



Published in final edited form as:

Mult Scler. 2013 November ; 19(13): 1760–1764. doi:10.1177/1352458513485147.

The MRI Rule of Five: Predicting the Occurrence of Relapse

Charity J. Morgan,

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama USA

Ashutosh Ranjan,

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL USA

Inmaculada B. Aban, and

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL USA

Gary R. Cutter

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL USA

Abstract

Background—Clinical intuition suggests that a sharp increase in the number of enhancing lesions should signal an increased risk of relapse. The Rule of Five recommends that subjects exhibiting at least five lesions over the baseline level be referred for closer monitoring. This rule has been used as an informal safety criterion with limited formal evaluation.

Objective—To determine the best threshold for the Rule and to demonstrate its predictive validity for risk of subsequent relapses for MS trials.

Methods—We used logistic regression modeling to apply the Rule to patient data from a Phase II clinical trial. Predictive validity was ascertained using rate ratios and ROC curves.

Results—We found that, for these data, a threshold of five lesions over the baseline constituted the best definition of a threshold. Overall, 35% of subjects broke the Rule at least once. Breaking the rule increased the odds of imminent relapse by a factor of 3.2 (95% Confidence Interval: 1.8 to 5.5).

Conclusion—Breaking the Rule of Five was found to be a significant predictor of an imminent relapse. Length of follow-up and the number of lesions discovered via MRI were also significant predictors of relapse.

Keywords

Multiple sclerosis; relapsing-remitting; magnetic resonance imaging scans; safety monitoring; logistic regression

Introduction

The value of MRI is well accepted in MS despite equivocal evidence of its exact place in the monitoring of patients or as an outcome measure. MRI has been accepted as an integral part of diagnosis for over a decade.¹ However, conflicting evidence concerning its potential role as a surrogate outcome abounds. Petkau et al. were unable to show that contrast enhancing lesions had sufficient concurrent or predictive validity to support developing evidence of surrogacy.² However, that study combined information across trials which may have introduced variations in study design such as entrance criteria or definitions. Sormani et al. have also shown that contrast enhancing lesions (CEL) are a significant predictor of changes in disability as measured by the expanded disability status scale (EDSS).³

CEL have been reduced in all approved drugs for Relapsing Remitting MS and used in Phase II studies as an indicator of potential treatment effects. Although the recent results of the Laquinimod studies had limited reductions in the relapse rate, nevertheless the relationship between CEL and relapse rate was consistent.⁴ A number of studies have used CEL as a marker of safety for the Data and Safety Monitoring Board (DSMB). In this context, such a safety monitoring tool has been called the “Rule of Five” or “Cutter rule”. The rationale was based on data made available from the Neuroimmunology Branch at NIH (McFarland, personal communication). The data for developing the rule emanated from patients on therapy and consecutive MRI's from each. The average within person standard deviation of the number of CELs was estimated. Then, using a rule of thumb for statistical outliers, 3.6 standard deviation units were added and rounded up to the nearest CEL yielding five lesions as the threshold where the increase in the number of lesions exceeds what would be expected by chance. The approach took all subjects who were on therapy, irrespective of their baseline lesion level, and looked for activity that seemed outside the normal range of variability. The approach was not trying to define a new clinical endpoint; however, it was proposed as a monitoring tool to ensure watchful attention at a patient level. The idea was to develop a conservative and easy to apply tool to alert a DSMB to a potential problem even in a small sample of participants.

The first instance that the rule was used was in the CombiRx safety study, which led to funding by the National Institute of Neurological Diseases and Stroke as a Phase III Trial. For the pilot CombiRx trial, the safety monitoring rule referred a patient to the DSMB if two occurrences of five or more CELs over the baseline level on consecutive monthly MRIs occurred. During the pilot trial, no subject fulfilled this criterion. While the Rule was initially developed based on treated groups, the rule should also be useful for placebo groups if it truly identifies risk. The rule has been utilized in many current trials, but its performance has been published in only one recent paper by Riddell et al. due to the lack of availability of raw data on which to document the performance and to possibly justify its use.⁵

In principle, this rule should indicate the likelihood of a subsequent MS clinical event, such as a relapse. In this paper, we use CELs as the risk variable on an MRI scan and propose that there is some threshold such that if the number of lesions on a scan crosses this threshold,

then a relapse is more likely to occur soon thereafter. We will define this threshold as a difference between CELs for a current scan and a scan taken at baseline.

Data were kindly made available for the unrestrictive purpose of addressing this question. In the next section, we give the background of the MN-166 trial with its key results. We then investigate the optimal threshold for the Rule and examine its performance in the presence of other covariates using logistic regression with random effect modeling. Finally, the predictive validity of the rule is assessed using an ROC curve.

Data

Data were provided by MediciNova, Inc. at the request of the authors. The study results were from a two year, randomized, double-blind, and placebo-controlled Phase II clinical trial of MN-166. MN-166 is an orally administered compound that increases neuronal growth factors, inhibits leukotriene activity, phosphodiesterases, and nitric oxide synthase. In order to be eligible for enrollment in the study, subjects were required to be between 18 and 55 years of age, with a diagnosis of relapsing MS. Participants were also required to have an EDSS of 5.5 or less and at least one CEL at baseline. For the first year of the study, subjects were randomized to one of the three treatment groups. They received either 30mg of MN-166 daily, 60mg daily, or placebo. In the second year of the study, all patients were on drug. Patients who received 30 or 60mg of MN-166 per day during the first year remained on the original assigned dose for the second year; patients who received placebo during the first year were randomized to receive either 30 or 60mg of MN-166 per day. Every two months, subjects returned to the study center for a neurological exam, MRI, and review of relapse history.

Data were provided for 297 patients. One subject was excluded from our analyses because the baseline scan for this subject was not available. Three more subjects were excluded because they were not followed after the baseline scan. The remaining 293 patients were followed for an average of 658.5 days (sd=179.7). The average number of scans per patient and the average number of CELs at baseline were 10.5 (2.9) and 3.8 (5.8), respectively. One hundred ninety-one (191) patients experienced at least one relapse during the course of the study; for those subjects experiencing relapse, the average number of relapses was 2.0 (1.1) and the maximum number of relapses per patient was five. For these subjects, the mean time to first relapse was 211.9 (sd=169.8) days. There was a significant but small positive correlation found between the number of CELs exhibited at baseline and the overall number of relapses (Kendall's $\tau = 0.16$, $p < 0.001$). While there was a significant negative correlation between the time to the first relapse from baseline and the overall number of relapses (Kendall's $\tau = -0.59$, $p < 0.001$), there was no relationship observed between the number of days followed and overall number of relapses (Kendall's $\tau = 0.07$, $p = 0.19$).

Methods

Defining a Threshold for the Rule

Since the MRI rule was initially developed for use with subjects receiving treatment and the era of placebo-controlled trials is waning as well as for the sake of simplicity, we consider

only the 192 participants (94 on 30 mg, 98 on 60 mg) who were initially randomized to the two active treatment arms. We begin validation of the Rule by assessing the threshold.

We define “imminent relapse” as a relapse that will occur within 30 days of the MRI. For each scan, we use as our response variable whether a relapse occurred within 30 days of that scan. For a given threshold (i.e., the number of CELs over the baseline level), θ , we say that the rule has been “broken” if the number of CELs on a scan is at least the threshold greater than the number present at baseline..

Our goal is to find a value of the threshold that will best predict imminent relapse. To do so, we fit a logistic regression model with a random intercept for each subject and fixed effects for the time of follow-up and the indicator, C , for breaking the rule to the relapse data, for threshold (θ) ranging from 2 to 7. To ensure that our model fit the data as well as possible, we include both linear and quadratic terms for the time of follow-up. We also considered models with no time effects and models with interactions between the Rule and both the linear and quadratic time effects. Models were compared using model fit assessments based on the second order bias correction to Akaike’s Information Criterion (AICc).⁶⁻⁷ The model showing smallest AICc (indicating the better fitting results) was chosen as best.

Examining the performance of the Rule in the presence of covariates

We next investigated whether the model could be improved by the addition of covariates and whether the Rule is still predictive of relapse after other characteristics were considered. In addition to linear and quadratic time effects, to determine the best model, we considered the following covariates: drug dosage, number of CELs on the current scan and CELs at baseline. Using a stepwise selection procedure in a logistic model with a random intercept for each subject, each of the covariates was added to the basic model. The covariate whose proposed model had the lowest AICc was then added to the model, and the process was repeated. If the addition of the covariate resulted in a decreased AICc, then this covariate was included and the current model was updated. We repeated this step until no more covariates could be added.

Results

The model with a threshold of $\theta = 5$ lesions for the rule and which included linear and quadratic time effects but no interaction terms was selected as providing the best fit, i.e., model with the smallest AICc (Figure 2). The need to include time effects to adequately model these data is discussed in further detail below. Using a five lesion threshold for the Rule, 35.4% of the participants broke the rule at least once. We next fit this model using generalized estimating equations (GEE) in order to estimate the population average effect of breaking the Rule. We found breaking the rule increased the odds of relapse by a factor of 3.12 (95% confidence interval for the odds ratio: 1.70 to 5.71) (Table 1).

The final model from the stepwise selection procedure is displayed in Table 2. Covariates are listed in the order in which they entered the model, along with the AICc of the model after each covariate was entered.

We found the number of lesions on a scan was the best predictor of whether a relapse was likely to occur in the next 30 days. Both the linear and quadratic terms for length of follow-up also contributed. Interestingly, the risk of imminent relapse increased for the first 300 days of follow-up, and then decreased from that point onward. This finding could possibly be due to selective drop-out or to a delayed treatment effect. When the linear and quadratic time effects and the counts of CELs were included in the model, adding the indicator for breaking the Rule of Five to the model did not decrease the AICc. That is, breaking the Rule of Five did not significantly improve the prediction of the risk of imminent relapse once length of follow-up and CEL count had been taken into consideration.

Predictive Value of Final Model

How good is the Rule at predicting imminent relapse? One measure is the concordance index as measured by the area under the curve of a Receiver Operating Characteristic (ROC). An ROC curve plots the false positive rate for a model against its true positive rate and can be used to assess how well a model can be used for prediction. A model with a greater concordance index indicates better predictive value. The concordance index for the final model from the stepwise selection procedure is 0.83 (Figure 3). A model with a concordance index between 0.80 and 0.90 is typically considered to have ‘excellent’ discrimination power.⁸ Interestingly, a model that uses only the indicator for breaking the Rule of Five to predict imminent relapse has a concordance index of 0.84, suggesting that the Rule alone is indeed useful at predicting imminent relapse.

Discussion

The justification for the choice of five CELs over baseline has been based on a rule of thumb for the detection of outliers in statistical analyses. It has been used effectively for monitoring in a number of trials. This paper provides empirical evidence that the risk associated with breaking this Rule can be quantified, thereby justifying the Rule’s use. Despite the heavily skewed distribution of enhancing lesion counts, we were able to show that five lesions is not an unreasonable threshold. Using the AICc, we concluded that five or more lesions was the best choice of threshold for these data. However, we note here that models that included time effects and used thresholds of four and six lesions had similar AICc values (ranging from 718.2 to 719.8) to the best-fitting model (AICc = 717.7), which used a five lesion threshold. Furthermore, changing the threshold to either four or six did not substantially affect prediction; models using a four- or six-lesion threshold had concordance indices ranging from 0.83 to 0.84. Thus the predictive value of the MRI Rule appears to be somewhat robust to the choice of threshold. For example, in another attempt to evaluate the Rule of Five, Riddell et al. showed that higher values also seemed to work with increasing success.⁵

While finding individual thresholds based on differing baseline values might be an alternative rule, the simplicity of a single number over baseline that shows an increased odds ratio and relative risk of nearly 3 for a relapse within 30 days should make acceptance of this rule for clinical trial monitoring easier. Furthermore the Rule translates easily to clinical

practice. As more therapies become available, such a threshold may indeed become useful for treatment monitoring.

We believe that the failure of prior studies to demonstrate a strong link between MRI and relapses may, in part, be due to the time horizon of the prediction. Taking the baseline MRI gadolinium counts and expecting them to predict which patients will have relapses over a year or two may have been the fatal flaw in the logic of clinical predictions from these counts. Here, we find the MRI Rule is useful for predicting “imminent” relapse, or relapse within 30 days. The predictive value of the Rule is considerably weakened, with the concordance index decreasing to 0.55, if imminent relapse is redefined as relapse within 60 days (results not shown). Thus, the predictive power of CEL diminishes with time. Over time there are likely other occurrences of high lesion counts in patients initially low in CELs. The misclassification induced by not considering these bursts of inflammation also reduces the correlation between baseline CELs and subsequent relapses.

While the use of the Rule has not been a major issue in trials, it has lacked documented evaluation of its performance. While we have limited our attention here to subjects with relapsing MS enrolled in the active treatment arms of one trial, we believe that the MRI Rule will work in a variety of settings as it has already been widely used by DSMBs across multiple MS clinical trials. The analyses conducted here should be applied to other cohorts in order to demonstrate the generalizability of the Rule. In this paper, we have provided a stronger basis for using this rule and present important information on the time course of the risk, which also can help explain the poor correlations found between MRI lesions and relapses in prior work.

Acknowledgments

The authors would like to thank Coretta Robinson for her assistance in editing the manuscript for non-technical content. They also thank MediciNova, Inc. and Dr. Rick Ladin for providing the data.

Partial support was received from NINDS 1 T32 NS054584-01A1 and NINDS 5 U01 NS045719.

References

1. O'Connor P. Key issues in the diagnosis and treatment of multiple sclerosis: An overview. *Neurology*. 2002; 59:S1–S33. [PubMed: 12448786]
2. Petkau J, Reingold SC, Held U, et al. Magnetic resonance imaging as surrogate outcome for multiple sclerosis relapses. *Mult Scler*. 2008; 14:770–778. [PubMed: 18535021]
3. Sormani MP, Li DK, Bruzzi P, et al. Combined MRI lesions and relapses as a surrogate for disability in multiple sclerosis. *Neurology*. 2011; 77:1684–1690. [PubMed: 21975200]
4. Comi G, Jeffery D, Kappos L, et al. Placebo-controlled trial of oral laquinimod for multiple sclerosis. *N Engl J Med*. 2012; 366:1000–1009. [PubMed: 22417253]
5. Riddell CA, Zhao Y, Li DKB, et al. Evaluation of safety monitoring guidelines based on MRI lesion activity in multiple sclerosis. *Neurology*. 2011; 77:2089–2096. [PubMed: 22094474]
6. Akaike H. A new look at the statistical model identification. *IEEE Trans on Autom Control*. 1974; 19:716–723.
7. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. *Biometrika*. 1989; 76:297–307.
8. Hosmer, DW.; Lemeshow, S. *Applied Logistic Regression*. 2nd ed.. New York: Wiley; 2000.

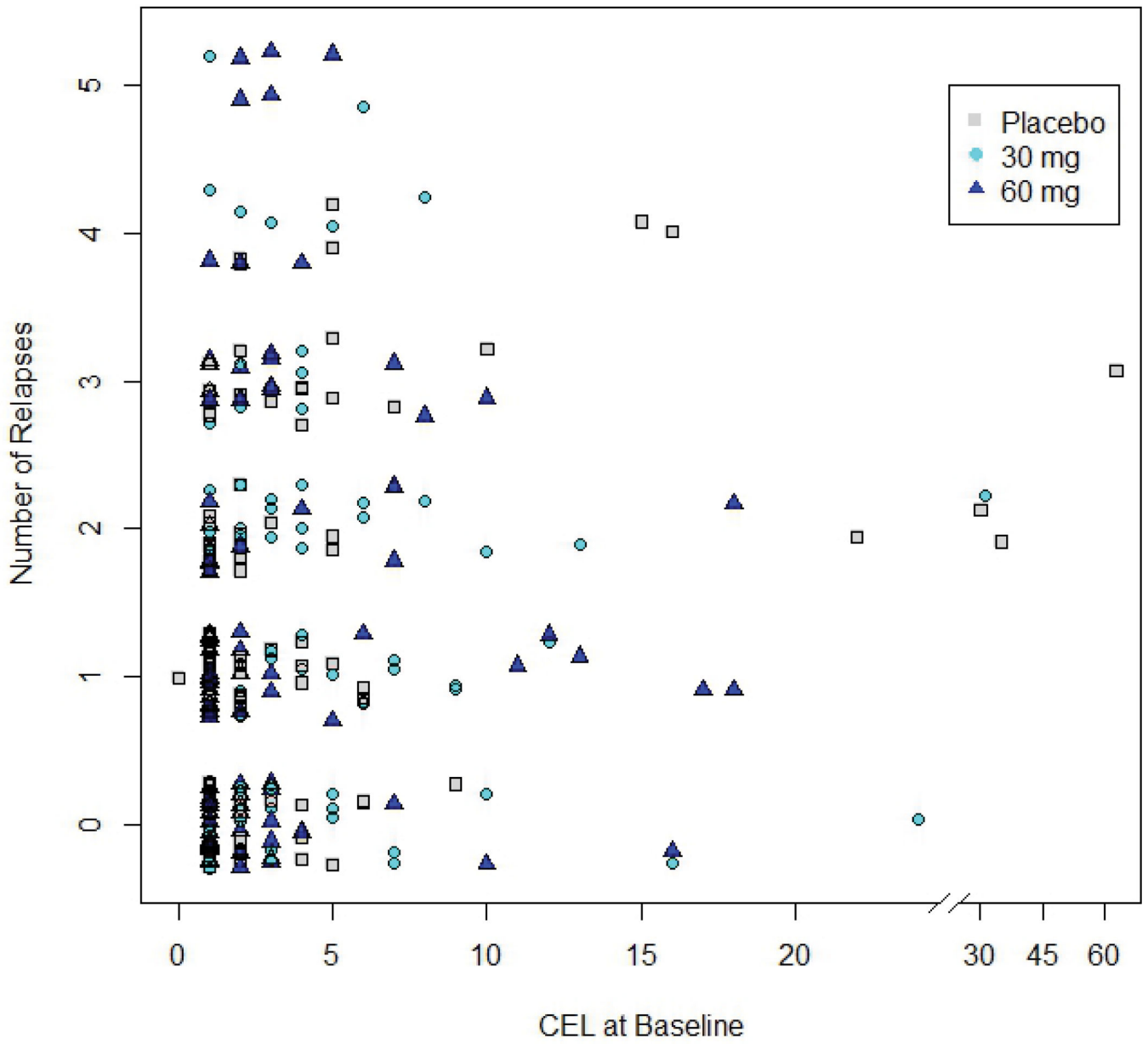


Figure 1.

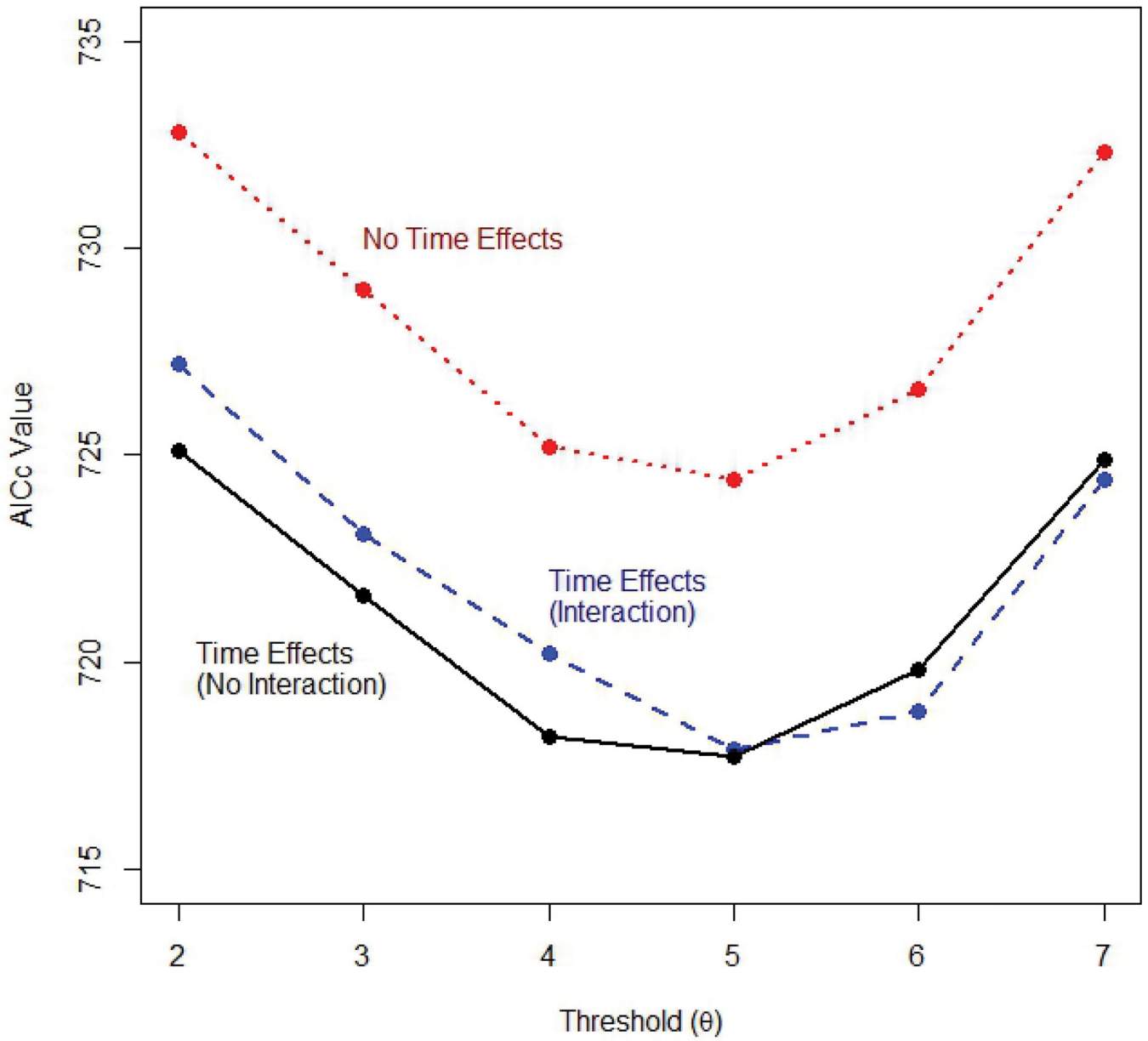


Figure 2.

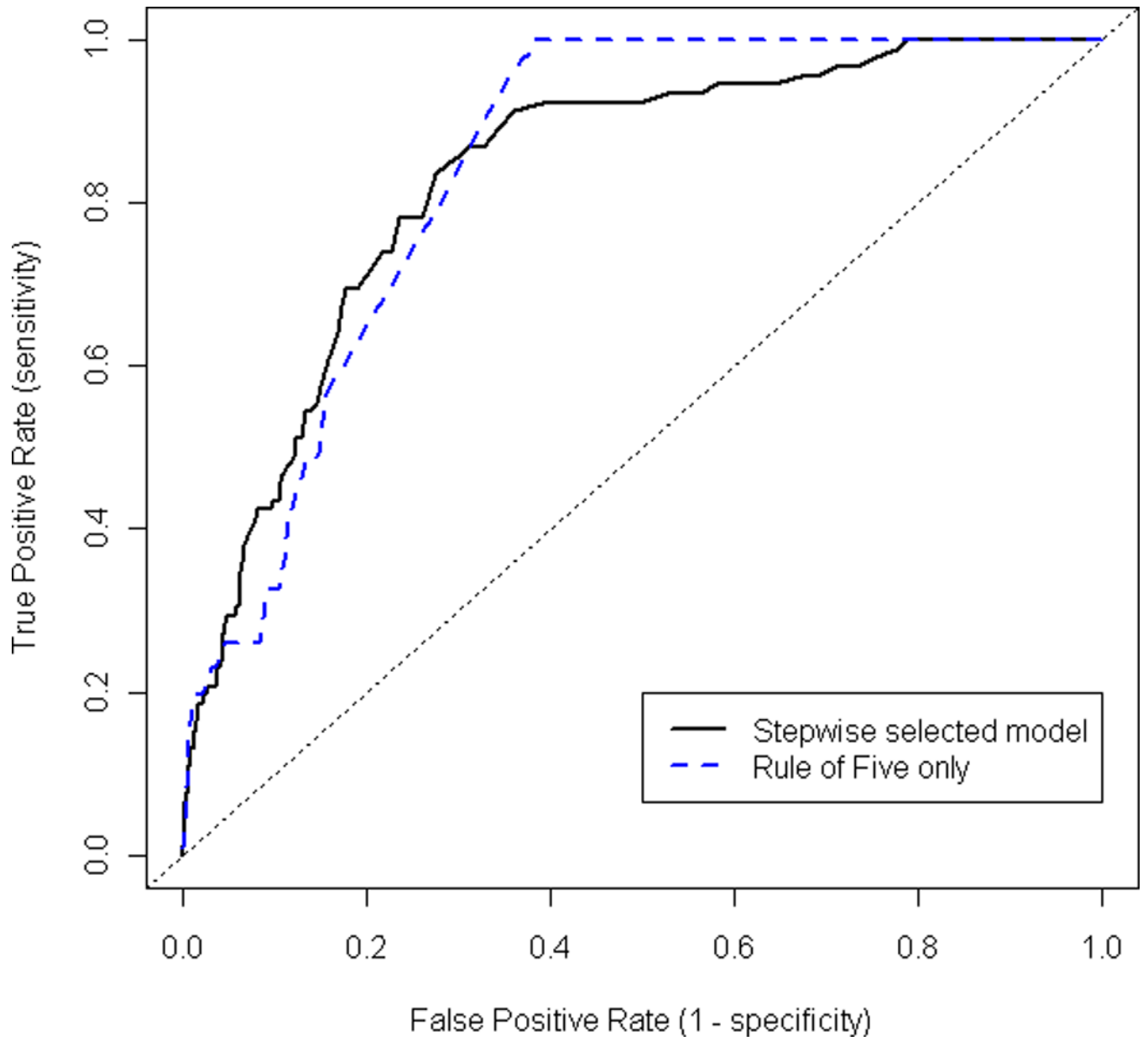


Figure 3.

Table 1

Best-fitting model for the selection of a threshold for the MRI Rule

Covariate ^a	Estimate	Std. Error	p-value	Odds Ratio (95% CI)
Intercept	-3.50	0.41	<0.001	
MRI Rule^b	1.14	0.31	<0.001	3.12 (1.70, 5.71)
Time	1.55	0.79	0.076	
Time²	-0.98	0.35	0.022	

^aTime is expressed in years.

^bA threshold of $\theta = 5$ lesions over the baseline CEL count provided the best fit.

Table 2Results of the stepwise model selection procedure^a

Covariate ^b	Estimate	Std. Error	p-value	AICc
Intercept	-3.86	0.166	<0.001	738.7
Number of CEL on scan	0.08	0.017	<0.001	714.7
Time²	-1.06	0.44	0.016	708.4
Time	1.67	0.89	0.061	706.7

^a Covariates are listed in the order in which they entered the model.^b Time is expressed in years.