
Article:
Bell Gorrod, H. orcid.org/0000-0001-8054-8073, Latimer, N.R., Damian, D. et al. (3 more authors) (2019) Assessing the long-term effectiveness of cladribine vs. placebo in the relapsing-remitting multiple sclerosis CLARITY randomized controlled trial and CLARITY extension using treatment switching adjustment methods. Advances in Therapy. ISSN 0741-238X

https://doi.org/10.1007/s12325-019-01140-z

Assessing the long-term effectiveness of cladribine vs. placebo in the relapsing-remitting multiple sclerosis CLARITY randomized controlled trial and CLARITY Extension using treatment switching adjustment methods

Helen Bell Gorrod, PhD[1]; Nicholas R. Latimer, PhD[1]; Doris Damian, PhD[2]; Robert Hettle, MMath[3]; Gerard T. Harty, MSc[2]; Schiffon L. Wong, MPH[2]

[1]School for Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK; [2]EMD Serono, Inc., Billerica, MA, USA; [3]PAREXEL International, London, UK

Corresponding author

Helen Bell Gorrod, PhD, Health Economics and Decision Modelling (HEDS), School for Health and Related Research (ScHARR), Regent Court, University of Sheffield, S1 4DA, UK

Tel: +44 (0)114 222 0890; E-mail: helen.bellgorrod@sheffield.ac.uk

Running head: Switching adjustment in CLARITY and extension

Word Count: 4,000

Abstract

Objectives: Treatment switching adjustment methods are often used to adjust for switching in oncology randomized controlled trials (RCTs). In this exploratory analysis, we apply these methods to adjust for treatment changes in the setting of an RCT followed by an extension study in relapsing-remitting multiple sclerosis.

Methods: The CLARITY trial evaluated cladribine tablets versus placebo over 96 weeks. In the 96-week CLARITY Extension, patients who received placebo in CLARITY received cladribine tablets; patients who received cladribine tablets in CLARITY were re-randomized to placebo or cladribine tablets. Endpoints were time to first qualifying relapse (FQR) and time to 3- and 6-month confirmed disability progression (3mCDP, 6mCDP). We aimed to compare the effectiveness of cladribine tablets to placebo over CLARITY and the extension. The rank preserving structural failure time model (RPSFTM) and Iterative Parameter Estimation (IPE) were used to estimate what would have happened if patients had received placebo in CLARITY and the extension, versus patients that received cladribine tablets and switched to placebo. To gauge whether treatment effect waned after the 96 weeks of CLARITY, we compared hazard ratios (HRs) from the adjustment analysis with HRs from CLARITY.

Results: The RPSFTM resulted in a HR 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. IPE algorithm results were similar. CLARITY HRs were 0.44 (95% CI 0.34-0.58), 0.60 (95% CI 0.41-0.87) and 0.58 (95% CI 0.40-0.83) for FQR, 3mCDP and 6mCDP respectively.

Conclusions: Treatment switching adjustment methods are applicable in non-oncology settings. Adjusted CLARITY plus CLARITY Extension HRs were similar to the CLARITY HRs, demonstrating significant treatment benefits associated with cladribine tablets versus placebo.

Keywords: multiple sclerosis, treatment switching, rank-preserving structural failure time model, iterative parameter estimation, adjustment methods, time-to-event

Key Points

This article reports on an exploratory analysis, where the authors apply treatment switching adjustment methods to adjust for treatment changes in the setting of a randomized controlled trial (CLARITY) followed by an extension study (CLARITY Extension) in relapsing-remitting multiple sclerosis. This analysis demonstrates that treatment switching adjustment methods are applicable in non-oncology settings, and when trial designs are non-typical - i.e. in the context of extension studies, where it is common for patients randomized to placebo in the initial trial to all switch onto the experimental treatment in the extension study for ethical reasons. Use of adjustment methods in this setting provides decision makers with longer-term evidence for relative treatment benefits in situations where there would otherwise be none.

1. Introduction

Treatment switching adjustment methods are often used to adjust for situations in oncology randomized controlled trials (RCTs) in which patients are permitted to switch from the control treatment to the intervention treatment after disease progression.[1-4] Results of these treatment switching adjustment analyses have been included as supporting evidence in a number of international health technology appraisals (HTAs) of cancer treatments.[5] However, these methods have rarely been applied outside of oncology.

The National Institute for Health and Care Excellence Decision Support Unit Technical Support Document 16 describes four commonly used methods for adjusting for treatment switching in RCTs.[4] These are the rank-preserving structural failure time model (RPSFTM),[6] the iterative parameter estimation (IPE) algorithm,[7] inverse probability of censoring weights (IPCW),[8] and a two-stage accelerated failure time adjustment method.[9] In the strictest sense, the IPE is a type of RPSFTM, although here we specify IPE to distinguish this method from the standard RPSFTM.

These methods for adjusting for treatment switching were examined using data from an RCT and extension study focused on assessing the efficacy of cladribine tablets (MAVENCLAD®, Merck KGaA), which are used to treat relapsing multiple sclerosis (MS). Relapsing–remitting MS is a chronic autoimmune neurodegenerative disease that progresses over a long period of time.[10] Management of the condition often involves the use of disease modifying treatments (DMTs), which can favorably change the course of the disease.[11] Many approved DMTs are administered parenterally, or through self-injection, whereas cladribine tablets represent one of the newer orally administered DMT options.

In this analysis, we apply treatment switching adjustment methods to an RCT followed by an extension study with re-randomization, in the context of MS. Extension trials involving treatment switching are common in MS due to ethical reasons; hence, these methods are applicable beyond this case-study. Our aim is to overcome the limitation in the data that patients in the placebo group of the trial switch onto an active treatment in the extension study, because this limits the assessment of long-term effectiveness of treatment compared with placebo using standard methods. To correct for this, we apply treatment switching adjustment methods to estimate the long term effectiveness of an active MS treatment compared to a counterfactual placebo arm over the duration of the RCT and extension period.

2. Methods

2.1.Trial design

The Cladribine Tablets Treating Multiple Sclerosis Orally (CLARITY, ClinicalTrials.gov: NCT00213135) trial compared low-dose cladribine tablets (3.5 mg/kg; now the approved dose), high-dose cladribine tablets (5.25 mg/kg), and placebo over a 96-week period.[12] Each treatment course of cladribine tablets consisted of two treatment weeks per 48 week period, administered during the beginning of the first month and the beginning of the second month of the respective treatment year. 1326 patients were randomized 1:1:1 to compare three lines of treatment during the 96 weeks of the study. 437 patients received a placebo treatment, 433 patients received low-dose cladribine tablets and 437 patients received high-dose cladribine tablets.[12] A full description of the CLARITY study methodology, including patient inclusion and exclusion criteria, has been published elsewhere.[12]

After completion of CLARITY, participants were permitted to enter into a 96-week extension study, in which former placebo patients were assigned to treatment with low-dose cladribine and former low-dose cladribine patients were randomized 2:1 to receive further treatment with low-dose cladribine tablets or placebo. Former high-dose patients were also included in the extension study, but the focus of this paper is on low-dose cladribine tablets, representing the dose approved by regulators.[13]

The CLARITY Extension study was not pre-specified at the initiation of CLARITY and enrollment to the extension study did not begin immediately as the first patients in CLARITY completed the core study. The median per-patient time between the end of CLARITY and start of CLARITY Extension was 40.3 weeks. 806 of the patients in the CLARITY study enrolled in

the CLARITY Extension. Figure 1 indicates the treatment pathways for patients randomized to receive placebo or low-dose cladribine tablets in CLARITY.

CLARITY and the CLARITY Extension were conducted in accordance with the Declaration of Helsinki (1964) and its later amendments [14] and the Good Clinical Practice guidelines in accordance with the International Conference of Harmonisation.[15] The protocols for CLARITY and CLARITY Extension were reviewed and approved by the relevant local review board or ethics committee at each participating study center. Further information has been published previously elsewhere. [12] Due to ethical considerations, patients who received placebo in CLARITY received low-dose cladribine tablets in the extension; thus, no group received placebo in both CLARITY and CLARITY Extension. Of the 437 patients randomized to receive placebo in CLARITY, 171 patients did not enroll in the CLARITY Extension, and 22 enrolled in the extension but did not receive treatment. The remaining 244 placebo group patients that enrolled into the extension trial received low-dose cladribine tablets. Of the 433 patients that received low-dose cladribine in CLARITY, 132 did not enroll in the CLARITY Extension. 284 of the CLARITY low-dose group patients that enrolled in the extension study were re-randomized to receive either low-dose cladribine tablets (186 patients) or placebo (98 patients). The remaining 17 low-dose cladribine patients enrolled in the extension for follow-up but were not randomized to receive treatment.

As cladribine tablets demonstrated sustained long-term effects (see section 5.1 of the Summary of Product Characteristics[13]), patients who received placebo (PP) after low-dose cladribine tablets (LL) still benefited from initial treatment with cladribine during placebo treatment in the extension phase. Our aim was to assess the efficacy of a low-dose cladribine, low-dose cladribine, placebo, placebo treatment arm (denoted LLPP), compared to a placebo, placebo,

placebo, placebo arm (denoted PPPP), with each treatment block lasting 48 weeks. Note that the first two 48-week blocks were in CLARITY and the second two occurred in CLARITY Extension, after the bridging interval between studies. The intention was that all patients treated with placebo during CLARITY trial switched onto cladribine tablets in the extension study for ethical reasons, and therefore the PPPP treatment strategy was unobserved. Therefore, we used treatment switching adjustment methods to estimate the hypothetical outcome of the CLARITY placebo arm if patients had remained on placebo during the extension study and had not switched onto low-dose cladribine tablet treatment. We sought to compare LLPP to PPPP for the outcomes of 3-month disability progression (3mCDP), 6-month disability progression (6mCDP) and first qualifying relapse (FQR).

2.2.Treatment switching adjustment methods

Among the four commonly used methods for adjusting for treatment switching in RCTs, the IPCW and the two-stage method both require a proportion of non-switchers to be present in the arm that is to be adjusted.[4, 9, 8] Following CLARITY, all placebo group patients who enrolled in the extension switched onto low-dose cladribine tablets. For this reason, we could not adjust for treatment switching using the IPCW or the two-stage method. However, we could apply the RPSFTM and IPE methods because these do not require there to be non-switchers in the estimation process, provided that the mean treatment duration differs between patients in the initially randomized groups. As these methods have been described in detail elsewhere,[2, 4] here we describe only their key characteristics.

2.2.1. Application of the RPSFTM and IPE

The applications of the RPSFTM and IPE involve two main steps. First, the value of the treatment effect, ψ, is estimated. Typically, the RPSFTM and IPE use the randomization of the trial to estimate counterfactual event times, i.e., event times that would have been observed if no treatment had been received.[6] It is assumed that if no treatment was received by patients in either arm of the trial, the event times for each arm would be, on average, equal. Counterfactual event times for each patient can be estimated as follows:

$$(\text{time to event})_i = (\text{time off treatment})_i + \exp(\psi) \times (\text{time on treatment})_i$$

where ψ represents the treatment effect, and i represents patient i.

Hence, the value for the treatment effect ψ can be identified at the point where the counterfactual event times are balanced in each arm. For the RPSFTM, this is achieved when the value of the log-rank Z test comparing counterfactual event times between arms is equal to zero.[6, 16] For the IPE, the treatment effect is determined at the point where the algorithm has converged.[7] The standard RPSFTM uses g-estimation to estimate the value of the treatment effect; this essentially consists of a grid search of possible values for ψ until one is found that results in equal counterfactual event times across treatment arms.[6] In contrast, the IPE uses an iterative parametric testing procedure to estimate the value of the treatment effect.[7] In both cases, the treatment effect is estimated in the form of an acceleration factor (AF), where $AF=\exp(-\psi)$.

Second, counterfactual survival times in the group(s) in which switching occurred are compared to observed survival times in the group(s) in which switching did not occur, to obtain adjusted estimates of the treatment effect.

2.2.2. RPSFTM and IPE assumptions

The RPSFTM and IPE rely on two key assumptions: the randomization assumption and the common treatment effect assumption.[7, 6] The randomization assumption assumes that, in the absence of treatment, outcomes would be equal in the randomized groups. This is likely to be reasonable in the context of an RCT. The common treatment effect assumption assumes that the effect of treatment is the same regardless of when it is received, relative to the amount of time for which it is taken. If the treatment effect received by switchers differs from that received by patients at initial treatment randomization, the RPSFTM and IPE will produce biased results. This assumption may be more problematic if patients who receive treatment later in the trial are considered to have a lower capacity to benefit from it.

We applied the RPSFTM and IPE methods to the CLARITY plus CLARITY Extension context by combining the study periods and using a treatment switch indicator to represent the time when low-dose cladribine was started in patients who were initially randomized to placebo. In these analyses, we assumed the treatment benefit could be maintained after treatment had been discontinued; therefore, the treatment indicator remained set to "1" (i.e., "on" treatment) for all time periods after treatment was initiated. Patients who followed the LLPP treatment arm were modelled to remain "on treatment" during both periods because the impact of the treatment may still affect prognosis after treatment has been discontinued, and patients who followed the a placebo, placebo, low-dose cladribine, low-dose cladribine treatment arm (PPLL) treatment arm were modelled as being "off treatment" during the CLARITY period and "on treatment" during the extension period of the trial. Applying the methods in this way allows for the fact that sustained benefit has been observed after cladribine treatment has been discontinued, [15] and

also ensures that mean treatment durations will differ in the initially randomized groups – a requirement for RPSFTM and IPE models to be applicable.

After counterfactual event times were obtained for switchers, hazard ratios (HRs) were estimated, using Cox proportional hazards regression models, for time from randomization to event (e.g., relapse or disability progression) for the LLPP arm compared with the counterfactual PPPP arm. For the RPSFTM, confidence intervals (CIs) were estimated for the adjusted HRs using a test-based procedure that retained the p-value from the unadjusted intention to treat (ITT) analysis of the combined CLARITY plus CLARITY Extension datasets (for the comparison of LLPP and PPLL).[16] For the IPE estimates, we performed 1000 bootstrapped iterations of the adjustment procedure to estimate CIs around the adjusted HR.

2.2.3. Selecting the preferred application

Although they use different processes to obtain a value of ψ, the RPSFTM and IPE perform in similar ways and are likely to produce similar results.[2, 4] However, because they use different estimation procedures, one method may produce less biased estimates than the other; for instance, g-estimation can be problematic if it identifies multiple possible solutions that provide equal counterfactual event times between treatment groups. Therefore, it is useful to run both RPSFTM and IPE analyses and compare the results. If the results are similar, this provides confidence that the results are not sensitive to the process used to identify ψ. To identify which of the two methods fits the data better, we compared the counterfactual event times in each arm (i.e. counterfactual untreated event times in the LLPP arm compared to counterfactual untreated event times in the PPLL arm) by estimating an HR. If the method has worked successfully, the counterfactual HR should be equal to one. This implies that the counterfactual event times were

balanced between randomized arms; thus, conditions for the successful estimation of the treatment effect were met.

2.2.4. Assessing the validity of the assumptions

The patients who received treatment in the PPLL arm were a subset of patients who were originally randomized in CLARITY, who had not previously experienced an event, and who chose to enroll in CLARITY Extension. This subset could have had a greater or diminished potential to benefit from treatment. Therefore, we tested the sensitivity of the results to violations of the common treatment effect in both directions. To assess the sensitivity of the results to the common treatment effect assumption, we repeated the analysis with a treatment effect decrement of 20% for switchers. We also tested the sensitivity of the results to a potentially larger treatment effect in the PPLL arm by running an analysis in which the treatment effect in the PPLL arm was assumed to be 20% higher than in the LLPP arm. The 20% threshold, while arbitrary, was chosen as it is large enough to enable us to determine whether violations of the common treatment effect assumption were likely to be important. To assess the plausibility of the randomization assumption, we compared characteristics observed at the end of the CLARITY study between the groups of patients who did versus did not enter the extension study, and who had not previously experience an event of interest. If the mean characteristics of these groups are similar, it is likely the dropout is non-informative and will not lead to bias in the results. Further analyses related to potential biases associated with dropout between CLARITY and CLARITY Extension are published elsewhere.[17] All analyses were performed in Stata 13[18] using the strbee command.[16] Financial support for this study was provided entirely by a contract with EMD Serono, Inc. (a business of Merck KGaA, Darmstadt, Germany). The funding agreement

ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. GH and SW are employed by the sponsor.

3. Results

Table 1 provides a descriptive summary of patient characteristics at the end of CLARITY for the subgroup of patients who had not experienced an event and did or did not enroll in the extension study. Note that those that received LLLL were grouped together with the LL arm patients that did not enter the extension study in table 1, because both groups were censored at the end of CLARITY in our analysis. These mean values were mostly similar between patient groups, however there were some statistically significant differences in terms of for EDSS, volume of T1 Hypointense lesions and volume of T2 lesions. The results of the time to event analyses for 3mCDP, 6mCDP, and FQR are presented in Table 2. Alongside the treatment switching adjusted HRs for the RPSFTM and IPE (LLPP versus PPPP), we present CLARITY ITT HRs, which compared the LL versus PP during CLARITY, and ITT HRs (LLPP versus PPLL), which compared the unadjusted arms of CLARITY plus CLARITY Extension. All treatment switching adjustment analyses produce numerically lower HRs than the ITT (LLPP versus PPLL) analyses.

For 3mCDP, the RPSFTM produced a HR of 0.62 (95% CI 0.46 to 0.84) over the CLARITY plus CLARITY Extension period. Figure 2 presents the Kaplan Meier plot of the observed and counterfactual datasets associated with this RPSFTM analysis. The RPSFTM results estimated a HR of 0.62 (95% CI 0.44 to 0.88) for 6mCDP and a HR of 0.48 (95% CI 0.36 to 0.62) for FQR. The estimates of the IPE analyses are similar to the RPSFTM estimates for all analyses. The counterfactual HR for the RPSFTM is closer to 1, which indicated the RPSFTM performed more successfully than the IPE.

Table 3 presents the results of the common treatment effect assumption sensitivity analysis. For each endpoint, the HR decreased by a maximum of 0.01 for a 20% increase in treatment benefit

for the PPLL arm and increased by a maximum of 0.01 for a 20% reduction in the treatment benefit. This analysis indicates the results were not highly sensitive to violations of the common treatment effect.

4. Discussion

This application of treatment switching adjustment methods to an RCT followed by an extension study demonstrated it is possible to apply the RPSFTM and IPE in a non-oncology context in a situation where all control group patients switched onto the experimental treatment. Although the IPE method has previously been applied in the context of an MS trial,[19] these analyses provide useful additional information to decision makers on longer-term treatment effect estimates. Treatment switching adjustment methods are relevant when patients in the trial switch treatments, resulting in a situation where the observed data do not represent the desired treatment comparison. However, while it is possible to apply adjustment methods in the setting investigated, because all placebo group patients who entered the extension study received cladribine tablets, the uncertainty around the adjustment analyses is high and these analyses should be considered exploratory. Yet the results of the current analysis appear to have face-validity, given that they produce similar estimates of the treatment effect as those observed at the end of the initial CLARITY study period.

The CLARITY plus CLARITY Extension study context in which we apply the RPSFTM and IPE is not typical, because we are not studying a single RCT, but an RCT with an extension study. The randomization assumption is critical for the RPSFTM and IPE and is more problematic in the context of an RCT followed by an extension in which not all RCT participants enrolled. Following CLARITY, 62% (806/1307) of patients enrolled in the extension study. This could potentially cause bias if the dropout was not random, i.e., if dropout is correlated with prognostic factors or other factors that influence the treatment effect. Assessment of the prognostic variables in each group indicated that the groups of patients who did and did not enter the extension study were mostly balanced at the end of the CLARITY (see table 1). The study

was not specifically powered to assess the statistical significance of differences in mean characteristics between these groups, hence p-values should be interpreted with caution, however there were some statistically significant differences in means for EDSS, volume of T1 Hypointense lesions and volume of T2 lesions. On average, those that had not previously experienced an event and did not enrol into the extension had a smaller volume of T2 lesions and higher EDSS than those that did enrol, hence the potential impact of this on the results is unclear. It is important to note that unmeasured differences could also be present between the two groups – this cannot be ruled out and could lead to biased results.[19] A detailed exploration of the impact of potentially informative dropout on the results of adjustment methods applied in the CLARITY plus CLARITY Extension context is published elsewhere – with findings suggesting that informative dropout did not appear to be an important issue, in terms of measured characteristics.[17] Hence, even though there did appear to be some differences in characteristics between patients who did and did not enter the extension study, this did not appear to impact on estimates of the long term treatment effect.[17]

The common treatment effect assumption required that we assume dropout between CLARITY and the extension study was not due to factors that influence the effectiveness of the treatment. Our analyses suggested adjusted HR estimates were robust to significant violations of the common treatment effect assumption (with a 20% increase or decrement in the treatment effect in switchers). Therefore, the results are unlikely to be substantially biased even if placebo group patients who did not enroll in the extension study had a different capacity to benefit from treatment to patients who did enroll. Doubt may be cast on the common treatment effect assumption due to the difference in time between patients in the LLPP group and patients in the PPLL group receiving cladribine tablets. However, given that relapsing-remitting MS is typically

a slowly progressive disorder, it may not be unreasonable to assume that treatment is similarly beneficial even if there are differences in the time of initiation. Additionally, our sensitivity analyses suggest this was uninfluential because the results were robust to violations of the common treatment effect assumption.

The bridging interval between the end of CLARITY and the beginning of the CLARITY Extension introduces additional considerations. In addition to the non-informative/random dropout assumption, we must assume that the effect of any other treatments received during the bridging interval is balanced in each arm of the trial. This seems reasonable, given that only 2 (2%) patients in the LLPP arm and 4 (2%) patients in the PPLL arm received treatment during the bridging interval. Ultimately, the effect of alternative treatment on outcomes was negligible. Furthermore, results were robust to the adjustment method used such that RPSFTM and IPE provided very similar adjusted HR estimates for LLPP compared to PPPP. The RPSFTM appeared to perform marginally better than the IPE given it produced slightly better matched counterfactual event times between treatment arms.

To gauge whether the effect of cladribine tablets appeared to wane over time, we compared the HRs from our treatment switching adjusted analyses (LLPP versus PPPP) with the HRs from the CLARITY ITT analyses (LL versus PP). For time to 3mCDP, 6mCDP and FQR, the point estimates of the adjusted (LLPP versus PPPP) HRs were less than 10% worse than the CLARITY ITT (LL versus PP) HRs (0.62, 0.62, and 0.48 versus 0.60, 0.58, and 0.44, respectively). This might indicate a slight decrease of the treatment effect over the subsequent 96-week placebo period of the extension study; however, there was no clear evidence to suggest that the clinical treatment benefit of cladribine tablets waned during the extension.

It is often appropriate to apply re-censoring to the data when the RPSFTM or IPE methods are used to break the dependency between treatment received and censoring times. [20] In this case, re-censoring was not necessary because all of the patients in the PPLL arm switched treatment. All patients in the control arm that continued onto the extension had their event times adjusted, so there should be no dependency between the treatment received and censoring time within the PPLL group. We tested the sensitivity of our results to the application of re-censoring and found it did not make any substantial difference to the results.

Finally, although we have undertaken rigorous analyses in an attempt to decipher as much as possible from the CLARITY and extension studies, additional information on longer term effects would be useful. For example, observational data collection could provide further evidence on the long-term effectiveness of cladribine tablets in real-world settings.

In conclusion, our analysis shows the RPSFTM and IPE treatment switching adjustment methods can be applied in non-oncology settings, within the context of a double-blinded extension trial combined with an initial RCT and 100% switching from control onto experimental treatment. This helps to provide decision makers with evidence of longer-term treatment effectiveness in situations where placebo group patients in extension studies must switch onto treatment due to ethical reasons. A key limitation of the study, which can add uncertainty and should be considered alongside the results, is the potential for bias from non-random drop out at the end of CLARITY in terms of observed and unobserved characteristics. The comparisons of LL versus PP HRs and adjusted LLPP versus PPPP HRs provided no statistical evidence to suggest that the treatment benefit of 3.5 mg/kg (low-dose) cladribine tablets wanes over the subsequent 96 weeks when patients receive placebo and were robust to violations of analytic assumptions.

Acknowledgments

The authors wish to thank the participants of the study. CLARITY and the CLARITY Extension were conducted in accordance with the Declaration of Helsinki (1964) and its later amendments [14] and the Good Clinical Practice guidelines in accordance with the International Conference of Harmonisation.[15] The protocols for CLARITY and CLARITY Extension were reviewed and approved by the relevant local review board or ethics committee at each participating study center. Further information has been published previously elsewhere.[12**]** Due to ethical considerations, patients who received placebo in CLARITY received low-dose cladribine tablets in the extension; thus, no group received placebo in both CLARITY and CLARITY Extension.

Medical Writing Assistance

The authors also wish to thank Jason Allaire, PhD of Generativity Solutions Group for his assistance with editing the paper. Medical writing assistance was funded by EMD Serono, Inc.

Funding Support

Financial support for this study, medical writing support, article processing charges were provided entirely by EMD Serono, Inc. (a business of Merck KGaA, Darmstadt, Germany). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. All authors had full access to all of the data in this study and take complete responsibility for the integrity of the data and accuracy of the data analysis.

Conflict of Interests

NRL reports having received consultancy fees for providing modelling advice to Astra Zeneca, BMS and Pfizer, and funding from EMD Serono, Inc. (a business of Merck KGaA, Darmstadt,

Germany) to undertake the analysis presented in this paper. NRL was supported by the National Institute for Health Research (NIHR Post Doctoral Fellowship, Dr Nicholas Latimer, PDF-2015-08-022) while contributing to this work. He is now supported by Yorkshire Cancer Research (Award reference number S406NL). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health, or Yorkshire Cancer Research.

HBG declares that The University of Sheffield has received funding for consultancy work for pharmaceutical companies, which covered the cost of their time spent working on projects. HBG received personal consulting fees from PharmaMar, fees for providing training courses on statistical methods to Merck EMD Serono and Pfizer, fees for attending an advisory board meeting, funding to present the work at a conference and have claimed back travel expenses from Merck EMD Serono.

RH was an employee of PAREXEL International, who received payment for consultancy support to EMD Serono, during the conduct of this analysis.

DD, GTH & SW are employees of EMD Serono, Inc. (a business of Merck KGaA, Darmstadt, Germany)

Compliance with Ethics Guidelines

CLARITY and the CLARITY Extension were conducted in accordance with the Declaration of Helsinki (1964) and its later amendments [14] and the Good Clinical Practice guidelines in accordance with the International Conference of Harmonisation.[15] The protocols for CLARITY and CLARITY Extension were reviewed and approved by the relevant local review board or ethics committee at each participating study center. Further information has been published previously elsewhere. [12] Due to ethical considerations, patients who received

placebo in CLARITY received low-dose cladribine tablets in the extension; thus, no group received placebo in both CLARITY and CLARITY Extension.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due publicly available due to the fact that this was a randomized clinical trial but are available from the corresponding author on reasonable request.

References

1. 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. doi:10.1002/cam4.643.
2. 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. doi:10.1017/S026646231600026X.
3. 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. doi:10.1634/theoncologist.2014-0429.
4. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. Sheffield: Decision Support Unit2014.
5. Latimer NR. Treatment switching in oncology trials and the acceptability of adjustment methods. Expert Rev Pharmacoecon Outcomes Res. 2015;15(4):561-4. doi:10.1586/14737167.2015.1037835.
6. 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.
7. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. Stat Med. 2002;21(17):2449-63. doi:10.1002/sim.1219.
8. 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.
9. Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials--an economic evaluation context: methods, limitations, and recommendations. Med Decis Making. 2014;34(3):387-402. doi:10.1177/0272989X13520192.
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. 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. N Engl J Med. 2010;362(5):416-26. doi:10.1056/NEJMoa0902533.
13. eMC. MAVENCLAD 10mg tablets. 2017. https://www.medicines.org.uk/emc/medicine/34044. 18 December 2017.
14. World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Researh Involving Human Subjects. 2017. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. 14 July 2017.
15. Dixon JR. The international conference on harmonization good clinical practice guideline. Qual Assure. 1999;6(2):65-74.
16. White IR, Walker S, Babiker AG. strbee: Randomization–based efficacy estimator. Strata J. 2002;2(2):140-50.

17. Bell Gorrod H, Latimer NR, Damian D, Hettle R, Harty GT, Wong LS. Impact of non-randomised drop-out on treatment switching adjustment in the relapsing-remitting multiple sclerosis CLARITY trial and the CLARITY Extension study. Value in Health. 2019;22(7):772-6.
18. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet (London, England). 2018;391(10130):1622-36. doi:10.1016/s0140-6736(18)30481-1.
19. Meng X, Chin PS, Hashmonay R, Zahur Islam M, Cutter G. Effect of switching from intramuscular interferon beta-1a to oral fingolimod on time to relapse in patients with relapsing-remitting multiple sclerosis enrolled in a 1-year extension of TRANSFORMS. Contemp Clin Trials. 2015;41:69-74. doi:10.1016/j.cct.2014.12.011.
20. White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. Stat Med. 1999;18(19):2617-34.

Table 1 Descriptive statistics of characteristics at the end of CLARITY for those that had not experienced an event prior to the end of CLARITY, by event and arm of initial randomization

	enrolled into extension and did not experience an event during CLARITY			did not enroll into extension, or were re-randomized to LLLL, and did not experience an event during CLARITY			
	N	Mean	Standard deviation	N	Mean	Standard deviation	p value
3-month confirmed disability progression	PP group						
age	203	38.58	9.37	133	38.17	10.58	0.72
sex	203	0.34	0.48	133	0.27	0.45	0.15
EDSS at last assessment	203	2.66	1.26	133	2.94	1.53	0.08*
T1 Gd-enhanced, volume in mm at 96 weeks	203	103.68	386.23	133	87.5	275.14	0.65
Number of T1 Gd-enhanced lesions at 96 weeks	203	0.48	1.17	133	0.45	1.35	0.83
T1 Hypointense lesions, volume in mm at 96 weeks	203	1462.38	2292.39	133	1366.4	2546.53	0.73
Number of T1 Hypointense lesions at 96 weeks	203	5.61	6.07	133	5.14	5.05	0.45
T2 lesions, volume in mm at 96 weeks	203	13380.57	12285.46	133	11087.41	9282.01	0.05**
3-month confirmed disability progression	LL group						
Age	81	37.67	10.46	284	37.31	10.23	0.79
Sex	81	0.33	0.47	284	0.31	0.46	0.69
EDSS at last assessment	81	2.64	1.21	284	2.60	1.23	0.78
T1 Gd-enhanced, volume in mm at 96 weeks	81	30.19	195.12	284	11.21	70.17	0.39
Number of T1 Gd-enhanced lesions at 96 weeks	81	0.07	0.31	284	0.06	0.34	0.79
T1 Hypointense lesions, volume in mm at 96 weeks	81	1761.1	2917.56	284	1252.97	2055.33	0.15
Number of T1 Hypointense lesions at 96 weeks	81	5.75	7.84	284	5.25	6.09	0.59
T2 lesions, volume in mm at 96 weeks	81	14195.09	13444.51	284	11597.11	12746.56	0.12
6-month confirmed disability progression	PP group						
Age	214	38.47	9.37	148	38.24	10.77	0.83
Sex	214	0.36	0.48	148	0.30	0.46	0.31
EDSS at last assessment	214	2.68	1.26	148	3.09	1.64	0.01***
T1 Gd-enhanced, volume in mm at 96 weeks	214	106.29	380.78	148	86.63	267.16	0.56
Number of T1 Gd-enhanced lesions at 96 weeks	214	0.50	1.17	148	0.47	1.35	0.83
T1 Hypointense lesions, volume in mm at 96 weeks	214	1444.24	2258.46	148	1465.55	2540.59	0.93
Number of T1 Hypointense lesions at 96 weeks	214	5.53	6.02	148	5.45	5.20	0.88
T2 lesions, volume in mm at 96 weeks	214	13471.18	12293.98	148	11922.09	11264.67	0.22
6-month confirmed disability progression	LL group						
Age	84	37.51	10.49	300	37.41	10.29	0.94
Sex	84	0.33	0.47	300	0.31	0.46	0.73
EDSS at last assessment	84	2.65	1.21	300	2.63	1.27	0.88
T1 Gd-enhanced, volume in mm at 96 weeks	84	29.11	191.64	300	10.61	68.31	0.39
Number of T1 Gd-enhanced lesions at 96 weeks	84	0.07	0.30	300	0.06	0.33	0.76
T1 Hypointense lesions, volume in mm at 96 weeks	84	1724.51	2872.63	300	1290.33	2101.62	0.20
Number of T1 Hypointense lesions at 96 weeks	84	5.67	7.72	300	5.37	6.28	0.75
T2 lesions, volume in mm at 96 weeks	84	14609.87	14081.25	300	11799.52	12894.68	0.10
First qualifying relapse	PP group						
age	161	39.64	9.68	111	39.50	10.20	0.91
sex	161	0.35	0.48	111	0.27	0.45	0.14
EDSS at last assessment	161	2.85	1.38	111	3.19	1.70	0.09*
T1 Gd-enhanced, volume in mm at 96 weeks	161	86.63	386.40	111	94.75	326.19	0.85
Number of T1 Gd-enhanced lesions at 96 weeks	161	0.37	0.88	111	0.49	1.86	0.55
T1 Hypointense lesions, volume in mm at 96 weeks	161	1459.85	2283.15	111	1375.63	2613.73	0.78
Number of T1 Hypointense lesions at 96 weeks	161	5.72	6.24	111	5.19	4.75	0.43
T2 lesions, volume in mm at 96 weeks	161	13231.12	12430.89	111	11081.81	9669.58	0.11
First qualifying relapse	LL group						

age	82	38.48	9.98	266	37.95	10.25	0.68
sex	82	0.32	0.47	266	0.30	0.46	0.78
EDSS at last assessment	82	2.95	1.45	266	2.68	1.32	0.13
T1 Gd-enhanced, volume in mm at 96 weeks	82	29.73	193.90	266	9.86	67.53	0.36
Number of T1 Gd-enhanced lesions at 96 weeks	82	0.07	0.31	266	0.06	0.33	0.67
T1 Hypointense lesions, volume in mm at 96 weeks	82	2138.57	3145.42	266	1359.83	2206.04	0.04**
Number of T1 Hypointense lesions at 96 weeks	82	6.67	8.03	266	5.64	6.68	0.29
T2 lesions, volume in mm at 96 weeks	82	15829.02	13924.13	266	12024.73	13332.88	0.03**

CLARITY: Cladribine Tablets Treating Multiple Sclerosis Orally (trial), EDSS: Expanded Disability Status Scale, Gd: gadolinium, LL: low-dose cladribine in CLARITY, LLPP low-dose cladribine in CLARITY followed by placebo in CLARITY Extension, N: Number of observations, PP: placebo in CLARITY, PPLL: placebo in CLARITY followed by low-dose cladribine in CLARITY Extension

Table 2 ITT and treatment switching adjusted HRs

Method	HR			CF HR test
	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 (LLPP vs PPPP)	0.62	0.46	0.84	1.00
IPE (LLPP vs PPPP)	0.62	0.45	0.83	0.94
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 (LLPP vs PPPP)	0.62	0.44	0.88	1.01
IPE (LLPP vs PPPP)	0.62	0.43	0.87	0.98
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 (LLPP vs PPPP)	0.48	0.36	0.62	1.00
IPE (LLPP vs PPPP)	0.48	0.37	0.62	0.95

CF HR Test: Counterfactual Hazard Ratio Test, CLARITY: Cladribine Tablets Treating Multiple Sclerosis Orally (trial), CI: confidence interval, HR; hazard ratio, IPE: Iterative Parameter Estimation, ITT: Intention to treat, LL: low-dose cladribine in CLARITY, LLLL: low-dose cladribine in CLARITY followed by low-dose in CLARITY Extension; LLPP low-dose cladribine in CLARITY followed by placebo in CLARITY Extension, 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), RPSFTM: Rank Preserving Structural Failure Time Model.

Table 3 RPSFTM – Common treatment effect sensitivity analysis

Method	HR		
	Point estimate	Lower 95% CI	Upper 95% CI
Time to 3-month progression			
treatment benefit for PPLL is reduced by 20%	0.63	0.46	0.85
treatment benefit for PPLL is increased by 20%	0.61	0.44	0.84
Time to 6-month progression			
treatment benefit for PPLL is reduced by 20%	0.63	0.45	0.88
treatment benefit for PPLL is increased by 20%	0.61	0.42	0.87
Time to first qualifying relapse			
treatment benefit for PPLL is reduced by 20%	0.48	0.37	0.63
treatment benefit for PPLL is increased by 20%	0.47	0.36	0.62

CI: confidence interval, HR: hazard ratio, PPLL: placebo in CLARITY followed by low-dose cladribine in CLARITY Extension, RPSFTM: Rank Preserving Structural Failure Time Model.

Figure Legends

Fig 1 CLARITY and CLARITY Extension arms used for analysis

Fig 2 Kaplan Meier plot for 3-month progression