

Personalized activity index, a new safety monitoring tool for multiple sclerosis clinical trials

Yinshan Zhao, Yumi Kondo, Anthony Traboulsee, David KB Li, Andrew Riddehough and A John Petkau

Multiple Sclerosis Journal –
Experimental, Translational
and Clinical

1: 1–15

DOI: 10.1177/
2055217315577829

© The Author(s), 2015.
Reprints and permissions:
[http://www.sagepub.co.uk/
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)

Abstract

Background: An abnormal increase of contrast-enhancing lesion (CEL) counts on frequent MRIs is interpreted as a signal of potential worsening in multiple sclerosis (MS) clinical trials. We demonstrate the utility of the MR personalized activity index (MR-pax) to identify such increases.

Methods: We analyzed a previous Phase II study in relapsing patients ($n = 167$) with MRIs at screening, baseline and months 1–6. We performed five consecutive reviews at 90-day intervals. At each review, we evaluate the MR-pax for each patient and also identify those who meet the rule-of-five (an ad-hoc guideline currently in use). To evaluate its clinical relevance, we assess the relation between having a small MR-pax (≤ 0.05 ; indicating an unexpected CEL increase) and relapse status in the 12 weeks post-review.

Results: Of the 399 patient reviews, 35 cases met the rule-of-five; 35 had an MR-pax ≤ 0.05 ; 18 met both criteria. The proportions experiencing clinical relapse are 63% among those meeting the rule-of-five, 61% among those with MR-pax ≤ 0.05 , and 83% for those meeting both criteria, more than double the rate of those meeting neither criterion (40%).

Conclusion: A guideline combining this new personalized index and the existing threshold-based criterion is able to better identify patients with a higher risk of experiencing relapses.

Keywords: Contrast-enhancing lesions, MRI, multiple sclerosis, safety monitoring in clinical trials

Date received: 22 October 2014; accepted: 30 January 2015

Introduction

Safety monitoring in clinical trials, both on the group and individual level, is of paramount importance. As safety outcomes can often be heterogeneous across participants of multiple sclerosis (MS) studies, personalized monitoring tools are desirable. In MS studies, especially Phase II trials with frequent magnetic resonance imaging (MRI) scanning, sudden increases of contrast-enhancing lesions (CELs) have been used by data safety monitoring boards (DSMBs) as a signal of potential disease worsening. Current DSMB guidelines to identify patients with increased CEL activity are ad-hoc in nature. One such criterion is the presence of ≥ 5 CELs above the baseline level on a follow-up scan (rule-of-five).¹ Previous studies showed that meeting the rule-of-five is associated with an elevated risk of a clinical relapse within a

month,^{1,2} confirming the merit of using CEL information to monitor patient safety in MS studies.

However, the rule-of-five approach relies on a predetermined threshold that does not account for the variability across individual patients and study cohorts, and does not use all available CEL information.^{2–4} To overcome these limitations, we developed a probability-based index.⁵ At each DSMB review, for every patient, this new procedure evaluates as an index the likelihood of observing CEL counts as large as those observed on the patient's recent scans given the patient's CEL data from previously reviewed scans. A small value of this personalized activity index (MR-pax) suggests that the observed count is unexpectedly large relative to the activity observed on previous scans, thus signaling a

Correspondence to:
Yinshan Zhao
MS/MRI Research Group,
University of British
Columbia, S195, 2211
Wesbrook Mall, Vancouver,
BC, V6T 2B5, Canada
[yinshan@
msmri.medicine.ubc.ca](mailto:yinshan@msmri.medicine.ubc.ca)

Yumi Kondo
Department of Statistics,
University of British
Columbia, Vancouver,
Canada

Anthony Traboulsee
Department of Medicine,
University of British
Columbia, Vancouver,
Canada

David KB Li
Department of Radiology,
University of British
Columbia, Vancouver,
Canada

Yinshan Zhao
Andrew Riddehough
MS/MRI Research Group,
University of British
Columbia, Vancouver,
Canada

A John Petkau
Department of Statistics,
University of British
Columbia, Vancouver,
Canada



possible change of the underlying disease activity level. In this paper, we demonstrate the utility of this new procedure and compare this probability-based approach to the threshold-based rule-of-five.

Patients and methods

Patients

The patients are from the Phase II study of lenercept, a recombinant tumour necrosis factor receptor p55 immunoglobulin fusion protein, in MS,⁶ previously examined by Riddell et al.² The cohort includes 167 patients between the ages of 19–51 years and with Expanded Disability Status Scale scores <6. The patients were diagnosed with clinically definite relapsing–remitting (83%) or secondary progressive MS (17%). All patients had at least two relapses in the preceding two years. Patients were randomly assigned to one of four treatment arms: 100 mg ($n = 40$), 50 mg ($n = 40$), 10 mg ($n = 44$) dose of lenercept, or placebo ($n = 43$).

MRI visits were scheduled at screening, at baseline and then on a four-week basis until week 24. Clinical assessments took place approximately at the same time as the MRI visits up to week 24 and on a 12-week basis thereafter until week 48. Patients were enrolled in the study over a period of about nine months. The study was terminated after 24 weeks of double-blind treatment because of a significant increase in the relapse rate among the 50 mg and 100 mg groups.⁶

The study was approved by the UBC Clinical Research Ethical Board. All patients gave written informed consent for the original study.

Evaluation of MR-pax

In a typical Phase II MS clinical trial setting, patients are enrolled in a staggered manner and are followed by MRI on a monthly basis, while the DSMB reviews the cumulative CEL data at scheduled intervals. We treated the study as if it was ongoing and monitored by the DSMB with the following schedule: The first review took place 120 days after the first patient's screening scan and subsequent reviews took place every 90 days until all scans were completed. At each review, scans will be referred to as either 'previous' or 'new'. A *previous scan* is either a scan performed before the treatment initiation (screening or baseline) or a follow-up scan that has been reviewed previously by the DSMB. A *new scan* is a follow-up scan that has not been reviewed previously.

The index relies on a mixed-effects negative binomial regression model to describe CEL counts on both the previous and new scans. Details on how the model was developed can be found in Zhao et al.⁵ In this model, we assume that the patient-specific random effects (a random intercept), representing the activity levels of individual patients, are independent. In the current analysis, we assume that these random effects follow a gamma distribution and given the activity level of a patient, the monthly CEL counts of the same patient are independent and follow a negative binomial distribution. The model enables us to predict the distribution of a patient's total CEL count on the new scans given his or her CEL counts on previous scans. The incorporation of patient-specific random effects implies this distribution will differ from patient to patient. By comparing the observed total CEL count (y) on the patient's new scans to this predicted distribution, we obtain the following conditional probability as our index, MR-pax: the chance to develop y or more CELs on the new scans given the CEL counts on the previous scans (x), i.e.:

$$\Pr(Y \geq y | X = x) = 1 - \sum_{t < y} \Pr(Y = t | X = x).$$

MR-pax takes values between 0 and 1, with a small value indicating more of an unexpected increase. For patients without CELs on their new scans, the MR-pax value is always 1. Table 1 illustrates the interpretation of MR-pax and Appendix e-1 provides a more detailed explanation.

The model is fitted by maximum likelihood to the available data at each review. As the overall activity level within each treatment group in a Phase II trial often changes over time, scanning time is included as a categorical covariate corresponding to three periods (fixed effects): pre-study (baseline and screening), months 1–3 and months 4–6. The three treated groups are considered as a single group. The pre-study mean level was assumed to be the same for all patients, whereas the mean levels during the second and third periods are allowed to differ for the treated and placebo patients. However, for monitoring purposes, all patients were treated as placebo patients when evaluating the indices, i.e. the evaluation was carried out only based on the estimated mean level of the placebo group. MR-pax can then be interpreted as the likelihood to observe such an increase under no active treatment. The implementation of this procedure is carried out using a freely available package lmeNB⁷ that we developed under

Table 1. Interpretation of MR-pax with illustrative cases.

MR-pax ^a	Illustrative examples ^b			Interpretation of MR-pax values and possible recommendations by the DSMB.
	Case	CELs on previous scans ^c	CELs on new scans ^d	
0.5–1	A	1/0	1/0	No extreme CEL activity on the new scans.
	B	5/0/10/5	1/0/10	
0.25–0.5	C	1/0/0/0	0/1/1	The CEL activity remains in the range expected for placebo patients.
	D	3/3	1/8/7	
0.1–0.25	E	0/0	1/1	The observed pattern of CEL activity is only slightly unexpected.
	F	3/1	5/3/6	
0.05–0.1	G	0/0	2/1/1	The observed pattern of CEL activity is somewhat unexpected. Closer future monitoring of these patients might be recommended.
	H	5/1	8/10	
0.05–0.01	I	0/0	3/3/2	The observed pattern is somewhat extreme. The DSMB might express concerns regarding the safety of these patients after considering other available data.
	J	4/8	25/30/10	
0.001–0.01	K	0/0/0/0	2/2/1	The observed pattern is fairly extreme. The DSMB might recommend withdrawal of the patient from the trial after considering other available data.
	L	4/8	25/30	
<0.001	M	0/0	10/15	The observed pattern is very extreme. If multiple patients have MR-pax values in this range, the DSMB might recommend an interim safety analysis or early termination of the trial after considering other available data.
	N	4/8	50/30	

^aThe MR-pax value, ranging from 0 to 1, represents the fraction of placebo patients with the same level of CEL activity on the previous scans who are expected to have at least that many CELs on the new scans. The ranges specified in this column are suggested potential ranges to correspond to the interpretation of the MR-pax values given in the last column.

^bThe MR-pax values of the example patients are computed based on the model fitted to the data of the Phase II study of lenercept at the final review.

^cA previous scan is either a scan performed before the treatment initiation (screening or baseline) or a follow-up scan that has been reviewed previously by the DSMB. The values $x_1/x_2/x_3$ represent the CEL counts on a patient's previous scans, i.e. three previous scans with x_1 , x_2 , and x_3 CELs, respectively.

^dA new scan is a follow-up scan that has not been reviewed previously. The values y_1/y_2 represent the CEL counts on a patient's new scans, i.e. two new-scans with y_1 and y_2 CELs, respectively.

MR-pax: magnetic resonance personalized activity index; CEL: contrast-enhancing lesion; DSMB: data safety monitoring board.

R, an open source environment for statistical computing and graphics.⁸

Relationship to clinical relapse

To demonstrate the clinical relevance of the CEL increases identified by our procedure, we investigated the relation between relapse status and having an extreme MR-pax value. A patient is **relapse free** if he or she is free from both an ongoing and any new relapses during the 12-week period immediately following the last new scan. As

DSMB meetings often take place every three to six months in Phase II trials, the 12-week window is chosen to reflect such a schedule. We vary the cut-off for an 'extreme' MR-pax and compute the relapse rates (i.e. the proportion of patient-reviews that are not post-review relapse free) for those below and above the cut-off.

To compare with the rule-of-five, the relapse rates are also computed for the following four categories of patient reviews: (1) not meeting the rule-of-five

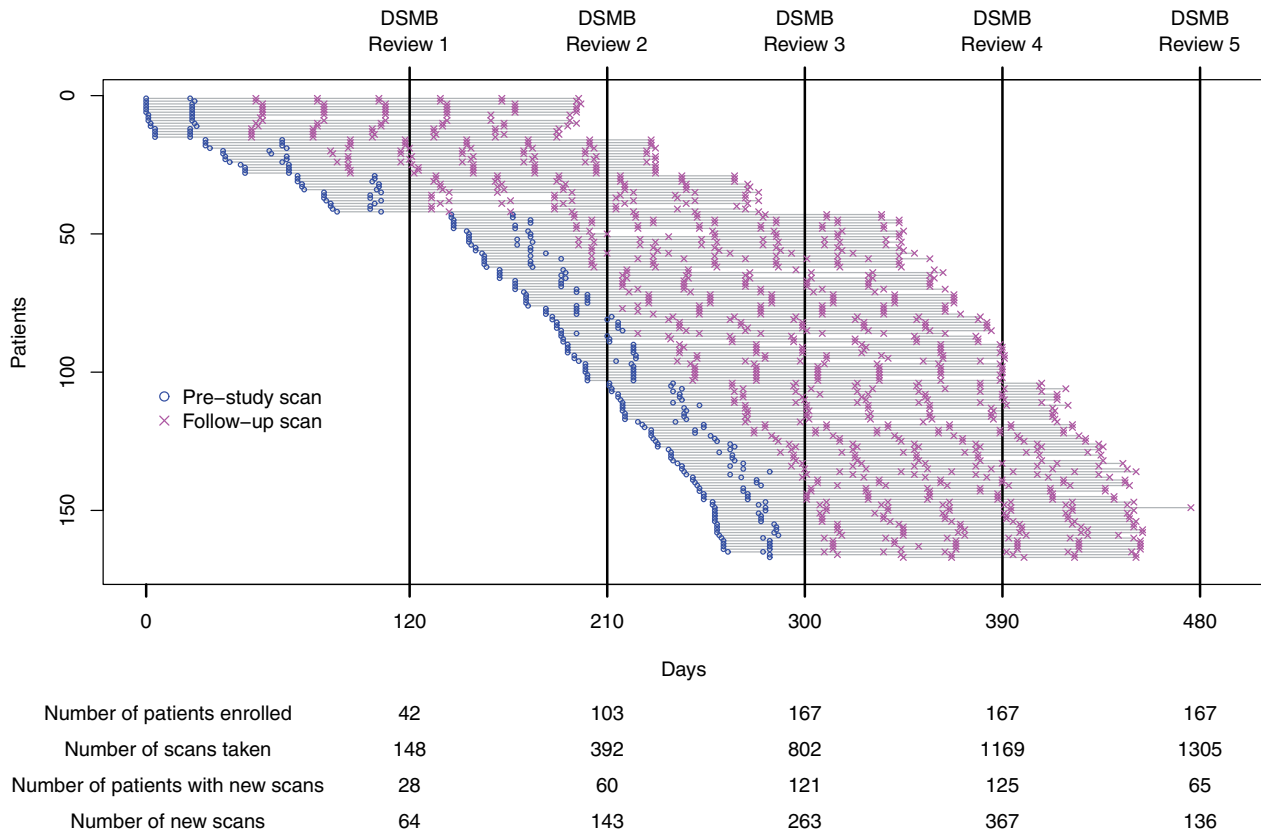


Figure 1. Timeline of patient recruitment, MRI scanning and DSMB reviews. (Each row represents a patient. Lines are disconnected at missing scans.)

MRI: magnetic resonance imaging; DSMB: data safety monitoring boards.

and $MR\text{-}pax > \text{cut-off}$, (2) not meeting the rule-of-five but $MR\text{-}pax \leq \text{cut-off}$, (3) meeting the rule-of-five but $MR\text{-}pax > \text{cut-off}$, and (4) meeting both criteria. A logistic regression with patient-specific random intercepts is used to compare the relapse risk of these four categories at selected cut-offs for $MR\text{-}pax$. As our initial analysis indicates that the overall relapse rate decreased after week 24, whether the post-review period was beyond week 24 was also included as a covariate. This analysis was performed using the R package lme4.⁹

Results

The patient characteristics have been reported previously.^{2,6} With the assumed review schedule, there were five reviews and a total of 399 person reviews. Except for one patient who was lost to follow-up after one review, all patients had two or three reviews. Figure 1 provides the timeline of patient recruitment, MRI scanning and reviews.

The estimated mean levels for the placebo group at each review are presented in Figure e-3. As an

example, Figure e-4 shows the output provided by lmeNB at Review 2, which includes a ranking of the patients according to their $MR\text{-}pax$ values. Figure 2 shows the $MR\text{-}pax$ values that are < 0.25 and their corresponding ranks at each review. The treated patients were predominant among the cases with extreme $MR\text{-}pax$ values at all reviews. For example, at Review 2, new scans from 60 patients were reviewed (15, 14, 17 and 14 from the 100 mg, 50 mg, 10 mg and placebo groups, respectively). Six patients, all treated, had an $MR\text{-}pax \leq 0.05$ (less than a 1-in-20 chance of observing a total CEL count as extreme or more extreme for a placebo patient with the same CEL counts on the previous scans). The most extreme case from the placebo group was ranked only 12th ($MR\text{-}pax = 0.17$). Seven treated patients and no placebo patients met the rule-of-five at this review; four of the seven patients who met the rule-of-five had an $MR\text{-}pax \leq 0.05$ while the $MR\text{-}pax$ values of the other three were 0.07, 0.14 and 0.30 and ranked 7th, 8th and 19th, respectively.

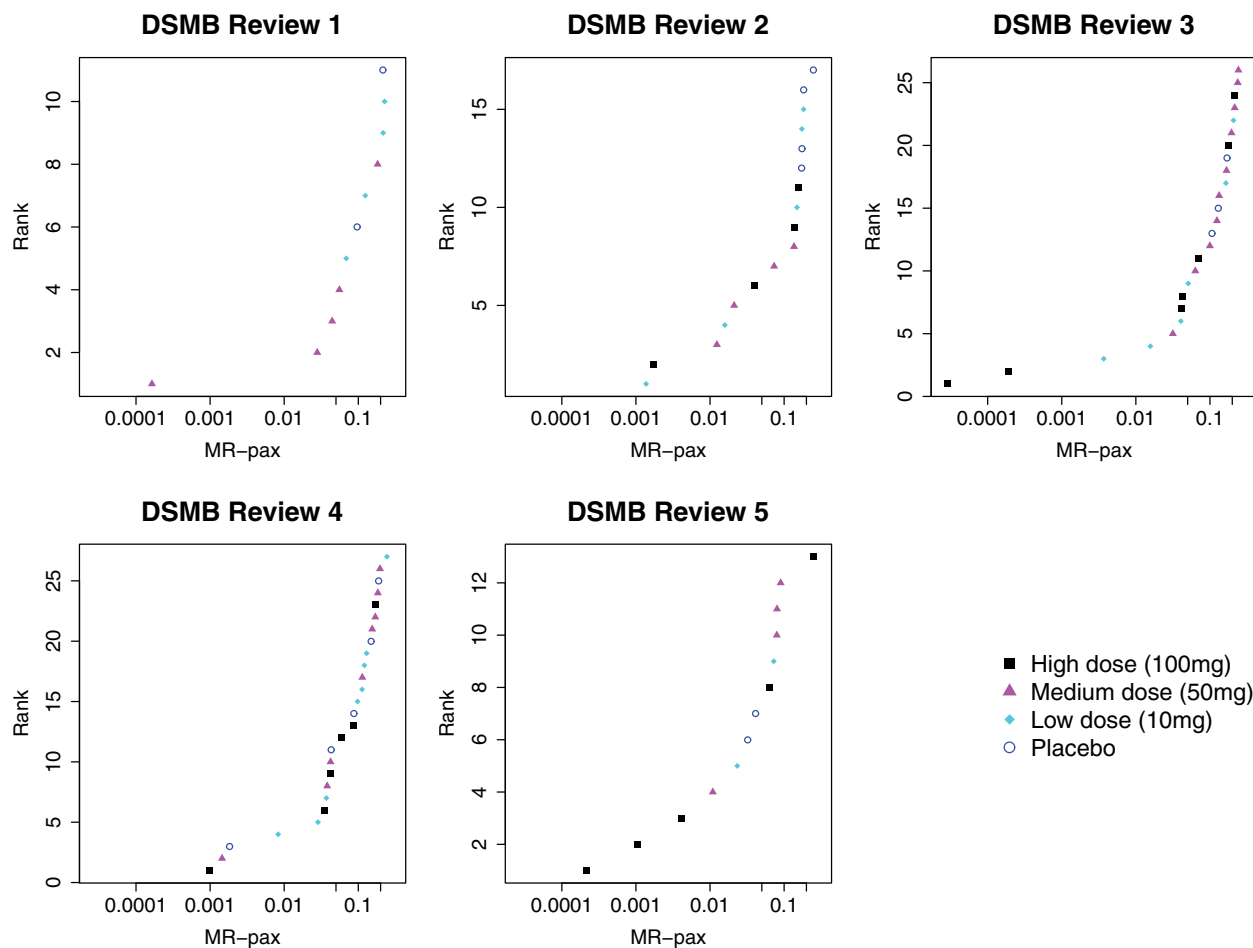


Figure 2. Cases having an MR-pax <0.25 at each DSMB review.
MR-pax: magnetic resonance personalized activity index; DSMB: data safety monitoring board.

Figure 3 shows the histograms of the MR-pax values, pooled across all reviews, by treatment group. The values of the placebo group are more evenly distributed between 0 to <1 , whereas greater proportions of small MR-pax values are observed in the treated groups, especially in the 100 mg group, indicating more cases with unusual CEL increases.

Overall, 35 patient reviews (9%) had an MR-pax ≤ 0.05 and 35 patient reviews met the rule-of-five; 18 met both criteria. Thirty-one patients (19%) met this index criterion at least once over the study (ten, eight, nine and four from the 100 mg, 50 mg, 10 mg and placebo group, respectively). In contrast, 23 patients met the rule-of-five at least once (seven, eight, six and two from the respective treatment groups). The CEL counts and MR-pax values of six illustrative patients are listed in Table 2. The MR-pax approach and the rule-of-five tend to agree

on cases exhibiting a clear increasing pattern (e.g. Patients 1 and 2 in Table 2). Patients with few lesions on the previous scans tend to have low activity on the new scans and are unlikely to meet the threshold of five. The MR-pax procedure is sensitive in detecting changes in such patients (e.g. Patient 3 in Table 2). Patients identified by the rule-of-five but not considered as extreme by their MR-pax values tend to have active previous scans. For example, Patient 6 had 12 and five CELs on the screening and baseline scans, respectively; at Review 1 this patient had just one new scan on which 11 CELs were observed. Although the patient met the rule-of-five, the MR-pax value (0.30) indicates that, considering both screening and baseline scans, the chance to observe ≥ 11 CELs on a new scan for such a placebo patient is about 30%, not highly unusual. See Zhao *et al.*⁵ for a more detailed comparisons of these two approaches.

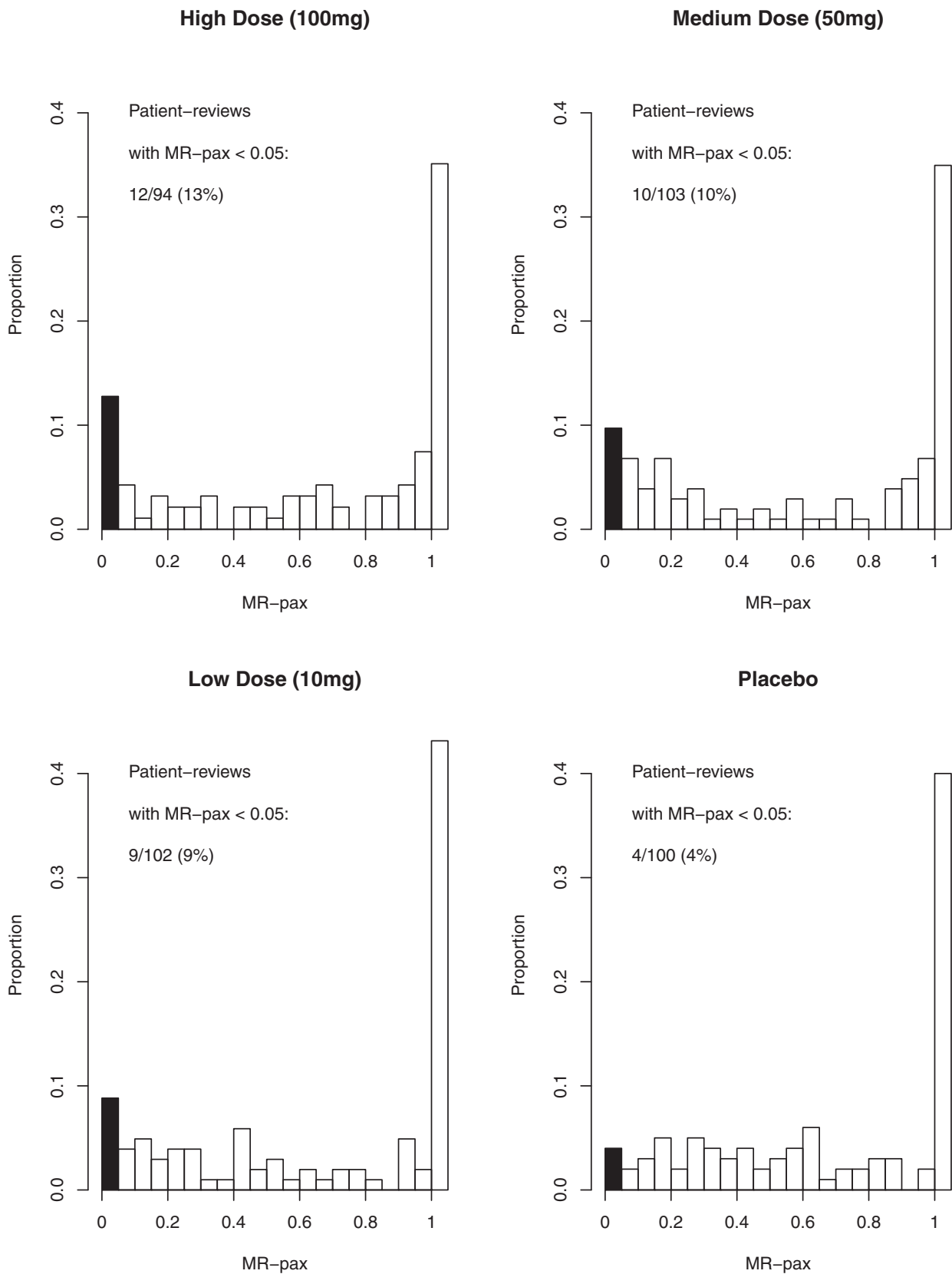


Figure 3. Distribution of the MR-pax values by treatment groups (all DSMB reviews combined; each patient may contribute more than one patient review. The black bar represents the proportion of patient reviews with MR-pax < 0.05.). MR-pax: magnetic resonance personalized activity index; DSMB: data safety monitoring board.

Table 2. Contrast-enhancing lesion counts (CELs) and MR-pax values for six selected patients at their successive reviews. (The times of an individual patient's reviews are determined by their study entry time as illustrated in Figure 1.).

Illustrative patient	Pre-study ^a CELs	CELs on new scans MR-pax (95% confidence interval)		
		Patient review 1	Patient review 2	Patient review 3
1	2/7	3/NA/40 ^b 0.0017 (9×10^{-5} –0.030)	49/NA/24 ^b 0.00019 (6×10^{-6} –0.0063)	
2	1/2	0/0/1 0.95 (0.91–0.97)	2/10/3 ^b 0.00099 (0.00023–0.0042)	
3	0/0	2/5 ^b 0.0037 (0.0014–0.0095)	1/0/0/0 0.99 (0.98–0.99)	
4	0/0	0/0 1 –	1/1/1 0.021 (0.0067–0.065)	2 0.063 (0.036–0.11)
5	2/0	3 0.18 (0.087–0.34)	NA/5/1 ^b 0.17 (0.10–0.27)	3/3 0.31 (0.21–0.43)
6	12/5	11 ^b 0.30 (0.10–0.61)	2/2/1 >0.99 (0.99–1.00)	1/1 0.99 (0.97–1.00)

^aScreening and baseline scans. ^bReviews that also met the rule-of-five. MR-pax: magnetic resonance personalized activity index; NA: missing scan.

The relapse status cannot be determined for three patient reviews because of lack of clinical follow-up. For the remaining patient reviews, the overall rate of not being relapse free in the 12-week period following the last new scan was 165/396 (42%). The rate for those who met the rule-of-five is 22/35 (63%). Figure 4 shows the post-review relapse rate for patients with an MR-pax value below different cut-offs (black solid line) ranging from 0.001 to 0.25. The relapse rate of those who met the MR-pax criterion was always much higher than the overall rate regardless of the cut-off value; it peaked at 73% around the cut-off of 0.02 and steadily declined for cut-off values larger than 0.15. The figure also shows the relapse rates for the subgroup of patients who met both the MR-pax criterion and the rule-of-five (grey solid line), which were in general higher than those who met the rule-of-five and peaked at 84% around the cut-off of 0.06.

Table 3 reports the observed rates and odds ratios (ORs) of relapse at the MR-pax cut-offs of 0.05 and 0.1. In both cases, the rate of clinical relapse among those patients who met both the MR-pax

cut-off and the rule-of-five is double the rate among those meeting neither criterion. In addition, patients who did not meet the rule-of-five and had an MR-pax between 0.05–0.10 were those experiencing moderate increases in CEL activity; they also had a rather high relapse rate (11/14, 79%).

In the fit of a mixed-effects logistic regression allowing an interaction between meeting the rule-of-five and having an MR-pax ≤ 0.05 , the interaction term approached significance ($p = 0.08$). Compared to the group which met neither criterion, the OR of relapse for the group meeting both criteria is 9.58 (95% confidence interval (CI): 2.25–55.4); the ORs for those who met only one criterion are close to one.

Discussion

MRI lesion activity in MS patients is known to be widely variable both between patients and studies, presenting a considerable challenge to determine whether the observed increases in individual patients are outside the normal range. Existing guidelines, such as the rule-of-five, are easy to implement, but may not fully meet this challenge as they do not recognize the

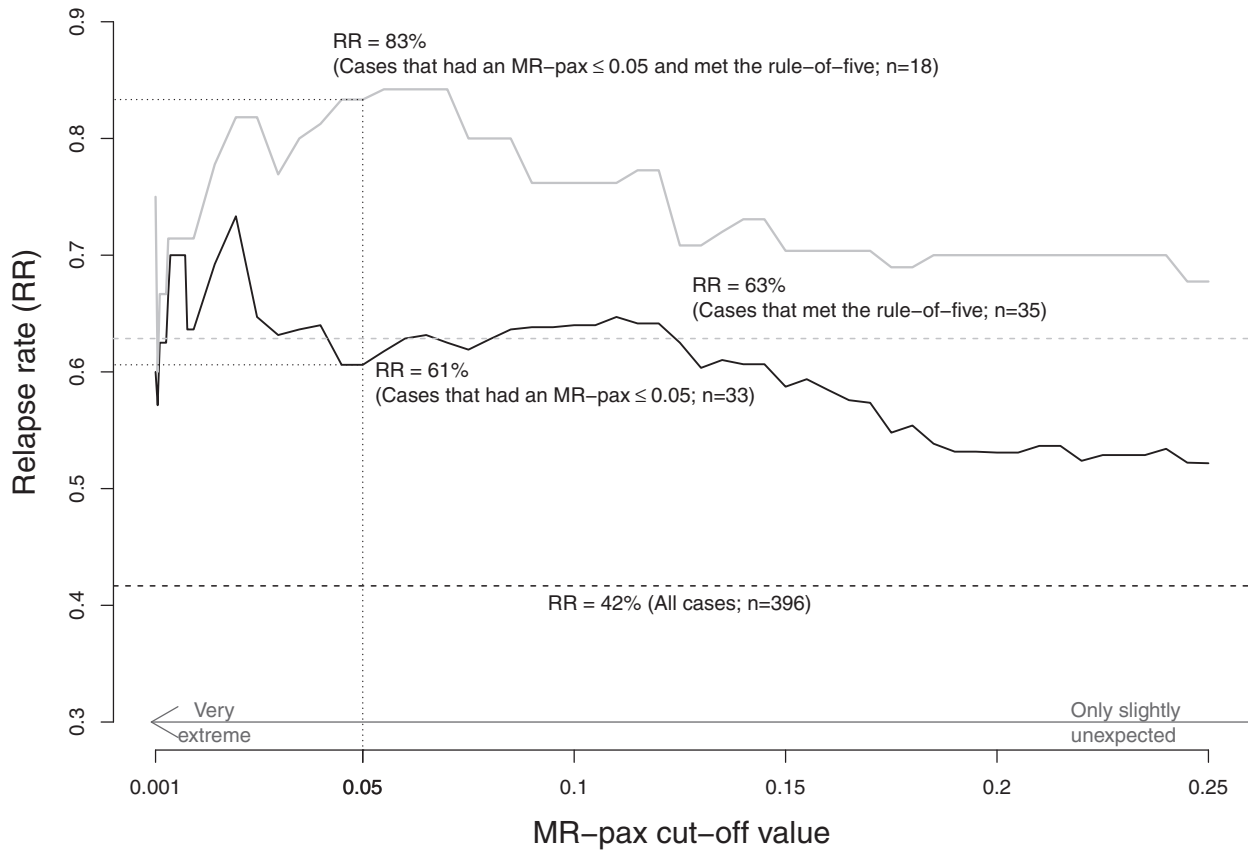


Figure 4. The post-review relapse rate for those who met the MR-pax criterion and the subgroup who also met the rule-of-five. **Black solid line:** relapse rate for patient reviews that met the MR-pax criterion as the cut-off value varies between 0.001 (extremely unlikely increase in CEL activity) to 0.25 (only slightly unlikely). The number of patient reviews that met the MR-pax criterion was five at the cut-off of 0.001 and increased to 92 at the cut-off of 0.25. **Grey solid line:** relapse rate for patient reviews that met both the MR-pax criterion and the rule-of-five. The number of patient-reviews that met both criteria was four at the cut-off of 0.001 and 31 at the cut-off of 0.25. **Black dashed line:** relapse rate for all patient-reviews. **Grey dashed line:** relapse rate for patient reviews that met the rule-of-five. MR-pax: magnetic resonance personalized activity index; CEL: contrast-enhancing lesion.

Table 3. Relapse rate in the following 12-week period by whether or not meeting the MR-pax criterion and the rule-of-five.

Meet rule-of-five	MR-pax ≤ 0.05	Proportion of patient reviews with relapse (%)	Odds ratio ^a (95% confidence interval)
(a) MR-pax cut-off at 0.05			
No	No	138/346 (40)	1
	Yes	5/15 (33)	1.11 (0.29–4.12)
Yes	No	7/17 (41)	1.10 (0.26–4.49)
	Yes	15/18 (83)	9.58 (2.25–55.4)
(b) MR-pax cut-off at 0.1			
No	No	127/332 (38)	1
	Yes	16/29 (55)	2.82 (1.03–8.22)
Yes	No	6/14 (43)	1.27 (0.31–5.16)
	Yes	16/21 (76)	6.74 (1.83–29.7)

^aBase on logistic regression with patient-specific random intercepts. MR-pax: magnetic resonance personalized activity index.

heterogeneity across patients. We developed a probability-based approach to evaluate the degree of abnormality of each patient according to his or her own data from previous scans.⁵ Our procedure relies on a mixed-effects negative binomial model that allows a different mean level for each patient and allows the within-patient variability to vary according to the patient's mean level. The model is fitted to data collected from the study under review and updated with the new available data at each review. Therefore, MR-pax is patient and cohort specific.

The MR-pax procedure is not restricted by a study's duration, scanning frequency, or size. Nor does it require that all patients have the same number of scans; patients can be at different points in follow-up or have some missing scans. These features ensure its usefulness in real-time safety monitoring.

We developed a package to implement this procedure⁷ that is freely available at <http://CRAN.R-project.org/package=lmeNB>. In addition to the model considered in this paper, the package offers a range of modelling choices that are considered by Zhao *et al.*⁵ In practice, a data center can implement our procedure using this package. The resulting patient list sorted by their MR-pax values (as shown in Figure e-4) can be provided to the DSMB, greatly simplifying the DSMB's task of identifying the extreme cases and reviewing their CEL activities.

In MS clinical trials, DSMBs are normally blinded to the treatment assignment. To maintain blinding, Zhao *et al.*⁵ previously proposed to analyze all patients as one group. Although this approach has the convenience of easy implementation, in practice it could lead to bias in the presence of a treatment effect. In this paper, we fitted the model allowing the group mean to differ between the treatment arms during the follow-up, but MR-pax was evaluated based on the placebo mean regardless of the treatment assignments. MR-pax can then be interpreted as the likelihood to observe such an increase if the patient had not received the treatment. This modification avoids the potential bias, yet does not unblind the DSMB as long as the evaluation of MR-pax is performed by an independent (unblinded) statistician. See Appendix e-3 for a comparison of these two approaches.

To eliminate the potential influence of the treatment, the DSMB may wish to include only pre-treatment scans as previous scans. Such strategies are easy to implement using our R package as it allows the user to customize the 'previous' and 'new' scans.

Threshold-based procedures, such as the rule-of-five, recognize cases with a substantial increase from the baseline level. However, a fixed threshold can be too high for patients who had no or few baseline lesions, and not high enough for those with a large baseline count. A guideline combining MR-pax and the rule-of-five can better identify patients with large increases and such patients tend to have the highest risk of experiencing post-review relapses. We also observed that patients with a moderately extreme MR-pax value (0.05–0.10) often did not meet the rule-of-five, but were also prone to have post-review relapses. Our results support the finding that a period of high CEL activities is associated with clinical worsening as suggested by previous studies.^{10,11} On the other hand, disease worsening might not always be manifested in clinical relapses. Unusual increases identified by MR-pax may be a precursor of other forms of worsening. Further validation of this procedure with larger cohorts will be useful.

MR-pax provides DSMBs with a rational basis to rank patients with different follow-up duration and activity levels. A DSMB may choose an MR-pax or rank threshold according to their level of safety concern. However, an extreme MR-pax is not meant to be an unequivocal indicator of clinical worsening; rather, it signals which patients might need more careful monitoring. As worsening may occur independent of CEL activity, it would be desirable to extend our method to monitor both clinical and MRI measures simultaneously.

MR-pax was developed to assess changes in individual patients, but it also has potential utility in group-based monitoring. In the original analysis of the cumulative number of newly active MRI lesions,⁶ a higher median count was observed in the 50 mg and 100 mg groups compared to the placebo; however, the differences were not statistically significant. Our MR-pax procedure identified more cases as extreme in all three treated arms compared to the placebo; this supports the clinical observation of unexpected toxicity with lenercept. It is possible that the treatment did not have the same adverse effect on all patients. In such circumstances, a group-based summary such as the median is not sensitive in reflecting such differences. Our procedure can be a sensitive tool to detect adverse treatment effects on CELs, especially when limited to a subset of patients.

To reliably estimate MR-pax, we rely on a statistical model that can accurately describe the pattern of longitudinal CEL counts. At the early reviews

when only a small number of patients are available, it is difficult to reliably fit the model and this limits the effectiveness of our procedure. We are currently extending our approach within the Bayesian framework and will implement this extension under R. This will enable the DSMB to incorporate prior information based on data from previous trials and based on their own expert knowledge, and thus enhance the performance of our MR-pax procedure, particularly in the early stages of a trial.

In this paper, we illustrated how a monitoring tool can be tailored to individual patients and its potential in safety monitoring on an individual level. The idea can be extended to other longitudinally collected safety outcomes, and, therefore, has broader utility in clinical trial monitoring.

Funding

This work was supported by the Multiple Sclerosis Society of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Milan & Maureen Ilich Foundation.

Conflict of interest

None declared.

Acknowledgements

The authors thank Dr Roger Tam for reviewing the manuscript.

References

1. Morgan CJ, Ranjan A, Aban IB, et al. The magnetic resonance imaging 'rule of five': Predicting the occurrence of relapse. *Mult Scler* 2013; 19: 1760–1764.
2. Riddell CA, Zhao Y, Li DK, et al. Evaluation of safety monitoring guidelines based on MRI lesion activity in multiple sclerosis. *Neurology* 2011; 77: 2089–2096.
3. Richert ND and Kryscio RJ. Does one guideline fit all? *Neurology* 2011; 77: 2080–2081.
4. Dong J, Zhao Y, Petkau AJ, et al. Further investigation of safety monitoring guidelines based on magnetic resonance imaging lesion activity in multiple sclerosis clinical trials. *Mult Scler* 2015; 21: 101–104.
5. Zhao Y, Li DK, Petkau AJ, et al. Detection of unusual increases in MRI lesion counts in individual multiple sclerosis patients. *J Am Stat Assoc* 2014; 109: 119–132.
6. The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. (1999) TNF neutralization in MS: Results of a randomized, placebo-controlled multicenter study. *Neurology* 1999; 53: 457–465.
7. Zhao Y and Kondo Y. lmeNB: Fit negative binomial mixed-effect regression model. R package version 1.2, <http://CRAN.R-project.org/package=lmeNB> (accessed 16 December 2014).
8. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/> (accessed 9 September 2013).
9. Bates D, Maechler M, Bolker B, et al. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0–5, <http://CRAN.R-project.org/package=lme4> (25 October 2013).
10. Grossman RI, Gonzalez-Scarano F, Atlas SW, et al. Multiple sclerosis: Gadolinium enhancement in MR imaging. *Radiology* 1986; 161: 721–725.
11. Smith ME, Stone LA, Albert PS, et al. Clinical worsening in multiple sclerosis is associated with increased frequency and area of gadopentetate dimeglumine-enhancing MRI lesions. *Ann Neurol* 1993; 33: 480–489.

Appendix e-1: Illustration of MR-pax with example cases

With our mixed-effects negative binomial model, we can predict the distribution of a patient’s total CEL count (Y) on the new scans given the CEL counts on their previous scans. For the model considered in our paper, this is the same as the distribution of Y given the **total** count on the previous scans (x), i.e., $\Pr(Y=y|X=x)$. This predictive distribution will differ depending on the total CEL count observed on the previous scans, the number of previous scans, and the number of new scans.

In Figure e-1, we display the predictive distributions of four cases that have three new scans but different numbers of previous scans (2 in the top row and 4 in the bottom row). (These predictive distributions were evaluated based on the model fitted to all the CEL data from the Phase II study of lenercept in MS.⁶ In

practice, the fitted model would vary from review to review.)

The top left panel of Figure e-1 presents the case with only one CEL on the two previous scans. This histogram predicts the distribution of the total CEL count on the next 3 (new) scans for placebo patients who had just one lesion on their first two scans. These patients are likely to have few CELs on the 3 new scans – about 24% will have no lesion activity, around 20% will have exactly one CEL, and roughly 5% are expected to have 11 or more CELs.

The top right panel illustrates the case where 5 CELs are observed on the first two scans. Patients with more CELs on the two previous scans are expected to have a higher lesion activity level as well as greater variability on the next three scans – only about 1% will have no lesion activity, a relatively

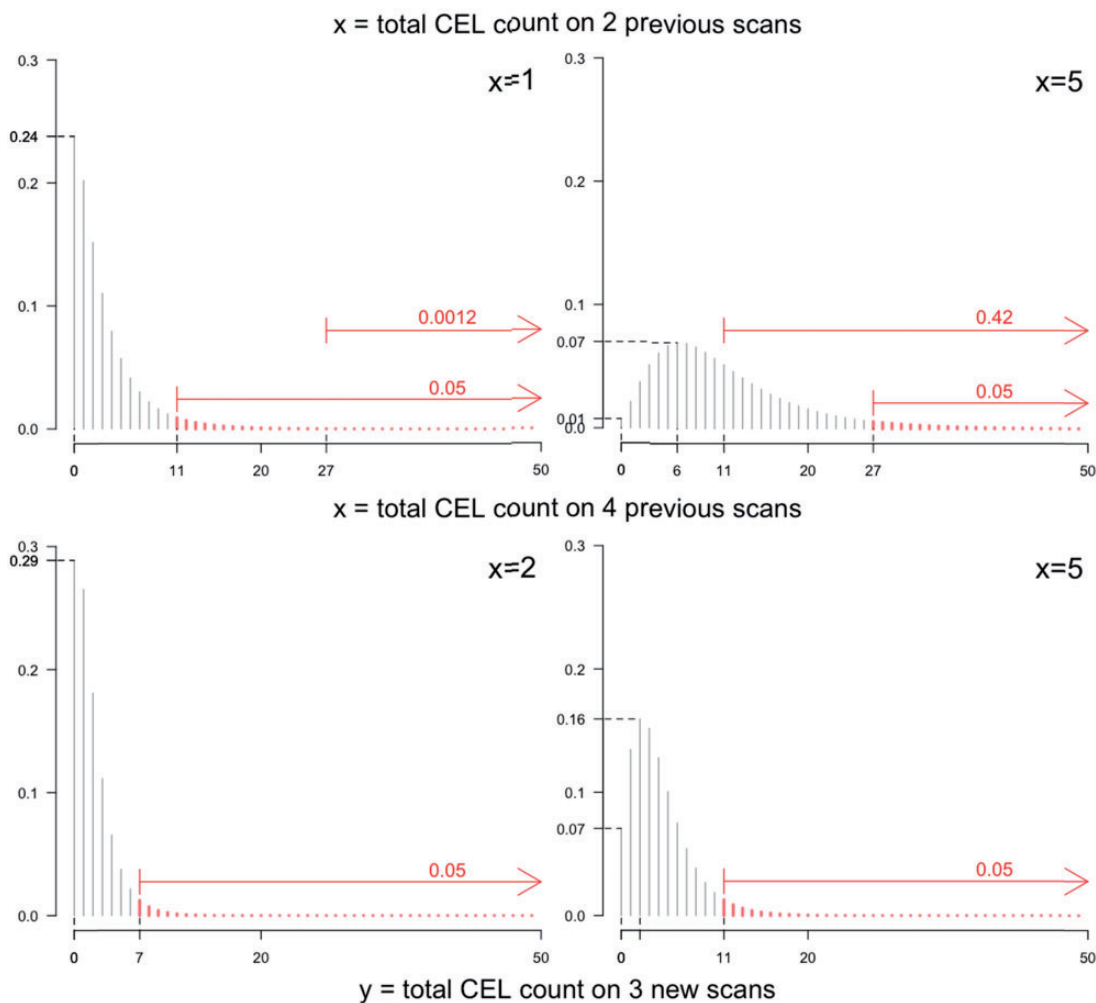


Figure e-1. The predictive distribution of the total CEL count on 3 new scans for four selected placebo cases.

large proportion will have 11 or more CELs (42%), and roughly 5% are expected to have 27 or more CELs which is highly unlikely for the previous case (0.12% or about one case per 1000 patients).

The bottom left panel presents the case with two CELs on four previous scans. These patients have the same average CEL count ($2/4 = 0.5$) on the previous scans as the patients described in top left panel ($1/2 = 0.5$). However, the two predictive distributions differ, as their numbers of previous scans are different. Compared to the previous case, these patients are likely to have even fewer CELs on the new scans – about 29% will have no lesion activity and only 5% would have 7 or more CEL counts.

The bottom right panel illustrates the case where 5 CELs are observed on four previous scans. These patients have the same total CEL count on their previous scans as those described in the top right panel, but are expected to have fewer CELs on the 3 new scans – 7% are expected to have no lesion activity and only 5% of patients are expected to have 12 or more CELs.

Based on the predictive distribution, we can evaluate the personalized lesion activity index, MR-pax:

$$\Pr(Y \geq y | X = x) = 1 - \sum_{t < y} \Pr(Y = t | X = x),$$

that corresponds to the area in the right tail of the distribution. When there is no CEL ($y=0$) on the new scans, the MR-pax value is always 1. As the observed total lesion count on the new scans (y) increases, the MR-pax value becomes smaller, indicating that it is less likely to observe such levels of activity among placebo patients with the same level of pre-scan activities. Therefore, MR-pax serves as a suitable measure for determining unusual increases in CEL activity.

Figure e-2 shows the MR-pax value as a function of the total CEL count on 3 new scans for selected pre-scan activity levels, demonstrating that the determination of “extremeness” is “personalized” according to the CEL information available from each patient. For example, for a placebo patient with no CEL activity on two previous scans (black), observing a total of ≥ 5 CELs on the 3 new scans is already somewhat extreme (MR-pax < 5%). On the other hand, for a placebo patient with one CEL on the two previous scans (orange), observing a total of 5 CELs on the 3 new scans is not considered outside the expected range because 22% of placebo patients with such pre-scan activity are expected to have ≥ 5

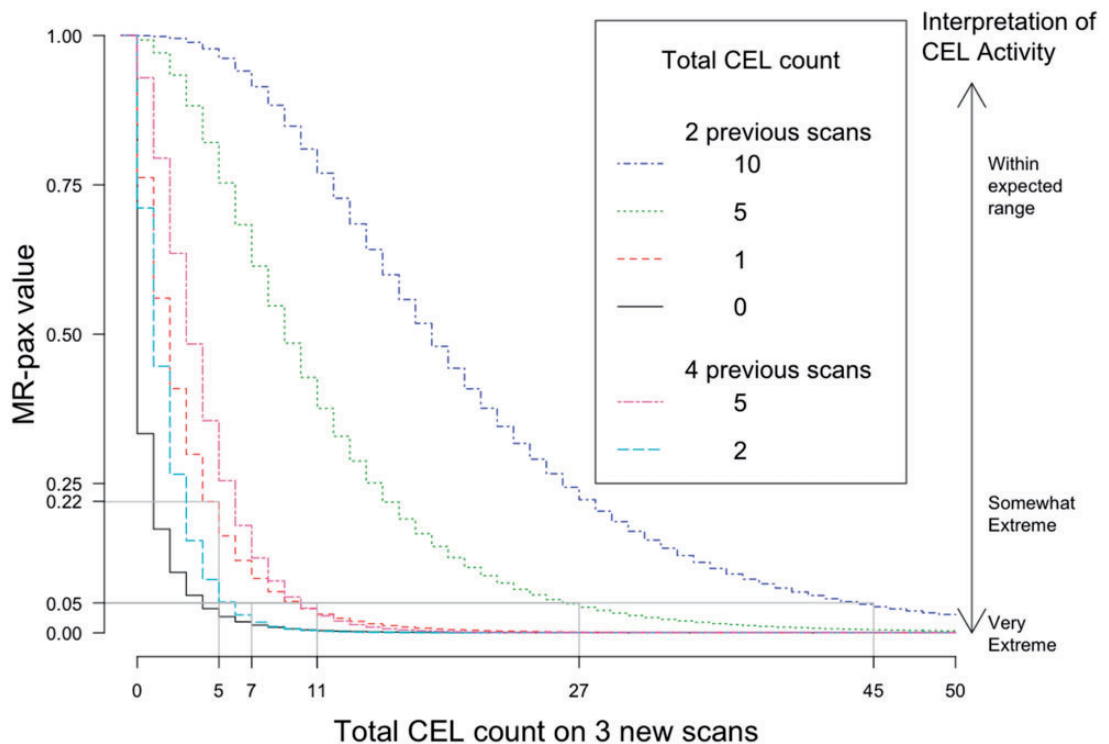


Figure e-2. MR-pax curves for selected levels of total CEL count on previous scans.

CELs. Furthermore, a placebo patient with a total of 10 CELs on two previous scans (blue) requires a much higher total CEL count on the 3 new scans to reach the same level of ‘extremeness’: at least 45 total CELs are required to yield a MR-pax <5%.

Figure e-2 also shows that MR-pax depends on the number of previous scans – patients with a total of 5 CELs on 4 previous scans (magenta) have a smaller MR-pax value for any non-zero total CEL count on the 3 new scans than patients with a total of 5 CELs on 2 previous scans (green). This indicates that the same CEL count on the 3 new scans is more “surprising” for those with a total of 5 CELs on 4 previous scans than for those with the same total CEL count on 2 previous scans.

While the rule-of-five only compares each single new scan to the baseline scan, the MR-pax is much more flexible – it can be calculated for any number of previous and new scans that are available. If a patient only has one new scan at the time of review, the index value is calculated based on this single new scan (and whatever previous scans are available for that patient).

Appendix e-2: Additional results

Figure e-3 presents the estimated mean levels for the placebo group at each DSMB review. Figure e-4 shows the output from the R package lmeNB at DSMB Review 2.

Appendix e-3: Comparison of the MR-pax estimates based on two models

We compare the two sets of MR-pax values:

1. Placebo: based on the placebo mean level estimated from a model allowing a treatment effect on the group mean. This index is considered in the main text.
2. Mixed: based on a model where all patients were treated as one group. This version of the index is considered by Zhao *et al.*⁵

As shown in Figure e-5, the differences between the two sets of indices are larger at the early reviews when the mean estimates are less reliable. The confidence intervals based on the ‘Placebo’ method are also much wider at the early reviews (results not shown).

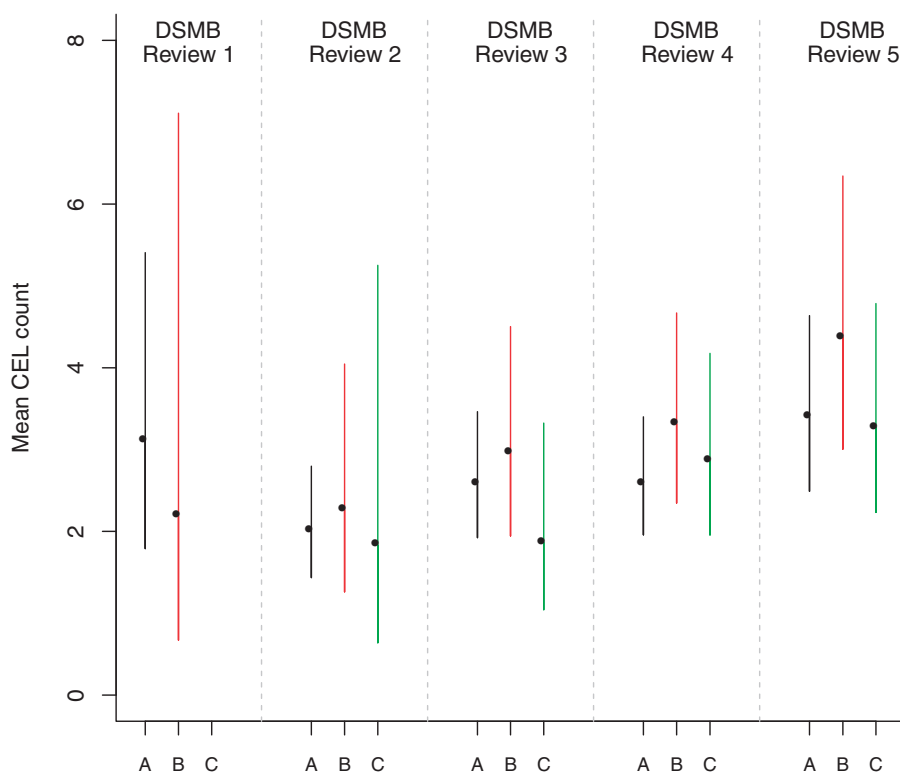


Figure e-3. Estimates and 95% confidence intervals of the mean contrast enhancing lesion (CEL) count for the placebo group in time periods A (pre-study, i.e., baseline and screening), B (months 1 – 3) and C (months 4 – 6) at each DSMB review.

Note: There were no months 4 – 6 scans available at DSMB Review 1.

Estimated conditional probability index and its 95 % CI.
The scalar function to summarize new scans: sum

	conditional.probablity	pre.scan	new.scan
090	0.001 (0.000,0.024)	0/0/0/0/0	1/2/2
061	0.002 (0.000,0.030)	2/7	3/40
026	0.012 (0.001,0.220)	7/9/14	4/14/31/32
054	0.016 (0.000,0.794)	5/8/8/9/9	8/20/15
057	0.021 (0.007,0.065)	0/0/0/0	1/1/1
016	0.039 (0.008,0.169)	7/3	15
095	0.073 (0.003,0.643)	5/10/26/25	27/16/24
025	0.136 (0.008,0.753)	19/15/25/22	25/27/18
062	0.138 (0.061,0.285)	1/0	4/0
053	0.149 (0.065,0.305)	0/0/0/0/0	0/0/1
039	0.158 (0.098,0.244)	0/0	1
006	0.172 (0.080,0.333)	0/0/1	2/0/2/0
003	0.173 (0.049,0.462)	0/0/0/1/0	1/1/0
089	0.173 (0.049,0.462)	1/0/0/0/0	0/1/1
038	0.183 (0.087,0.345)	1/1	3
101	0.183 (0.087,0.345)	2/0	3
009	0.248 (0.068,0.599)	2/6	3/8/7
059	0.283 (0.106,0.567)	1/4	3/5
041	0.297 (0.103,0.609)	12/5	11
024	0.298 (0.009,0.954)	13/13/12/3/13	12/10/9
029	0.318 (0.216,0.441)	0/0	1/0/0
032	0.318 (0.216,0.441)	0/0	0/1/0
093	0.324 (0.108,0.656)	3/0/1/1	1/0/4
084	0.394 (0.193,0.639)	4/2	4
001	0.414 (0.042,0.920)	4/7/0/0/3	5/2/1
055	0.429 (0.209,0.681)	0/0/0/0/1	0/0/1
015	0.439 (0.311,0.576)	0/1	1
040	0.439 (0.311,0.576)	1/0	1
085	0.447 (0.182,0.747)	7/11	9
096	0.519 (0.211,0.813)	4/4	3/5/4
031	0.592 (0.184,0.903)	9/14	13/7/8
060	0.627 (0.485,0.750)	0/1	1/0
099	0.648 (0.424,0.822)	3/2	2
005	0.725 (0.236,0.957)	7/11/10	4/5/7/9
008	0.739 (0.516,0.882)	2/1	0/0/3
004	0.772 (0.383,0.948)	5/6/1	1/5/1/3
030	0.920 (0.776,0.975)	4/0	1/1/0
052	0.923 (0.107,0.999)	6/8/15/4/5	3/3/3
097	0.975 (0.883,0.995)	5/1	0/1/1
027	0.989 (0.900,0.999)	2/4/3/1	0/0/1

Patients are ordered by decreasing conditional probability.
Patients with no pre-scans or no new scans are not reported.
Patients whose new scans are all zero are not reported as conditional probability of such patients are always one.

Figure e-4. Output from the R package lmeNB (version 1.2) for the data at DSMB Review 2
Patients 061, 057, 101 and 041 are also illustrated in Table 2, where they are labelled as patients 1, 4, 5 and 6, respectively.
As DSMB Review 2 was the first time patients 061, 101 and 041 were reviewed, their data appear under ‘Patient Review 1’ in Table 2. Patient 057 was reviewed previously at DSMB Review 1, so the data for this patient appears under ‘Patient Review 2’.

The number of patient-reviews with an MR-pax < 0.05 for each treatment group are comparable by the two methods (Table e-1). The relation between relapse rate and having a small MR-pax value is also similar based on these two methods (results not shown).

In conclusion, the ‘Placebo’ method reduces the impact of the potential treatment effects. The advantage of the ‘Mixed’ approach is that the fitted model is more stable at the early reviews and it does not require any knowledge of the treatment assignment. On the other hand, the two methods

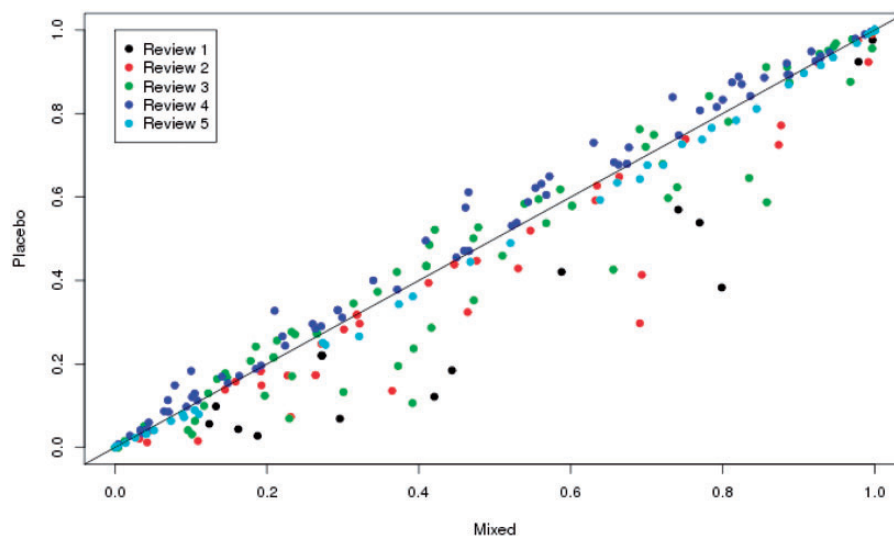


Figure e-5. Comparison of the ‘Placebo’ and ‘Mixed’ MR-pax values.

Table e-1. Number of patient-reviews with an MR-pax < 0.05 by treatment group.

Group	‘Placebo’	‘Mixed’
Placebo	3/100	4/100
Low Dose (10 mg)	9/102	9/102
Medium Dose (50 mg)	7/103	10/103
High Dose (100 mg)	12/94	12/94

tend to identify the same patients with the most extreme values in each review. The ranks based on the MR-pax values are fairly robust to the evaluation methods.

In principle, one could also compute MR-pax based on the mean level of the treated group, however, this may lead to an overly sensitive procedure in the

presence of a positive treatment effect and vice versa when there is a harmful effect. Furthermore, the resulting MR-pax does not have a clear interpretation since the MR-pax value then depends on the treatment effect in the current study. Therefore, from the safety point of view, it is more logical to use the placebo group to establish the reference.