Impact of Non-Randomised Dropout on Treatment Switching Adjustment in the Relapsing-Remitting Multiple Sclerosis CLARITY Trial and the CLARITY Extension Study

Running Title: Switch and dropout in CLARITY and extension

Précis: A novel combination of methods are applied to adjust for treatment switching and non-randomised dropout in the CLARITY trial and the CLARITY extension study

Acknowledgements: Not applicable

Word Count: 4000/4000

Number of pages: 20

Number of figures: 1

Number of tables: 1

Appendix: 3 pages, 0 figures and 1 table

Abstract: 270/250 words

Objectives: Statistical methods to adjust for treatment switching are commonly applied to randomised controlled trials (RCTs) in oncology. However, RCTs with extension studies incorporating non-randomised dropout require consideration of alternative adjustment methods. The current study utilized a recognised method and a novel method to adjust for treatment switching in relapsing–remitting multiple sclerosis (MS).

Methods: The CLARITY RCT evaluated the efficacy of cladribine versus placebo over 96 weeks. Many (but not all) CLARITY participants enrolled in the 96-week CLARITY extension study; placebo-treated patients from CLARITY received cladribine (PP→LL) and cladribine-treated patients were re-randomised to placebo (LL→PP) or continued cladribine (LL→LL). End points were time to first qualifying relapse (FQR), time to 3-month, and 6-month confirmed disability progression (3mCDP, 6mCDP). We aimed to estimate the effectiveness of the LL→PP treatment strategy compared to a counterfactual (unobserved) PP→PP strategy. We applied the commonly-used rank preserving structural failure time model (RPSFTM) and a novel approach that combined propensity score matching (PSM) with inverse probability of censoring weights (IPCW).

Results: The RPSFTM resulted in LL→PP versus PP→PP hazard ratios (HRs) of 0.48 (95% confidence interval [CI] 0.36-0.62) for FQR, 0.62 (95% CI 0.46-0.84) for 3mCDP, and 0.62 (95% CI 0.44-0.88) for 6mCDP. The PSM+IPCW resulted in HRs of 0.47 (95% CI 0.38-0.63) for FQR, 0.61 (95% CI 0.43-0.86) for 3mCDP, and 0.63 (95% CI 0.40-0.87) for 6mCDP.

Conclusions: The PSM+IPCW HRs were consistent with those from the RPSFTM, suggesting the results were not substantially biased by informative dropout, assuming that all relevant

confounders were controlled for. There was no statistical evidence of a reduction in the

cladribine treatment effect during the extension period.

Highlights

- In our previous paper we applied RPSFTM and IPE analyses to adjust for treatment switching between the CLARITY trial and CLARITY extension study.

- This paper applies an alternative combination of statistical methods for adjusting for treatment switching, which may be useful for situations in which trial characteristics hamper the application of standard adjustment techniques. Results were consistent with those produced by the RPSFTM and IPE methods.

- There was no statistical evidence of a reduction in the cladribine treatment effect during the extension period.

1. Introduction

Statistical methods to adjust for treatment switching are often applied to data from oncology trials,(1-3) but are rarely applied to clinical trials in other disease areas. In oncology trials, a switch typically occurs when patients in the control group are permitted to switch onto the intervention treatment following disease progression (1-3) while a proportion of non-switching patients remain on the control treatment.(1, 2) For this reason, a majority of studies have discussed treatment switching adjustment methods with this context in mind.(4)

To adjust for this treatment switching, the NICE Decision Support Unit (DSU) Technical support document (TSD) 16 suggests four methods that can be used to adjust time to event efficacy estimates for treatment switching: the rank preserving structural failure time model (RPSFTM), the iterative parameter estimation (IPE) algorithm, inverse probability of censoring weights (IPCW), and a two-stage accelerated failure time adjustment method.(4) The RPSFTM is the most commonly applied to randomised controlled trials (RCT) and requires assumptions of randomisation and common treatment effect are met.(4, 5) Unfortunately, these assumptions do not strictly hold in more complex designs in which informative drop-out occurs or if the average treatment effect received by patients who switch is different than the effect seen in patients originally randomised to the treatment group.

Beyond the methods recommended by DSU TSD 16,(4) it is possible to apply alternative methods, specific to the circumstances of the data, to adjust for treatment switching and other sources of confounding. Unlike the RPSFTM, propensity score matching (PSM) coupled with inverse probability of censoring weights (IPCW), together referred to as PSM+IPCW, can be used to account for potential informative drop-out. The PSM+IPCW method does not rely on a

common treatment effect assumption, but it does require that all relevant confounders be included in the matching process.(6, 7)

The current study presents an application of the RPSFTM and PSM+IPCW to an RCT(8) and subsequent extension study(9) assessing the efficacy of cladribine tablets (MAVENCLAD®, Merck KGaA), which are used to treat relapsing–remitting multiple sclerosis (MS). Relapsing–remitting MS is a chronic autoimmune neurodegenerative disease that progresses over a long period of time.(10) Guidelines for the management of this condition advises early treatment with the use of disease modifying therapies (DMTs), which can favourably alter the course of the disease.(11) Many approved DMTs are administered parenterally, or through self-injection, whereas cladribine tablets represent one of a number of orally administered DMT options. Given the possibility that informative dropout occurred between the original RCT and the extension study as not all patients who entered the original RCT went on to the extension study, the assumption of randomization is potentially violated. Consequently, the aim of the current study was to supplement the RPSFTM analysis with an alternative analysis (PSM+IPCW) that could account for the potential that the drop-out is informative, in order to test the sensitivity of the results of the RPSFTM method.

2. Methods

Trial Design

The Cladribine Tablets Treating Multiple Sclerosis Orally (CLARITY) trial(8) and subsequent extension study(9) provide an example of a DMT trial with a complex follow-up study design. In the CLARITY trial patients were randomised 1:1:1 to receive low-dose (3.5 mg/kg) cladribine tablets (n=433), high-dose (5.25 mg/kg) cladribine tablets (n=437), or placebo (n=437). As the

purpose of DMTs is to delay or prevent disability progression and relapses in MS patients, assessed outcomes were time to 3-month disability progression (3mCDP), 6-month disability progression (6mCDP) and first qualifying relapse (FQR). See Giovannoni and colleagues for a full description of the CLARITY study methodology, including outcomes as well as patient inclusion and exclusion criteria.(8)

After the completion of the initial 96 weeks of CLARITY, there was a gap period (median of 40.3 weeks) before the start of the CLARITY extension study.(9) The CLARITY extension study consisted of 806 patients from the CLARITY study. Figure 1 provides an overview of the assignment of CLARITY patients to the CLARITY extension. Of the 437 patients randomised to receive placebo in CLARITY, 171 patients did not enroll in CLARITY extension, and 22 entered the extension for follow-up but did not receive treatment. The remaining 244 patients in the CLARITY placebo group (PP) who enrolled in the extension received low-dose cladribine (LL) in the extension study period (this cohort is labelled PP→LL). Of the 433 patients who received low-dose cladribine in CLARITY, 132 did not enroll in CLARITY extension, and 284 of the CLARITY low-dose group patients who enrolled in the extension study were re-randomised to receive either low-dose (LL→LL; n=186 patients) or placebo (LL→PP; n=98 patients) in the extension period. The remaining 17 patients entered the extension for follow-up but were not randomised to receive treatment.

Simple intention-to-treat (ITT) analyses of the extension study revealed similar outcomes for the PP→LL and LL→PP groups: 3mCDP [hazard ratio (HR)=0.86 (95% confidence interval (CI) 0.40-1.83)], 6mCDP [HR=0.88 (95% CI 0.40-1.93)], and FQR [HR=0.53 (95% CI 0.26-1.10)]. These findings suggest some degree of carryover of initial treatment with cladribine but do not speak to the long-term treatment effect of cladribine compared to placebo. Ideally, time to event

outcomes for LL➔PP could be compared with outcomes for PP➔PP. In the absence of a placebo comparator that continued across both 96-week periods, it was necessary to create a counterfactual PP➔PP arm by applying complex statistical methods to adjust for the treatment switching from placebo to low-dose cladribine (PP➔LL). The creation of the counterfactual arm allowed the estimation of what would have happened if patients had remained on placebo during the extension study.

Insert figure 1

Application of RPSFTM

RPSFTM was applied to the combined CLARITY plus CLARITY extension study datasets to estimate HRs for a treatment arm of low-dose cladribine in CLARITY followed by placebo in the extension study (LL➔PP) compared to placebo in CLARITY and the extension study (PP➔PP). The gap period time between the CLARITY and extension study was included as part of the time to event, and was assumed to be part of the placebo (off-treatment) period in both arms. In the context of CLARITY and the extension study, the randomisation assumption may have not been satisfied because not all patients who were included in the LL and PP groups of CLARITY chose to take part in the extension study. Consequently, for the results of the RPSFTM to be valid, it must be assumed that patients who did not experience an event of interest during CLARITY and who subsequently entered the extension study were representative of all patients who had not experienced an event of interest during CLARITY. In addition, the common treatment effect assumption is a strong assumption to make for several reasons. First, although the efficacy of low-dose treatment is expected to be sustained over a four-year period for LL➔PP patients, the treatment effect may not be perfectly maintained during the placebo

9

phase of the trial. Secondly, patients receiving the treatment later may have less capacity to benefit from treatment. Finally, if drop-out between CLARITY and the extension study was informative, patients who received cladribine in the extension study phase may have had a different capacity to benefit compared to the patients who were initially randomised to cladribine in CLARITY.(5)

Application of PSM+IPCW

Due to the limitations associated with the RPSFTM, the sensitivity of the estimated treatment effects was examined using an alternative adjustment method that did not require the same assumptions as RPSFTM. To relax the common treatment effect assumption, PSM was used to create samples of placebo and low-dose cladribine patients at CLARITY baseline. These samples closely resembled the characteristics, measured at baseline of the extension study, of the placebo group patients that went on to receive low-dose cladribine in the extension. Using these matched samples, an acceleration factor was estimated using a Weibull accelerated failure time model specific to this group of patients in order to represent the treatment effect of low-dose cladribine observed in CLARITY for this group. This acceleration factor was then used to shrink observed event times for those placebo group patients that received low-dose cladribine in the extension study and had not experienced an event of interest during CLARITY. This resulted in creation of a placebo counterfactual arm for the entire CLARITY plus CLARITY extension period. More specifically, counterfactual survival times for the PPLL group were obtained using Equation 1;

$$U_i = T_{A_i} + \frac{T_{B_i}}{\mu_B} \qquad (1)$$

Where T_{A_i} represents time until first event while in CLARITY and the gap period, T_{B_i} represents time until first event while in the extension study, and μ_B is the estimated acceleration factor. We

estimated HRs for the LL→PP versus PP→PP comparison using Cox proportional hazard models applied to the counterfactual dataset.

Although this analysis does not rely on the common treatment effect assumption, it does rely on the conditional independence assumption (CIA) and common support assumption.(6, 7) The CIA states that, conditional on a set of variables, the outcome of interest is independent of treatment.(6, 7), According to the CIA, if all prognostic characteristics are controlled for by matching patients who received low-dose cladribine in CLARITY with patients who received low-dose cladribine in the extension study, there will be no difference in the observed event times between the two groups. Hence, the assumption is essentially equivalent to the no unmeasured confounders assumption required by the IPCW. In most situations this assumption is untestable, because we do not usually have comparison groups that both receive the same treatment. However, in this case, we are able to test the validity of this assumption by comparing outcomes in patients who received low-dose cladribine in the extension study to outcomes in patients who received low-dose cladribine in CLARITY; a HR close to 1 would indicate the CIA holds. For the CLARITY placebo matched group, this assumption cannot be directly tested because the extension counterfactual for placebo is unobserved. However, the balance of the characteristics using standardised differences and likelihood ratio tests from logistic regressions on this matched sample can be assessed, these would provide an indication of how well the matching performed.

The common support assumption requires that for each patient who was treated with low-dose cladribine in the extension study and did not experience an event of interest in CLARITY, there is a patient in the CLARITY comparison group with similar characteristics. Hence, it is not possible to perform matching if all patients in the low-dose cladribine group in the extension

study are fundamentally different, in terms of a characteristic that could affect prognosis, from all patients at CLARITY baseline. This assumption may be challenging to satisfy given patients who entered CLARITY were required to have had at least one relapse within the previous 12 months, whereas by definition, the patients in our CLARITY extension sub-sample had not previously experienced an event for at least 96 weeks. This difference could impact a patient's capacity to benefit from treatment. However, if the HR of low-dose in extension versus low-dose in CLARITY is close to 1, we can be reasonably confident that the capacity to benefit is not greatly affected by this difference.

To ensure the estimated acceleration factor closely reflected the treatment effect received by extension group patients, a range of different matching algorithms was applied. To identify the preferred application of matching for each endpoint, the validity of the CIA and covariate balance of the matched samples was assessed. The number of unmatchable patients in the extension group sample was also examined and the maximum weights were applied to patients to assess whether the matched sample relied on few repeated observations. More details on these assessments are provided in the supplementary materials.

The proportion of patients that dropped out between the CLARITY and extension studies is a threat to meeting the randomisation assumption. Although this assumption is intrinsic to the RPSFTM, the PSM adjustment method does not rely upon it to obtain the adjustment factor used to estimate counterfactual survival times in PP➔LL patients. However, estimates of the HR using the PSM adjustment method may still, potentially, be subject to bias from informative drop-out. To mitigate this potential risk, the inverse probability of censoring weights (IPCW) was used to adjust for potentially informative dropout. The IPCW method weights patients who enrolled in the extension study and those with similar characteristics to be similar to those who

12

did not enroll in the extension study, thus removing bias from informative drop-out.(12) The

IPCW method relies on a "no unmeasured confounders" assumption, which requires that all

relevant prognostic characteristics are included in the model.(12) The covariates included in the

IPCW and PSM models were age, sex, region, time since first attack, prior use of any DMTs,

expanded disability status scale (EDSS), T1 Gd-enhancing lesion volume, T1 hypointense lesion

volume, T2 lesion volume, binary indicators of the number of T1 Gd-enhancing lesions, and the

number of T1 hypointense lesions (1 if 10 or more, and 0 if otherwise). All analyses were

performed in Stata 13.(13)

4. Results

The time to event results for 3mCDP, 6mCDP, and FQR are presented in Table 1. HRs for the RPSFTM, PSM adjusted and PSM+IPCW adjusted approaches are presented in addition to ITT (LL→PP vs PP→LL) and CLARITY ITT HRs. The ITT (LL→PP vs PP→LL) represents a comparison of the observed trial data without adjustment for treatment switching. All treatment switching adjustment analyses produced numerically lower HRs than the ITT (LL→PP vs PP→LL) analysis.

Insert table 1

For 3mCDP, the RPSFTM results indicate a LL→PP versus PP→PP HR of 0.62 (95% CI 0.46 to 0.84) over the entire CLARITY plus CLARITY extension period. For 6mCDP, the RPSFTM estimated a LL→PP versus PP→PP HR of 0.62 (95% CI 0.44 to 0.88); for FQR, a LL→PP versus PP→PP HR of 0.48 (95% CI 0.36 to 0.62).

The results of the preferred PSM and PSM+IPCW analyses are presented with bootstrapped confidence intervals. For 3mCDP, the preferred PSM adjustment method resulted in a LLPP versus PPPP HR of 0.60 (95% CI 0.44 to 0.83), for 6mCDP, a LLPP versus PPPP HR of 0.62 (95% CI 0.40 to 0.84), and for FQR, a LLPP versus PPPP HR of 0.48 (95% CI 0.37 to 0.63). Each of these point estimates are similar in magnitude to the RPSFTM estimates.

To adjust for potential informative dropout due to non-enrollment in the extension study after CLARITY, IPCW to the PSM was applied to the preferred PSM analyses. The results indicate that the PSM+IPCW HR point estimates were very similar to the HRs for the PSM adjustment and RPSFTM for each end point.

5. Discussion

Adjusting for treatment switching involves substantial methodological uncertainty because the different adjustment methods, can be applied in different ways. Decision makers may be reluctant to use results of standalone adjustment analyses due to concern that the adjustment analyses reported by manufacturers might be those that produce the most favourable results for the experimental treatment.(14) In this study, alternative adjustment methods have been used to validate the results and assumptions of a commonly used treatment switching adjustment method (i.e. RPSFTM).

The application of treatment switching adjustment methods in the case of CLARITY and CLARITY extension is complicated, as the switching mechanism differs from the standard case and there was dropout between the end of CLARITY and enrolment in CLARITY extension.. To address risks to the common treatment effect and randomisation assumptions associated with the RPSFTM, alternative PSM, and PSM+IPCW adjustment methods, which rely on different assumptions, were utilized. The PSM and PSM+IPCW applications that performed most successfully, in terms of matching algorithm performance, produced results that were very similar to those from the RPSFTM analyses for each of the three outcomes of interest. The estimated HRs did not differ by more than 0.02 for any outcome, and further confidence in the results was provided by the similarity of the adjusted CLARITY plus CLARITY extension HRs with the CLARITY ITT HRs.

The PSM adjusted analyses primarily relied on the CIA. The HRs derived by comparing events in the "LL" period in the "PP→LL" group to events experienced in the matched CLARITY LL sample were used to test the validity of the CIA for the matched samples. For 3-month progression and 6-month progression, some of the matching techniques that were tested resulted

in CLARITY LL matched samples which could closely replicate the time to event outcomes in the "PP→LL" group. However, it should be noted that the performance in this measure was not consistent across all matching techniques. This lack of consistency suggests the CIA may have not been fully met. One difference that could not be controlled for between CLARITY baseline and CLARITY extension baseline arises from the fact that only patients with one or more relapse within the previous 12 months entered CLARITY. By definition, patients in the PP→LL sample of interest had not experienced an event for at least 96 weeks at extension baseline, hence this could represent an important difference between the samples.

Although the limitations in the matching attempts indicate there may be some bias in the results of the PSM analysis, it is important to note that for the two disability progression outcomes the matching did appear to perform well. For the third outcome, time to first relapse, we would expect that the bias would likely cause an underestimation of the treatment benefit of low-dose cladribine compared with placebo; a comparison of the effectiveness of cladribine in the PP→LL arm with the LL arm from CLARITY indicates that the treatment effect may have actually been higher in switchers than in the group initially randomised to cladribine, even when these groups are matched on observable characteristics at extension and baseline respectively. However, this finding should be interpreted with caution because matching may be imperfect.

Although the PSM analysis accounts for potential differences in the treatment effect in those who enrolled into the extension study compared to those who were originally randomised into CLARITY, informative censoring from the dropout of patients that did not enroll in the extension study could remain an issue when estimating the LLPP versus PPPP treatment effect over the entire CLARITY plus extension time period. Therefore, the IPCW was applied to the counterfactual datasets of our preferred matched analyses to adjust for informative censoring.

For each end point, the HRs from our preferred IPCW analyses differed from our analyses that were not adjusted for informative censoring by only 0.01. This indicates censoring at the end of CLARITY and the start of the extension study did not cause substantial bias in the analyses, assuming all important confounders were controlled for in the model. The IPCW relies on an untestable "no unmeasured confounders" assumption. In an attempt to include all relevant confounders, the models incorporated data on all characteristics assumed to affect prognosis. The maximum stabilised weights derived from the models were relatively low, hence it is unlikely that the analysis would have been biased due to the application of a large weighting to a time point for a patient that is non-representative of the sample in general. Another potential source of bias arises from the fact that time-dependent data on prognostic characteristics was not collected during the gap period. For the patients who had not experienced an event before the end of CLARITY, the last observation was carried forward until the next observation at the start of the extension. For those who experienced an event during the gap, the last observation was carried forward until an event occurred. This could explain the relatively low stabilised weight because there were a number of time points during the gap period in which variables have values that did not change, even though the values may have changed in reality. This missing information is a limitation of the analysis, however given the data available, our analyses advance the best attempt to address the potential for informative censoring due to non-informative dropout.

In conclusion, this study applied a novel PSM+IPCW method to adjust for treatment switching in the context of a trial combined with an extension study, with less than full enrolment and in which control group patients switch treatments in the extension. The study has shown how to systematically assess different PSM approaches and how to identify which applications appear

17

to be most appropriate. The end results are similar to those from the RPSFTM, which is helpful

for decision makers who may be unsure about the validity of a particular method or

assumptions. The similarity of the adjusted HRs indicates that the RPSFTM results are not

substantially affected by bias from informative dropout, assuming that all relevant patient

characteristics were incorporated into the IPCW model and that the results were not affected by

data collection limitations during the gap period. In addition, comparisons of the adjusted

CLARITY plus CLARITY extension HRs with CLARITY HRs indicate that there is no

statistical evidence for waning of the 3.5 mg/kg dose cladribine treatment effect during the

extension period.

6. References

1.   Latimer NR, Abrams KR, Amonkar MM, Stapelkamp C, Swann RS. Adjusting for the Confounding Effects of Treatment Switching-The BREAK-3 Trial: Dabrafenib Versus Dacarbazine. Oncologist. 2015;20(7):798-805.
2.   Latimer NR, Bell H, Abrams KR, Amonkar MM, Casey M. Adjusting for treatment switching in the METRIC study shows further improved overall survival with trametinib compared with chemotherapy. Cancer Med. 2016;5(5):806-15.
3.   Latimer NR, Henshall C, Siebert U, Bell H. Treatment Switching: Statistical and Decision-Making Challenges and Approaches. Int J Technol Assess Health Care. 2016;32(3):160-6.
4.   Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. Sheffield: Decision Support Unit, 2014.
5.   Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. Comm Stats Theory Method. 1991;20(8):2609-31.
6.   Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv. 2015;22(1):31-72.
7.   Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for casual effects. Biometrika. 1983;70(1):41-55.
8.   Giovannoni G, Comi G, Cook S, Rammohan K, Rieckmann P, Soelberg Sorensen P, et al. A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis. The New England journal of medicine. 2010;362(5):416-26.
9.   Comi G, Cook S, Rammohan K, Soelberg Sorensen P, Vermersch P, Adeniji AK, et al. Long-term effects of cladribine tablets on MRI activity outcomes in patients with relapsing-remitting multiple sclerosis: the CLARITY Extension study. Therapeutic advances in neurological disorders. 2018;11.
10.   Weinshenker BG. Epidemiology of multiple sclerosis. Neurol Clin. 1996;14(2):291-308.
11.   Goodin DS, Frohman EM, Garmany GP, Jr., Halper J, Likosky WH, Lublin FD, et al. Disease modifying therapies in multiple sclerosis: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and the MS Council for Clinical Practice Guidelines. Neurology. 2002;58(2):169-78.
12.   Robins JM, 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(3):779-88.
13.   Stata statistical software, Version 13.0, Texas, USA: 2013.
14.   Latimer NR. Treatment switching in oncology trials and the acceptability of adjustment methods. Expert Rev Pharmacoecon Outcomes Res. 2015;15(4):561-4.

Tables and figures

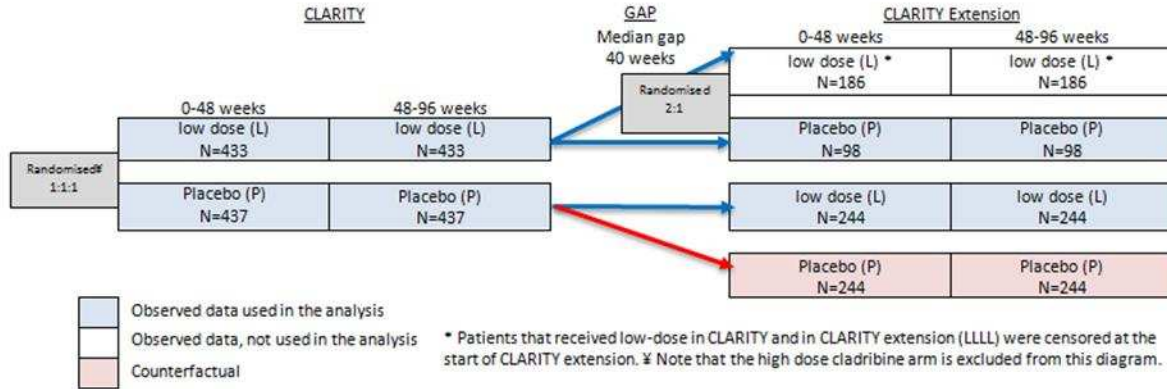**Figure 1**: CLARITY and CLARITY extension arms used for analysis

Table 1: ITT and treatment switching adjusted HRs

Method	HR		
	Point estimate	Lower 95% CI	Upper 95% CI
Time to 3-month progression			
ITT (LLPP vs PPLL)	0.67	0.52	0.87
CLARITY ITT (LL vs PP)	0.60	0.41	0.87
RPSFTM treatment group no re-censoring (LLPP vs PPPP)	0.62	0.46	0.84
PSM (LLPP vs PPPP)	0.60	0.44	0.83
PSM + IPCW (LLPP vs PPPP)	0.61	0.43	0.86
Time to 6-month progression			
ITT (LLPP vs PPLL)	0.67	0.50	0.90
CLARITY ITT (LL vs PP)	0.58	0.40	0.83
RPSFTM treatment group no re-censoring (LLPP vs PPPP)	0.62	0.44	0.88
PSM (LLPP vs PPPP)	0.62	0.40	0.84
PSM + IPCW (LLPP vs PPPP)	0.63	0.40	0.87
Time to first qualifying relapse			
ITT (LLPP vs PPLL)	0.53	0.43	0.67
CLARITY ITT (LL vs PP)	0.44	0.34	0.58
RPSFTM treatment group no re-censoring (LLPP vs PPPP)	0.48	0.36	0.62
PSM (LLPP vs PPPP)	0.48	0.37	0.63
PSM + IPCW (LLPP vs PPPP)	0.47	0.38	0.63

CLARITY: Cladribine Tablets Treating Multiple Sclerosis Orally (trial), HR: Hazard Ratio, IPCW: Inverse Probability of Censoring Weights, ITT: Intention to treat, LL: low-dose Cladribine in CLARITY, LLPP low-dose Cladribine in CLARITY followed by placebo in CLARITY extension, NN: nearest-neighbour, PP: placebo in CLARITY, PPLL: placebo in CLARITY followed by low-dose Cladribine in CLARITY extension, PPPP: placebo in CLARITY and CLARITY extension (counterfactual arm), PS: Propensity score, RPSFTM: Rank Preserving Structural Failure Time Model.