

BMJ Open Treatment crossovers in time-to-event non-inferiority randomised trials of radiotherapy in patients with breast cancer

Sameer Parpia,¹ Jim A Julian,¹ Lehana Thabane,^{2,3} Chushu Gu,¹ Timothy J Whelan,^{1,4} Mark N Levine^{1,4}

To cite: Parpia S, Julian JA, Thabane L, *et al.* Treatment crossovers in time-to-event non-inferiority randomised trials of radiotherapy in patients with breast cancer. *BMJ Open* 2014;**4**:e006531. doi:10.1136/bmjopen-2014-006531

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-006531>).

Received 3 September 2014
Revised 1 October 2014
Accepted 3 October 2014



CrossMark

¹Ontario Clinical Oncology Group, Department of Oncology, McMaster University, Hamilton, Ontario, Canada

²Centre of Evaluation of Medicines, St Joseph's Hospital, Hamilton, Ontario, Canada

³Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

⁴Juravinski Cancer Centre, Hamilton, Ontario, Canada

Correspondence to
Dr Sameer Parpia;
parpia@mcmaster.ca

ABSTRACT

Background: In non-inferiority trials of radiotherapy in patients with early stage breast cancer, it is inevitable that some patients will cross over from the experimental arm to the standard arm prior to initiation of any treatment due to complexities in treatment planning or subject preference. Although the intention-to-treat (ITT) analysis is the preferred approach for superiority trials, its role in non-inferiority trials is still under debate. This has led to the use of alternative approaches such as the per-protocol (PP) analysis or the as-treated (AT) analysis, despite the inherent biases of such approaches.

Methods: Using simulations, we investigate the effect of 2%, 5% and 10% random and non-random crossovers prior to radiotherapy initiation on the ITT, PP, AT and the combination of ITT and PP analyses with respect to type I error in trials with time-to-event outcomes. We also evaluate bias and SE of the estimates from the ITT, PP and AT approaches.

Results: The AT approach had the best performance in terms of type I error, but was anticonservative as non-random crossover increased. The ITT and PP approaches were anticonservative under all percentages of random and non-random crossover. Similarly, lowest bias was seen with the AT approach; however, bias increased as the percentage of non-random crossover increased. The ITT and PP had poor performance in terms of bias as crossovers increased.

Conclusions: If minimal crossovers were to occur, we have shown that the AT approach has the lowest type I error rates and smallest opportunity for bias. Results of trials with a high number of crossovers should be interpreted with caution, especially when crossover is non-random. Attempts to prevent crossovers should be maximised.

INTRODUCTION

The non-inferiority randomised trial design is frequently used to compare novel experimental breast radiation regimens with standard breast irradiation for the prevention of

Strengths and limitations of this study

- Using simulations, we investigated the effect of crossovers on various analysis populations in non-inferiority trials with time-to-event outcomes.
- We considered random and non-random crossovers.
- Simulations were limited to clinically plausible non-inferiority trials of radiotherapy in women with early stage breast cancer.
- Adjustment of baseline covariates was not considered in the analysis.

local recurrence in patients with breast cancer who have undergone breast conserving surgery. For example, hypofractionated radiotherapy that delivers a high dose of radiation per fraction and therefore requires a shorter duration of treatment resulting in greater convenience for the patient is often compared with standard radiotherapy using a non-inferiority design.^{1–4} The challenges in the design, conduct and analysis of such trials have been discussed by several authors.^{5–9} These include the determination of the non-inferiority margin and issues related to assay sensitivity, biocreeep and the choice of the analysis population.^{5 7 10–14}

Typically, in breast cancer radiotherapy trials, prior to beginning treatment, the patient undergoes a planning process to establish the treatment fields to target the tumour and avoid radiating normal tissue. Such planning generally occurs after randomisation. Sometimes, planning may reveal that it is not possible to deliver the experimental regimen, and therefore the patient is treated with standard therapy. In some cases, after being randomised to experimental radiotherapy, the patient decides to be treated with standard radiotherapy instead. In a trial of

1234 women comparing hypofractionated radiotherapy with standard radiotherapy for the prevention of local recurrence, the crossover percentage was 1.2%.³ Generally, in trials evaluating new experimental radiotherapy techniques (that are not currently available as part of usual care), patients are not permitted to cross over from standard therapy to experimental therapy.¹⁵

In randomised superiority trials, it is well established that the analysis should be performed based on the *intention-to-treat* (ITT) principle—which states that patients are analysed according to the group they were randomised to regardless of the treatment they received. An ITT analysis tends to produce diluted treatment effect estimates and therefore is considered a conservative approach in analysis of superiority trials, but is anticonservative in demonstrating non-inferiority.¹⁶ This has led to the use of a *per-protocol* (PP) analysis where the analysis set comprises only patients who fully comply with their assigned treatment,¹¹ or an *as-treated* (AT) analysis which groups patients according to the treatment they actually received,¹² despite the inherent bias of such analyses.^{17 18}

Several researchers have investigated the effect of non-compliance on the ITT and PP analyses in non-inferiority trials with binary or continuous outcomes.^{11 19–22} This research has focused mainly on issues such as dropouts, missing data and treatment discontinuations. Literature on the effect of crossover bias in non-inferiority trials is limited. Sheng and Kim²³ showed that both the ITT and PP can be biased in trials with binary outcomes. Matsuyama¹⁶ studied the effect of crossovers to the other treatment after initiating the assigned treatment in the time-to-event situation, and suggested that the PP analysis should not be used. Similarly, others have studied the effect of switching treatments mid-trial (ie, after receiving at least some of their original allocated treatment) in the context of superiority trials.^{24–29} However, the effect of crossovers prior to treatment initiation such as in radiotherapy trials is unknown.

Some authors have suggested performing the analysis using the ITT and PP populations, and that non-inferiority should only be concluded if the null hypothesis is rejected using both analyses.^{5 22 30} The Committee of Proprietary Medicinal Products Points-to-Consider states that ‘... similar conclusions from both the ITT and PP are required in a non-inferiority trial’.³¹ However, this has not been investigated comprehensively for trials with time-to-event outcomes with crossovers.

In this paper, we focus on non-inferiority trials of radiotherapy with a time-to-event outcome in patients with early stage breast cancer. Using simulation, we investigate the effect of patient crossover from the experimental to standard therapy (prior to initiation assigned therapy) on the analysis using the ITT, PP and AT and the combination of ITT and PP analysis sets with respect to type I error, bias and SEs.

METHODS

Intention-to-treat

The ITT approach uses all patients who were randomised to the study and analyses them according to their assigned treatment group regardless of whether they actually received or complied with the treatment. This is advantageous because it preserves the integrity of randomisation and therefore ensures that, on average, the treatment groups are comparable. In addition, since it includes all patients, it helps prevent bias which may occur when excluding patients.¹² Furthermore, the ITT approach will produce results that are likely to be observed in the clinic.¹⁰ The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use also states: “however, in an equivalence or non-inferiority trial use of the full analysis set (ITT) is generally not conservative and its role should be considered very carefully”.¹⁰

Per-protocol

The PP approach excludes patients who have not completed their assigned treatment based on the study protocol. This approach aims to measure the ‘pure’ treatment effect by including only patients who comply with the protocol and excludes those who have crossed over.³² The use of the PP analysis in non-inferiority trials has increased because of the apparent anticonservative nature of the ITT in this setting.¹¹

As-treated

This approach analyses patients according to the treatment they actually received, and not the treatment assigned. Therefore, crossover patients are included in the analysis and are grouped with the treatment arm to which they crossed over.

Combination of ITT and PP analyses

This combined approach requires that both the ITT and PP analyses are performed and that the null hypothesis is rejected (ie, declaring non-inferiority) only if both analyses reject the null hypothesis.

Hypotheses and assessing non-inferiority

Following D’Agostino *et al.*,⁵ let λ_E and λ_S represent the constant hazard rates for the experimental and standard therapy, respectively, and $\theta = \lambda_E/\lambda_S$ be the HR. Let M be the non-inferiority margin, that is, the maximum tolerable amount by which λ_E can be worse than λ_S ($M > 1$). Then the null and alternative hypotheses are:

$$H_0 : \theta = \lambda_E/\lambda_S \geq M$$

$$H_1 : \theta = \lambda_E/\lambda_S < M$$

To test the hypothesis of non-inferiority, we compute the $100(1 - 2\alpha)\%$ CI for $\hat{\theta}$. If the upper bound of the CI is less than M , then we can conclude that the experimental

therapy is no worse than the standard therapy by a maximum of M , and hence is non-inferior to the standard therapy at a significance level of α .

Simulation

The 5-year local recurrence rate following radiotherapy in women with early stage breast cancer who have undergone breast conserving surgery was approximately 5%, or $\lambda_S = -\log(0.95)/5$.³ In recent trials of radiotherapy in women with breast cancer, non-inferiority margins of 1.5 and 1.7 have been used.^{2 3} Also, it is recommended that a one-sided $\alpha=0.025$ be used for non-inferiority studies.^{10 31} We considered two non-inferiority trials with a 5-year local recurrence rate of 5%, a one-sided $\alpha=0.025$, 90% power, 4 years of accrual and an additional 3 years of follow-up. On the basis of these parameters, we calculated total sample sizes of 5134 and 3004 for trials with non-inferiority margins of 1.5 and 1.7, respectively, assuming a 1:1 allocation ratio.³³

Patients undergoing breast conserving therapy have a varying risk of recurrence, and therefore we considered two risk subgroups (high, low) and assumed the HR of high versus low risk to be 1.4, similar to that of a grade III versus grade I/II tumours.³⁴ In addition, we assumed that 20% of the patients were high risk and 80% were low risk.

To evaluate the type I error rates, data were simulated under the null hypothesis where E is inferior to S with a true HR of θ being equal to the non-inferiority margin ($\theta=M$). For each trial, we simulated data for two randomly generated treatment groups of equal size. For each patient, the baseline covariate of risk (high vs low) was generated using the binomial distribution with probability 0.2. Local recurrence-free survival times were generated using the formula³⁵:

$$T = -\frac{\log(U)}{\lambda_S \exp(\beta'x)}$$

where U is a random variable following a uniform distribution on the interval from 0 to 1, $\lambda_S = -\log(0.95)/5$ is the baseline hazard function, β is the vector of the regression coefficients and x is the vector of covariates. The regression coefficient for treatment (E vs S) was $\log(M)$ and $\log(1.4)$ for the risk (high vs low) covariate. Enrolment times were generated using the uniform distribution from 0 to 4 corresponding to 4 years of accrual. Patients were censored at the end of the trial if they remained event-free at that time.

We evaluated scenarios where the percentage of patients who crossed over from experimental to standard therapy was 2%, 5% and 10%. For each of these, we considered the following situations: (1) the crossover of patients was random, and (b) the high-risk patients were more likely to cross over (ie, non-random). This was simulated assuming that 50% of the crossover patients were high-risk patients.

For each approach, computation of the $100(1-2\alpha)$ CI of the estimated HR, $\hat{\theta}$, was performed using the Cox proportional hazards (Cox-PH) model with $\alpha=0.025$. We carried out 10 000 replications for each trial giving an SE of the estimate of type I error of 0.15%. The one-sided type I error was calculated as the proportion of trials that had the null hypothesis of inferiority rejected, that is, the proportion of trials in which the upper CI was less than M . Bias for each of the ITT, PP and AT analyses was calculated as the percentage difference between $\hat{\theta}$ and θ , and averaged over the number of simulations. The SE was also averaged over the total number of simulations. All analyses were performed in R 3.0 (<http://www.r-project.org>).

RESULTS

Impact on type I error

The results of the type I errors for the four approaches are shown in [table 1](#), and graphically in [figure 1](#). The results showed that the AT approach had the best performance with type I errors closer to nominal for 2% and 5% crossovers, 0.028 and 0.027, respectively ([figure 1A](#)). We observed that the combined ITT+PP approach performed better than the separate ITT and PP analyses, and that the ITT and PP approaches had comparable overall type I errors. However, these approaches had type I errors that were greater than the nominal value, regardless of the crossover percentage. In general, overall type I errors increased as the crossover percentage increased for all approaches.

For scenarios with a random crossover ([figure 1B](#)), the AT approach had nominal or close to nominal type I errors for all crossover percentages. The ITT+PP approach had close to nominal type I error when the random crossover was 2%, but performed poorly as the random crossover percentage increased. The individual ITT and PP approaches had greater than nominal type I errors under all scenarios of random crossover.

Under non-random crossover scenarios ([figure 1C](#)), all approaches performed poorly irrespective of the crossover percentage with the exception of the AT analyses when the true HR was 1.5 and the crossover was 2% ([table 1](#)). In general, the PP approach had the worst performance under scenarios of non-random crossover.

Impact of bias

The AT approach also had the best performance in terms of overall bias of the HR estimates, whereas the ITT and PP approaches perform similarly ([figure 2A](#)). As the percentage of crossover patients increased, the overall per cent bias also increased for all approaches. When the crossovers were random ([figure 2B](#)), the AT approach had comparable bias across all levels of crossover percentages, whereas the ITT and PP approaches had greater bias as the crossover percentage increased. The ITT and PP approaches behaved similarly under the random and non-random scenarios, but their bias was

Table 1 Results of type I error, bias and SE for each approach by non-inferiority margin, crossover percentage and crossover type

True HR (NI margin)	Crossover (%)	Crossover type	Type I error			Bias (%)			SE			
			ITT	PP	AT	ITT+PP	ITT	PP	AT	ITT	PP	AT
1.5	2	Random	0.0317	0.0312	0.0283	0.0288	1.1868	1.1935	0.5350	0.1106	0.1111	0.1103
		Non-random	0.0330	0.0354	0.0313	0.0316	1.1835	1.4029	0.8987	0.1107	0.1112	0.1103
	5	Random	0.0366	0.0367	0.0251	0.0319	2.3329	2.3343	0.6509	0.1108	0.1120	0.1101
		Non-random	0.0348	0.0367	0.0291	0.0313	2.1389	2.7165	1.4218	0.1109	0.1122	0.1100
	10	Random	0.0507	0.0464	0.0227	0.0394	4.1908	4.2561	0.6910	0.1112	0.1137	0.1100
		Non-random	0.0547	0.0653	0.0436	0.0483	4.5385	5.6278	3.0124	0.1112	0.1140	0.1099
1.7	2	Random	0.0302	0.0298	0.0262	0.0273	1.8692	1.8839	0.9979	0.1416	0.1421	0.1410
		Non-random	0.0306	0.0318	0.0270	0.0290	1.6345	1.8810	1.1791	0.1416	0.1422	0.1410
	5	Random	0.0345	0.0341	0.0236	0.0300	2.7744	2.8141	0.6719	0.1420	0.1434	0.1407
		Non-random	0.0373	0.0400	0.0311	0.0334	3.2014	3.7458	2.0289	0.1420	0.1436	0.1407
	10	Random	0.0474	0.0465	0.0246	0.0381	5.3500	5.4620	1.0507	0.1424	0.1454	0.1402
		Non-random	0.0581	0.0643	0.0386	0.0512	5.8370	6.8523	3.2413	0.1425	0.1458	0.1402

AT, as-treated; ITT+PP, intention-to-treat and per-protocol combination; PP, per-protocol; TT, intention-to-treat.

larger under the non-random crossover scenario with the PP approach having a larger bias compared with the ITT (figure 2C). Similar to the ITT and PP approaches, the AT approach also showed increased bias as non-random crossover increased, albeit with smaller bias. For each approach, bias was greater when the true HR was 1.7 compared with 1.5.

Impact on SE

The three approaches had comparable overall SEs across all scenarios (table 1). However, a slight trend was observed where the AT approach had the smallest SEs, followed by the ITT, and then the PP approach. Furthermore, we observed that for each approach, the SEs were comparable under scenarios of random and non-random crossovers. SEs were greater for the trials where the true HR was 1.7 compared with 1.5.

DISCUSSION

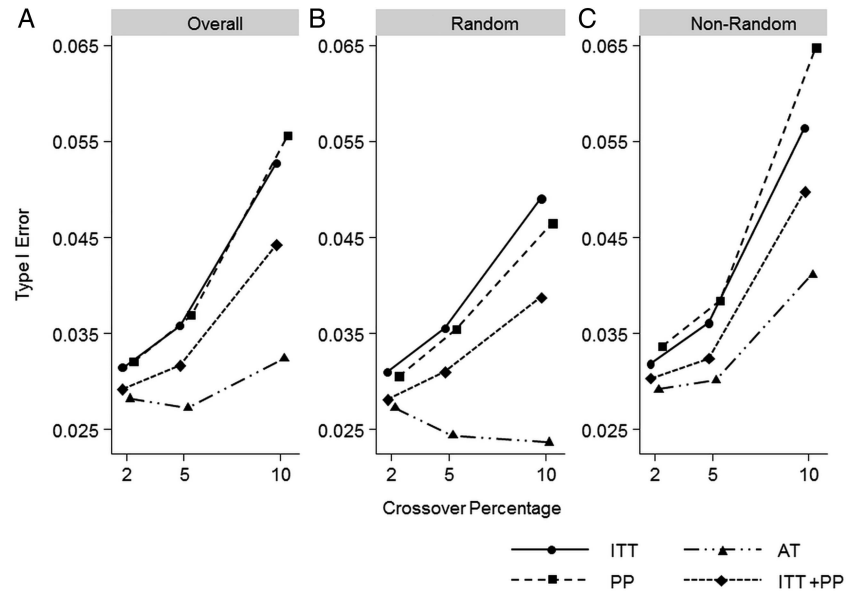
In randomised non-inferiority trials of radiotherapy regimens in women with early stage breast cancer, it is inevitable that some patients will cross over from the experimental arm to the standard arm prior to treatment initiation due to complications in experimental radiotherapy planning or patient or physician preference. In such situations, the ideal population for the final analysis is unclear. We evaluated the performance of the ITT, PP, AT and combined ITT+PP approach under various crossover scenarios.

The AT approach had the best performance under all scenarios in terms of type I error rate. However, it can only be recommended for situations where the crossover is random. Patients who crossed over had their hazard of outcome determined by that of the standard group, and were analysed accordingly by the AT approach. Considering this and the fact that the crossover was random, it is not surprising that the AT approach had near nominal type I error rates under these situations.

Moreover, the combined ITT+PP approach performed better than the ITT and PP approaches separately. This is due to the fact that the ITT+PP approach requires both analyses to reject the null hypothesis prior to non-inferiority being concluded, hence adding an extra level of testing compared to individual ITT and PP approaches, and therefore making it *harder* to conclude non-inferiority. Interestingly, neither the ITT nor the PP approach can be recommended under simulated scenarios, adding to the literature that both approaches could provide increased erroneous results.^{16 23}

We also observed that the AT approach had the lowest bias of the HR estimate across all crossover percentages. Moreover, the biases of the ITT and PP approaches were comparable across all scenarios. For all three approaches, the bias is in the negative direction, and generally increases as the crossover percentage increases, except for the AT approach under the random crossover scenarios where it is not affected by the percentage

Figure 1 Type I error rates for the ITT, PP, AT and combined ITT+PP approaches by crossover type and percentage. (A) Overall; (B); Random crossover; (C) Non-random crossover. ITT, intention-to-treat; PP, per-protocol; AT, as-treated; ITT+PP, intention-to-treat and per-protocol combination.



crossover. Reasons for this observation are similar to that of its performance in terms of type I error under the same scenarios.

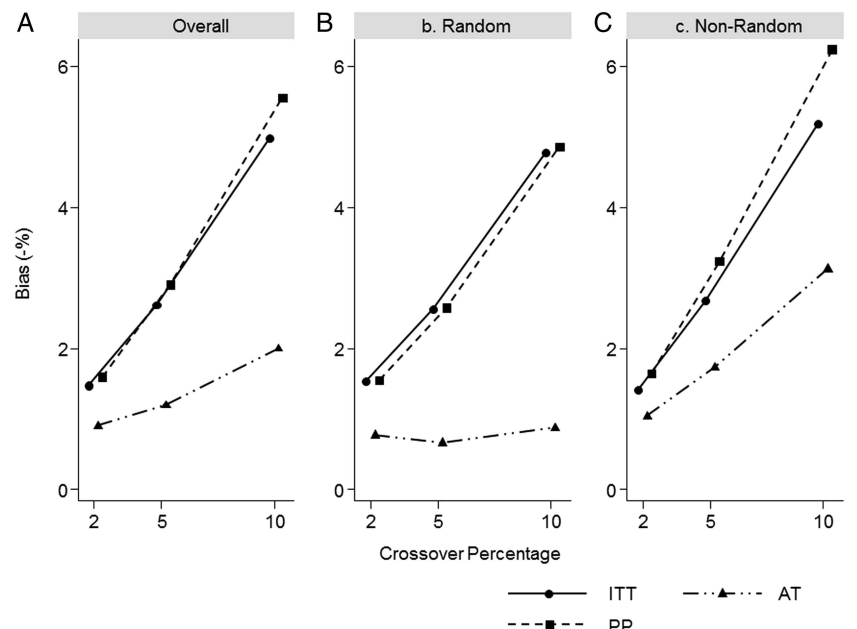
The biases for all methods are larger in scenarios where the true HR is larger because this reflects a greater hazard of events in the experimental arm. Therefore, the crossover patients have a greater impact on the estimated HR, driving it closer to the null than in situations where the true HR is smaller.

Since the assessment of non-inferiority is based on the CI approach, a combination of greater bias in the negative direction and smaller SEs would yield a lower upper limit of the 95% CI, which is more likely to fall within the non-inferiority margin. Therefore, it is no coincidence that in general the scenarios with the greater bias and smaller SEs corresponded to the scenarios with

larger type I error rates. We observed that within each approach, the SEs were comparable for random and non-random crossovers, but the bias was larger for non-random crossover, suggesting that the bias had a greater influence on the type I error rate when comparing non-random versus random crossover within each method.

Our study had some limitations. The generalisability of our findings may be limited since we studied trials with event rates that are pertinent to radiotherapy trials in patients with early stage breast cancer. However, our methodology and results can be applied to other clinical settings where crossovers occur prior to initiation of treatment. In diseases where the event rates differ from the ones evaluated in this research, further simulations would be required to evaluate these approaches. For simplicity, we assumed that non-random crossover was

Figure 2 Bias for the ITT, PP and AT approaches by crossover type and percentage. (A) Overall; (B) Random crossover; (C) Non-random crossover. ITT, intention-to-treat; PP, per-protocol; AT, as-treated.



based on a single covariate. However, non-random crossover can occur for several reasons and, depending on the reason the for crossover, the hazards may also differ. We did not consider adjusting for baseline covariates in the analysis, which may improve the estimation of the treatment effect. However, this is less likely in large RCTs. We did not evaluate the causal proportional hazards estimator³⁶ because it is not readily available in standard statistical software. Other complex methods based on the counterfactual framework, such as the randomised-based analysis which employs *g*-estimation methods, were also not evaluated.^{15 16}

The choice of analysis population for non-inferiority trials is a difficult issue. We have shown that the AT approach preserves type I error under scenarios of random crossover. However, it is difficult to prove that crossover is random, and therefore assuming a random crossover may not be appropriate leading to concerns about the validity of the inference test. Moreover, the PP approach, which excludes patients, is likely to disturb the prognostic balance achieved by randomisation, which can also cause erroneous trial results. The advantage of the ITT approach is that it preserves the advantages of randomisation and mirrors what will happen in practice, and therefore is pragmatic. On the other hand, it can be anticonservative in situations where crossover is high. In our experience, the crossover percentage in radiotherapy trials in patients with early stage breast cancer is less than 2%, and we have shown that the AT and combined ITT+PP approaches are better at handling crossovers than the ITT and PP approaches.

CONCLUSION

The design, conduct and analysis of non-inferiority trials should be performed with extra rigour and to the highest standards. Attempts to prevent crossovers and other protocol deviations such as dropouts and losses to follow-up should be maximised. If a minimal percentage of crossovers were to occur, we have shown that the AT approach had the lowest type I error rates and smallest bias. A sensitivity analysis using the combined ITT+PP approach may also be warranted. In addition, both the ITT and PP results should be reported with details of the patients who crossed over.

Contributors SP, JAJ, CG, LT, TJW and MNL conceived the study. SP conducted the literature review, designed and implemented the simulation and wrote the initial draft of the manuscript. All authors reviewed and revised the draft version of the manuscript, and they also read and approved the final version of the manuscript.

Funding This research was funded in part by funds from the CANNeCTIN Program.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-

commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Bentzen SM, Agrawal RK, Aird EG, *et al*; START Trialists' Group. The UK Standardisation of Breast Radiotherapy (START) trial A of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *Lancet Oncol* 2008;9:331–41.
2. Bentzen SM, Agrawal RK, Aird EG, *et al*; START Trialists' Group. The UK Standardisation of Breast Radiotherapy (START) trial B of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *Lancet* 2008;371:1098–107.
3. Whelan TJ, Pignol JP, Levine MN, *et al*. Long-term results of hypofractionated radiation therapy for breast cancer. *N Engl J Med* 2010;362:513–20.
4. Olivetto IA, Whelan TJ, Parpia S, *et al*. Interim cosmetic and toxicity results from RAPID: a randomized trial of accelerated partial breast irradiation using three-dimensional conformal external beam radiation therapy. *J Clin Oncol* 2013;31:4038–45.
5. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Stat Med* 2003;22:169–86.
6. James Hung HM, Wang SJ, Tsong Y, *et al*. Some fundamental issues with non-inferiority testing in active controlled trials. *Stat Med* 2003;22:213–25.
7. Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008;27:317–32.
8. Fleming TR, Odem-Davis K, Rothmann MD, *et al*. Some essential considerations in the design and conduct of non-inferiority trials. *Clin Trials* 2011;8:432–9.
9. DeMets DL, Friedman L. Some thoughts on challenges for noninferiority study designs. *Drug Info J* 2012;12:420–7.
10. ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials—E9. 1998.
11. Garrett AD. Therapeutic equivalence: fallacies and falsification. *Stat Med* 2003;22:741–62.
12. Wiens BL, Zhao W. The role of intention to treat in analysis of noninferiority studies. *Clin Trials* 2007;4:286–91.
13. Brown D. Noninferiority trials in regulatory guidance and marketing authorization applications: huge advances over the last 20 years but problems still to be solved. *Stat Biopharm Res* 2013;5:223–8.
14. James Hung HM, Wang SJ. Statistical considerations for noninferiority trial designs without placebo. *Stat Biopharm Res* 2013;5:239–47.
15. Morden JP, Lambert PC, Latimer N, *et al*. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol* 2011;11:4.
16. Matsuyama Y. A comparison of the results of intent-to-treat, per-protocol, and *g*-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. *Stat Med* 2010;29:2107–16.
17. Lee YJ, Ellenberg JH, Hirtz DG, *et al*. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991;10:1595–605.
18. McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Stat Med* 2009;28:2639–52.
19. Rohmel J. Therapeutic equivalence investigations: statistical considerations. *Stat Med* 1998;17:1703–14.
20. Robins JM. Correction for non-compliance in equivalence trials. *Stat Med* 1998;17:269–302.
21. Hauck WW, Anderson S. Some issues in the design and analysis of equivalence trials. *Drug Info J* 1999;33:109–18.
22. Matilde Sanchez M, Chen X. Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Stat Med* 2006;25:1169–81.
23. Sheng D, Kim MY. The effects of non-compliance on intent-to-treat analysis of equivalence trials. *Stat Med* 2006;25:1183–99.
24. Robins JM, Tsiatis A. Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. *Commun Stat Theory Methods* 1991;20:2609–31.
25. Law MG, Kaldor JM. Survival analyses of randomized clinical trials adjusted for patients who switch treatments. *Stat Med* 1996;15:2069–76.
26. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. *Stat Med* 2002;21:2449–63.

27. Shao J, Chang M, Chow SC. Statistical inference for cancer trials with treatment switching. *Stat Med* 2005;24:1783–90.
28. Odondi L, McNamee R. Performance of statistical methods for analysing survival data in the presence of non-random compliance. *Stat Med* 2010;29:2994–3003.
29. White IR. Uses and limitations of randomization-based efficacy estimators. *Stat Methods Med Res* 2005;14:327–47.
30. Gomberg-Maitland M, Frison L, Halperin JL. Active-control clinical trials to establish equivalence or noninferiority: methodological and statistical concepts linked to quality. *Am Heart J* 2003;146:398–403.
31. Committee on Proprietary Medical Products Point-to-Consider. Points to consider on switching between superiority and non-inferiority. 2000.
32. Brittain E, Lin D. A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Stat Med* 2005;24:1–10.
33. Jung SH, Kang SJ, McCall LM, *et al.* Sample size computation for two-sample noninferiority log-rank test. *J Biopharm Stat* 2005;15:969–79.
34. Voduc KD, Cheang MC, Tyldesley S, *et al.* Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* 2010;28:1684–91.
35. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med* 2005;24:1713–23.
36. Loeyes T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics* 2003;59:100–5.