598	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.552	-0.557	-0.547	0.045	-7.605	57.079	0.008	0.080	0.083	0.923
ITT	-0.396	-0.401	-0.391	0.202	-33.784	241.910	0.048	0.083	0.085	0.355
CENSORED	-0.348	-0.354	-0.343	0.250	-41.736	278.541	0.070	0.090	0.091	0.205
TD - COV	-0.251	-0.256	-0.245	0.347	-58.093	411.233	0.128	0.084	0.086	0.017
Loeys & Goetghebeur	-0.559	-0.828	-0.305	0.039	-6.443	31.287	0.017	0.123	0.133	0.955
RPSFTM	-0.593	-0.826	-0.358	0.005	-0.886	4.170	0.016	0.127	0.120	0.965
IPCW (unadjusted)	-0.379	-0.385	-0.373	0.219	-36.605	246.412	0.056	0.089	0.090	0.315
IPCW (adjusted by prognosis)	-0.601	-0.607	-0.595	-0.003	0.553	-3.344	0.010	0.099	0.095	0.933

Table 4.5. Results from scenario 5 (% at good prognosis = 30; crossover probabilities among patients at: good prognosis = 50, poor prognosis = 50)

True effect		95% CI								
β = -0.223	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.209	-0.212	-0.205	0.015	-6.554	25.493	0.004	0.057	0.059	0.954
ITT	-0.149	-0.153	-0.146	0.074	-33.060	126.057	0.009	0.059	0.060	0.776
CENSORED	-0.210	-0.214	-0.206	0.013	-5.920	20.817	0.004	0.063	0.066	0.960
TD - COV	-0.211	-0.215	-0.207	0.012	-5.493	20.041	0.004	0.061	0.064	0.962
Loeys & Goetghebeur	-0.206	-0.419	-0.001	0.017	-7.782	17.095	0.011	0.102	0.107	0.957
RPSFTM	-0.238	-0.425	-0.020	-0.015	6.547	-13.900	0.011	0.105	0.103	0.952
IPCW (unadjusted)	-0.202	-0.206	-0.198	0.021	-9.296	32.810	0.004	0.063	0.066	0.944
IPCW (adjusted by prognosis)	-0.224	-0.228	-0.220	-0.001	0.247	-0.819	0.005	0.067	0.066	0.947

Table 4.6. Results from scenario 6 (% at good prognosis = 30; crossover probabilities among patients at: good prognosis = 20, poor prognosis = 80)

True effect		95% CI								
β = -0.223	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.209	-0.212	-0.205	0.015	-6.554	25.493	0.004	0.057	0.059	0.954
ITT	-0.122	-0.126	-0.119	0.100	-45.104	171.548	0.014	0.059	0.060	0.611
CENSORED	-0.016	-0.020	-0.012	0.208	-93.007	313.614	0.047	0.066	0.071	0.150
TD - COV	0.042	0.038	0.046	0.265	-118.908	420.413	0.074	0.063	0.069	0.019
Loeys & Goetghebeur	-0.216	-0.394	-0.041	0.007	-3.007	7.826	0.007	0.086	0.090	0.955
RPSFTM	-0.243	-0.406	-0.059	-0.020	9.081	-21.847	0.009	0.093	0.088	0.954
IPCW (unadjusted)	-0.059	-0.063	-0.055	0.164	-73.589	248.918	0.031	0.066	0.070	0.344
IPCW (adjusted by prognosis)	-0.222	-0.227	-0.217	0.001	-0.547	1.639	0.006	0.075	0.074	0.954

Table 4.7. Results from scenario 7 (% at good prognosis = 70; crossover probabilities among patients at: good prognosis = 50, poor prognosis = 50)

True effect		95% CI								
β = -0.223	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.203	-0.208	-0.199	0.020	-8.946	28.628	0.005	0.070	0.076	0.961
ITT	-0.145	-0.149	-0.141	0.078	-35.043	110.143	0.011	0.070	0.077	0.835
CENSORED	-0.207	-0.212	-0.203	0.016	-7.071	20.655	0.006	0.076	0.084	0.963
TD - COV	-0.209	-0.214	-0.205	0.014	-6.221	18.840	0.006	0.074	0.081	0.964
Loeys & Goetghebeur	-0.204	-0.438	0.018	0.019	-8.700	18.420	0.011	0.105	0.117	0.968
RPSFTM	-0.225	-0.456	0.023	-0.001	0.673	-1.227	0.015	0.122	0.122	0.959
IPCW (unadjusted)	-0.197	-0.202	-0.192	0.026	-11.754	34.458	0.006	0.076	0.084	0.957
IPCW (adjusted by prognosis)	-0.227	-0.232	-0.221	-0.003	1.533	-4.122	0.007	0.083	0.084	0.952

Table 4.8. Results from scenario 8 (% at good prognosis = 70; crossover probabilities among patients at: good prognosis = 20, poor prognosis = 80)

True effect		95% CI								
β = -0.223	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.203	-0.208	-0.199	0.020	-8.946	28.628	0.005	0.070	0.076	0.961
ITT	-0.141	-0.145	-0.137	0.082	-36.809	114.963	0.012	0.071	0.077	0.834
CENSORED	-0.003	-0.008	0.001	0.220	-98.506	282.992	0.054	0.078	0.084	0.235
TD - COV	0.093	0.089	0.098	0.317	-141.879	428.269	0.106	0.074	0.080	0.009
Loeys & Goetghebeur	-0.209	-0.439	0.013	0.014	-6.333	13.546	0.011	0.104	0.115	0.968
RPSFTM	-0.216	-0.449	0.019	0.007	-3.094	6.122	0.013	0.113	0.119	0.948
IPCW (unadjusted)	-0.033	-0.038	-0.029	0.190	-85.020	246.121	0.042	0.077	0084	0.377
IPCW (adjusted by prognosis)	-0.225	-0.231	-0.220	-0.002	1.048	-2.705	0.007	0.086	0.087	0.946

Table 4.9. Results from scenario 9 (% at good prognosis = 30; crossover probabilities among patients at: good prognosis = 50, poor prognosis = 50)

True effect		95% CI								
$\beta = 0$	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.002	-0.005	0.002	-0.002	-	-2.927	0.003	0.055	0.057	0.959
ITT	0.000	-0.004	0.003	0.000	-	-0.638	0.003	0.053	0.057	0.966
CENSORED	-0.001	-0.004	0.003	-0.001	-	-0.902	0.004	0.060	0.063	0.966
TD - COV	-0.001	-0.004	0.003	-0.001	-	-1.050	0.003	0.058	0.062	0.961
Loeys & Goetghebeur	0.001	-0.196	0.193	0.001	-	0.706	0.009	0.097	0.099	0.962
RPSFTM	-0.028	-0.184	0.246	0.028	-	27.297	0.011	0.101	0.110	0.968
IPCW (unadjusted)	0.007	0.003	0.011	0.007	-	11.636	0.004	0.059	0.063	0.963
IPCW (adjusted by prognosis)	-0.001	-0.005	0.003	-0.001	-	-1.266	0.004	0.063	0.063	0.958

Table 4.10. Results from scenario 10 (% at good prognosis = 30; crossover probabilities among patients at: good prognosis = 20, poor prognosis = 80)

True effect		95% CI								
$\beta = 0$	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	-0.002	-0.005	0.002	-0.002	-	-2.927	0.003	0.055	0.057	0.959
ITT	0.001	-0.002	0.004	0.001	-	1.663	0.003	0.055	0.057	0.965
CENSORED	0.189	0.185	0.193	0.189	-	294.866	0.040	0.064	0.068	0.188
TD - COV	0.247	0.244	0.251	0.247	-	405.783	0.065	0.061	0.066	0.020
Loeys & Goetghebeur	-0.001	-0.167	0.165	-0.001	-	-0.900	0.006	0.079	0.085	0.968
RPSFTM	0.016	-0.165	0.192	0.016	-	22.123	0.005	0.072	0.091	0.962
IPCW (unadjusted)	0.147	0.143	0.151	0.147	-	229.179	0.026	0.064	0.068	0.419
IPCW (adjusted by prognosis)	-0.001	-0.005	0.004	-0.001	-	-0.897	0.005	0.072	0.071	0.945

Table 4.11. Results from scenario 11 (% at good prognosis = 70; crossover probabilities among patients at: good prognosis = 50, poor prognosis = 50)

True effect		95% CI								
$\beta = 0$	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	0.001	-0.003	0.005	0.001	-	1.274	0.005	0.068	0.072	0.963
ITT	0.000	-0.004	0.004	0.000	-	-0.203	0.005	0.069	0.072	0.955
CENSORED	0.000	-0.004	0.005	0.000	-	0.525	0.006	0.078	0.081	0.959
TD - COV	0.001	-0.004	0.005	0.001	-	0.908	0.006	0.075	0.078	0.963
Loeys & Goetghebeur	0.003	-0.211	0.206	0.003	-	2.667	0.010	0.099	0.106	0.959
RPSFTM	0.003	-0.243	0.243	0.003	-	1.975	0.019	0.139	0.124	0.958
IPCW (unadjusted)	0.011	0.006	0.015	0.011	-	13.594	0.006	0.077	0.081	0.952
IPCW (adjusted by prognosis)	0.001	-0.004	0.006	0.001	-	1.230	0.007	0.084	0.081	0.942

Table 4.12. Results from scenario 12 (% at good prognosis = 70; crossover probabilities among patients at: good prognosis = 20, poor prognosis = 80)

True effect		95% CI								
$\beta = 0$	$\bar{\hat{\beta}}$	Lower	Upper	Bias	Percentage BIAS	STD Bias	MSE	Empirical SE	Average SE	Coverage
In absence of SCO	0.001	-0.003	0.005	0.001	-	1.274	0.005	0.068	0.072	0.963
ITT	0.001	-0.003	0.005	0.001	-	1.731	0.005	0.069	0.072	0.956
CENSORED	0.202	0.197	0.206	0.202	-	260.421	0.047	0.077	0.081	0.289
TD - COV	0.299	0.294	0.303	0.299	-	403.700	0.095	0.074	0.078	0.021
Loeys & Goetghebeur	-0.001	-0.214	0.205	-0.001	-	-1.133	0.010	0.102	0.107	0.956
RPSFTM	-0.010	-0.240	0.232	-0.010	-	-7.257	0.018	0.133	0.120	0.850
IPCW (unadjusted)	0.172	0.167	0.176	0.172	-	224.612	0.035	0.076	0.081	0.430
IPCW (adjusted by prognosis)	0.003	-0.003	0.008	0.003	-	3.129	0.007	0.086	0.084	0.952

Chapter 5

Discussion, conclusions and hints for future research

1.1. Discussion

When a RCT is designed to compare the difference in (disease-free) survival time between two treatment groups, letting patients switch treatment may critically complicate the statistical analysis and interpretation of results.

As emerged from the literature review in Chapter 1, this phenomenon is quite common among trials assessing the efficacy of biological drugs and hormonal therapies for breast cancer, as over one out of five let patients in the control arm cross over to receive the intervention under study, at a certain point after randomisation. The ITT analysis was always reported; by the way, as already explained, this approach may not be appropriate, tending to underestimate the actual treatment effect, in the hypothesis that the treatment is more effective than the regimen administered in the control. The review highlighted also the fact that, if present, other types of analyses conducted were usually based on naïve methods, i.e. the censored analysis. Only in a very few cases, results from a more complex method, i.e. IPCW, were reported.

Chapter 2 presented these methods along with others that can be helpful to deal with SCO: naïve methods, i.e. treatment considered as a time dependent variable, and more complex methods, i.e. Loeys and Goetghebeur estimator, and the RPSFTM by Robins and Tsiatis.

The results of the simulation study, presented in Chapter 4, confirmed that the ITT analysis, although the most used approach, tends to underestimate the true treatment effect across all the considered scenarios, but it performs quite well in the extreme case that the

treatment under study does not have an impact on survival if compared to control regimen. Other naïve methods – such as censoring switching patients at the time of crossover, or incorporating treatment received as a time-dependent covariate – performed well, with low levels of bias, when the probability of crossing over was assumed to be the same for patients with different prognosis; if patients with a poor prognosis were assumed to have a higher probability to cross over, these methods led to biased results, underestimating the true treatment effect.

As far as more complex methods is concerned, the estimator proposed by Loeys & Goetghebeur did perform quite well across all scenarios, even if it tended to present wider confidence intervals if compared to the other methods. By the way, this approach is based on the "all-or-nothing" assumption, which is unlikely to hold in a context as the one considered.

The RPSFTM approach led to low biased estimates across the various scenarios, but tending to have wider confidence intervals, as for the Loeys & Goetghebeur method. The common treatment effect assumption, considered by this method, which states that the effect does not change in relation to the time in which a patient starts receiving the treatment, is strong and can be difficult to justify in a real context.

The unadjusted IPCW analysis could estimate the true treatment effect with low bias when the probability of crossover was the same for the two prognosis classes, otherwise it did not perform well, in particular with if the treatment has a moderate or no effect. Across all the scenarios, when adjusted by prognosis, the IPCW method had always a good performance; this highlights the importance of knowing the characteristics that may favour the crossover, in order to obtain an unbiased and accurate estimate of the treatment effect. By the way it is not easy, as emerged from the literature review, to understand from the papers reporting trials' results which are the characteristics that may influence the choice of a patient to switch arm.

1.2. Conclusions and hints for future research

The present work aimed to highlight the need for attention being paid on the phenomenon of SCO in RCTs. This thesis, in fact, underlines the concerns related to it, but further research is still necessary.

A more comprehensive medical literature research would be helpful to better estimate the prevalence of SCO in trials evaluating interventions for breast cancer, but it would be important to explore other medical fields, too. A systematic review of all methodological papers dealing with the methods to adjust for SCO and published in the literature may lead to have other methods to consider, not analysed in this thesis.

The simulations performed for this work can be adapted to create scenarios in which there are more than two prognosis categories, with different probabilities to cross over. It could be appropriate to overcome the limitation for which, when present, the crossover occurs after 1 year, letting this time change among patients, and to see the impact this modification can have on the performances of the different methods.

Although simulations are necessary, it would be interesting to apply the described methods to a real dataset from a trial in which SCO has occurred.

Clinical decisions are based on RCTs' results, so appropriate analyses and interpretations are crucial. Systematic reviews and meta-analyses, from which clinical guidelines and recommendations derive, may include trials in which SCO has occurred [3,4]: further research should evaluate the impact of SCO on the summary results of the meta-analysis.

References

1. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Annals of Internal Medicine 2010; 152:726–32.
2. Fundamentals of Clinical Trials, Friedman LM., Furberg CD, DeMets D, Reboussin DM, Granger CB. Springer International Publishing 2015. Ch Participant adherence 297-318.
3. Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, D'Amico R. Trastuzumab containing regimens for early breast cancer. Cochrane Database Syst Rev. 2012 Apr 18;4:CD006243.
4. Balduzzi S, Mantarro S, Guarneri V, Tagliabue L, Pistotti V, Moja L, D'Amico. Trastuzumab containing regimens for metastatic breast cancer. Cochrane Database Syst Rev. 2014 Jun 12;6:CD006242.
5. Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomized controlled trials: a simulation study. BMC Med Res Methodol.2011 11:4.
6. Robins JM and Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics – Theory and Methods 1991 20(8):2609-2631.
7. Loeys T and Goethebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics 2003 59:100-105.
8. Robins JM and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000; 56:779–788.

9. Apolone G, Joppi R , Bertelè V, Garattini S. Ten years of marketing approvals of anticancer drugs in Europe: regulatory policy and guidance documents need to find a balance between different pressures. British Journal of Cancer 2005 93, 504–509. doi:10.1038/sj.bjc.6602750.
10. Rimawi M, Hilsenbeck SG. Making sense of clinical trial data: is inverse probability of censoring weighted analysis the answer to crossover bias? Journal of Clinical Oncology 2012; 30(4).
11. Montori VM, Devereaux PJ, Adhikari NKJ, Burns KEA, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized Trials Stopped Early for Benefit - A Systematic Review. The Journal of the Medical Association 2005; 294(17).
12. Kim LG and White IR. Compliance-adjusted intervention effects in survival data. The Stata Journal 2004, 4(3):257-264.
13. White IR and Walker S. strbee: Randomization-based efficacy estimator. The Stata Journal 2002, 2(2):140-150.
14. Burton A, Altman DG, Royston P, Holder RL: The design of simulation studies in medical statistics. Statistics in Medicine 2006, 25:4279-4292.

Studies from the medical literature review of Chapter 1 in which SCO has occurred:

IMPACT

Smith IE, Dowsett M, Ebbs SR, Dixon JM et al. Neoadjuvant Treatment of Postmenopausal Breast Cancer With Anastrozole, Tamoxifen, or Both in Combination:

The Immediate Preoperative Anastrozole, Tamoxifen, or Combined With Tamoxifen (IMPACT) Multicenter Double-Blind Randomized Trial. Journal of Clinical Oncology 2005; 23(22):5108-5116.

HERA

Dowsett M, Procter M, McCaskill-Stevens W, De Azambuja E, Dafni U, Rueschoff J, et al. Disease-free survival according to degree of HER2 amplification for patients treated with adjuvant chemotherapy with or without 1 year of trastuzumab: the HERA trial. Journal of Clinical Oncology 2009; 27:2962–9.

Gianni L, Dafni U, Gelber RD, Azambuja E, Muehlbauer S, Goldhirsch A, Untch M, Smith I, Baselga J, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. Lancet Oncology 2011;12(3):236-44.

Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. New England Journal of Medicine 2005; 353(16):1659–72.

Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M, et al. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet 2007; 369 (9555): 29–36.

Suter TM, Procter M, Van Veldhuisen DJ, Muscholl M, Bergh J, Carlomagno C, et al. Trastuzumab-associated cardiac adverse effects in the Herceptin adjuvant trial. Journal of Clinical Oncology 2007; 25 (25):3859–65.

Untch M, Gelber RD, Jackisch C, Procter M, Baselga J, Bell R, et al. Estimating the magnitude of trastuzumab effects within patient subgroups in the HERA trial. Annals of Oncology 2008; 19(6):1090–6.

MA17

Muss HB, Tu D, Ingle JN, Martino S, Robert NJ, Pater JL, Whelan TJ et al. Efficacy, Toxicity, and Quality of Life in Older Women With Early-Stage Breast Cancer Treated With Letrozole or Placebo After 5 Years of Tamoxifen: NCIC CTG Intergroup Trial MA.17. Journal of Clinical Oncology 2008; 26(12):1956-1964.

Goss PE, Ingle JN, Pater JL, Martino S, Robert NJ, Muss HB, Piccart MJ, Castiglione M et al. Late Extended Adjuvant Treatment With Letrozole Improves Outcome in Women With Early-Stage Breast Cancer Who Complete 5 Years of Tamoxifen. Journal of Clinical Oncology 2008; 26(12):1948-195

Ingle JN, Tu D, Pater JL, Muss HB et al. Intent-to-treat analysis of the placebo-controlled trial of letrozole for extended adjuvant therapy in early breast cancer: NCIC CTG MA.17. Annals of Oncology 2008; 19:877-882.

Goss PE, Ingle JN, Martino S, Robert NJ, Muss HB et al. Efficacy of Letrozole Extended Adjuvant Therapy According to Estrogen Receptor and Progesterone Receptor Status of the Primary Tumor: National Cancer Institute of Canada Clinical Trials Group MA.17. Journal of Clinical Oncology 2007; 25(15): 2006-2011.

Goss PE, Ingle JN, Martino S, Robert NJ et al. A Randomized Trial of Letrozole in Postmenopausal Women after Five Years of Tamoxifen Therapy for Early-Stage Breast Cancer. New England Journal of Medicine 2003; 349(19):1793-1802.

Jin H, Tu D, Zhao N, Shepherd LE, and Goss PE. Longer-Term Outcomes of Letrozole Versus Placebo After 5 Years of Tamoxifen in the NCIC CTG MA.17 Trial: Analyses Adjusting for Treatment Crossover. 2012; 30(7):718-721.

NSABP-B-33

Mamounas EP, Jeong JH, Wickerham DL, Smith RE et al. Benefit From Exemestane As Extended Adjuvant Therapy After 5 Years of Adjuvant Tamoxifen: Intention-to-Treat Analysis of the National Surgical Adjuvant Breast and Bowel Project B-33 Trial. Journal of Clinical Oncology 2008; 26(12): 1965-1971.

BIG 1-98

The Breast International Group (BIG) 1-98 Collaborative Group. A Comparison of Letrozole and Tamoxifen in Postmenopausal Women with Early Breast Cancer. New England Journal of Medicine 2005; 353(26):2747-2757.

Coates AS, Keshaviah A, Thürlimann B, Mouridsen H et al. Five Years of Letrozole Compared With Tamoxifen As Initial Adjuvant Therapy for Postmenopausal Women With Endocrine-Responsive Early Breast Cancer: Update of Study BIG 1-98. Journal of Clinical Oncology 2007; 25(5):486-492.

The BIG 1-98 Collaborative Group. Letrozole Therapy Alone or in Sequence with Tamoxifen in Women with Breast Cancer. New England Journal of Medicine 2009; 361:766-776.

Colleoni M, Giobbie-Hurder A, Regan MM et al. Analyses Adjusting for Selective Crossover Show Improved Overall Survival With Adjuvant Letrozole Compared With Tamoxifen in the BIG 1-98 Study. Journal of Clinical Oncology 2011; 29(9):1117-1124.

Regan MM, Neven P, Giobbie-Hurder A et al. Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: the BIG 1-98 randomised clinical trial at 8·1 years median follow-up. Lancet Oncology 2011;12:1101-1108.

NOAH

Gianni L, Eiermann W, Semiglazov V, Manikhas A, Lluch A, Tjulandin S, et al. Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER2-negative cohort. Lancet 2010; 375:377–84.

BCIRG-006

Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, et al. Adjuvant Trastuzumab in HER2-Positive Breast Cancer. New England Journal of Medicine 2011; 365:1273-1283.

TEAM

Van de Velde CJH, Rea D, Seynaeve C, Putter H et al. Adjuvant tamoxifen and exemestane in early breast cancer (TEAM): a randomised phase 3 trial. Lancet 2011; 377:321-331.

B31+N9831

Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. The New England journal of medicine 2005;353(16):1673-84.

Perez EA, Romond EH, Suman VJ, Jeong JH, Sledge G, Geyer CE Jr, Martino S, Rastogi P, Gralow J, Swain SM, Winer EP, Colon-Otero G, Davidson NE, Mamounas E, Zujewski JA, Wolmark N. Trastuzumab Plus Adjuvant Chemotherapy for Human Epidermal Growth Factor Receptor 2–Positive Breast Cancer: Planned Joint Analysis of Overall Survival From NSABP B-31 and NCCTG N9831. Journal of Clinical Oncology 2014; 32(33): 3744-3752.

Slamon 2001

Slamon DJ, Leyland-Jones B, Shak S et al. Use of chemotherapy plum monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. New England Journal of Medicine 2001; 344(11):783-792.

Mouridsen 2003

Lipton A, Ali SM, Leitzel K, Demers L et al. Serum HER-2/neu and Response to the Aromatase Inhibitor Letrozole Versus Tamoxifen. Journal of Clinical Oncology 2003; 21(10): 1967-1972.

Mouridsen H, Gershanovich M, Sun Y et al. Phase III Study of Letrozole Versus Tamoxifen as First-Line Therapy of Advanced Breast Cancer in Postmenopausal Women:

Analysis of Survival and Update of Efficacy From the International Letrozole Breast Cancer Group. Journal of Clinical Oncology 2003; 21(11): 2101-2109

TANDEM

Kaufman B, Mackey JR, Clemens MR et al. Trastuzumab Plus Anastrozole Versus Anastrozole Alone for the Treatment of Postmenopausal Women With Human Epidermal Growth Factor Receptor 2–Positive, Hormone Receptor–Positive Metastatic Breast Cancer: Results From the Randomized Phase III TAnDEM Study. Journal of Clinical Oncology 2009; 27(33): 5529-5537.

AVADO

Miles DW, Chan A, Dirix LY, Corte´s J, Pivot X et al. Phase III Study of Bevacizumab Plus Docetaxel Compared With Placebo Plus Docetaxel for the First-Line Treatment of Human Epidermal Growth Factor Receptor 2–Negative Metastatic Breast Cancer. Journal of Clinical Oncology 2010; 28(20): 3239-3247.

EGF104900

Blackwell KL, Burstein HJ, Storniolo AM, Rugo H et al. Randomized Study of Lapatinib Alone or in Combination With Trastuzumab in Women With ErbB2-Positive, Trastuzumab-Refractory Metastatic Breast Cancer. Journal of Clinical Oncology 2010; 28(7): 1124-1130.

Blackwell KL, Burstein HJ, Storniolo AM, Rugo H et al. Overall Survival Benefit With Lapatinib in Combination With Trastuzumab for Patients With Human Epidermal Growth Factor Receptor 2–Positive Metastatic Breast Cancer: Final Results From the EGF104900 Study. Journal of Clinical Oncology 2012; 30(21): 2585-2592.

RIBBON-1

Robert NJ, Diéras V, Glaspy J et al. RIBBON-1: Randomized, Double-Blind, Placebo-Controlled, Phase III Trial of Chemotherapy With or Without Bevacizumab for First-Line

Treatment of Human Epidermal Growth Factor Receptor 2–Negative, Locally Recurrent or Metastatic Breast Cancer. Journal of Clinical Oncology 2011; 29(10): 1252-1260.

NCT00435409

Crown JP, Dieras V, Staroslawska E, Yardley DA, Bachelot T, Davidson N, Wildiers H, Fasching PA, Capitain O, Ramos M, Greil R, Cognetti F, Fountzilas G, Blasinska-Morawiec M, Liedtke C, Kreienberg R, Miller WH Jr, Tassell V, Huang X, Paolini J, Kern KA, Romieu G. Phase III Trial of Sunitinib in Combination With Capecitabine Versus Capecitabine Monotherapy for the Treatment of Patients With Pretreated Metastatic Breast Cancer. Journal of Clinical Oncology 2013; 31(23): 2870-2878.

NCT00938652

O'Shaughnessy J, Schwartzberg L, Danso MA, Miller KD, Rugo HS, Neubauer M, Robert N, Hellerstedt B, Saleh M, Richards P, Specht JM, Yardley DA, Carlson RW, Finn RS, Charpentier E, Garcia-Ribas I, Winer EP. Phase III Study of Iniparib Plus Gemcitabine and Carboplatin Versus Gemcitabine and Carboplatin in Patients With Metastatic Triple-Negative Breast Cancer. Journal of Clinical Oncology 2014; 32(34): 3840-3847.

NCT00075764

Mehta RS, Barlow WE, Albain KS, Vandenberg TA, Dakhil SR, Tirumali NR, Lew DL, Hayes DF, Gralow JR, Livingston RB, Hortobagyi GN. Combination Anastrozole and Fulvestrant in Metastatic Breast Cancer. New England Journal of Medicine 2012; 367: 435-44.

CONFIRM

Di Leo A, Jerusalem G, Petruzelka L, Torres R, Bondarenko IN, Khasanov R, Verhoeven D, Pedrini JL, Smirnova I, Lichinitser MR, Pendergrass KP, Malorni L, Garnett S, Rukazenkov Y, Martin M. Final Overall Survival: Fulvestrant 500 mg vs 250 mg in the Randomized CONFIRM Trial. Journal of the National Cancer Institute 2014; 106(1).