3.1

Persistent T1 Hypointensity as an MRI Marker for Treatment Efficacy in Multiple Sclerosis

I.J. van den Elskamp
J. Lembcke, V. Dattola
K. Beckmann
C. Pohl
W. Hong
R. Sandbrink
K. Wagner
D. L. Knol
B.M.J Uitdehaag
F. Barkhof

Mult Scler. 2008;14(6):764-9.

ABSTRACT

Background: MRI is often used as primary outcome measure in phase II clinical trials in Multiple Sclerosis (MS). Since persistent T1 hypointense lesions are a surrogate parameter for axonal damage and demyelination, they may serve as a marker for monitoring the efficacy of neuroprotective drugs. At present, a power analysis using black hole (BH) evolution as primary outcome measure has not been performed. *Objective:* to assess the feasibility of using BH evolution on serial brain MR images as primary outcome measure in proof of concept studies in MS.

Methods: MRI-data obtained from 169 active, RRMS patients were analysed for BH evolution by determining the cumulative number of contrast enhancing lesions (CEL) evolving into a persistent black hole (PBH) after 3 months. With a parametric simulation procedure, based on a statistical distribution fitting the data, sample sizes were calculated.

Results: 21.2% of the total number of CELs observed during the study period evolved into a PBH. Ring enhancing lesions evolved most frequently into a PBH (59.4%), followed by lesions larger than 10 mm (57.4%) and Periventricular CELs (30.6%). The simulation procedure, based on the statistical Negative Binomial (NB) model resulted in a sample sizes between 200 subjects and 30 subjects per arm, for treatment effects ranging from 50% to 90% reduction of the number of CELs evolving into a PBH, respectively.

Conclusion*:* To perform a MRI monitored phase II clinical trial with a feasible sample size, using the evolution of CELs into PBHs as primary outcome parameter, a potent drug is required to obtain sufficient power.

INTRODUCTION

Serial brain Magnetic Resonance Imaging (MRI) is routinely used to monitor the efficacy of experimental treatments in Multiple Sclerosis (MS). To demonstrate a treatment effect of immunosuppressant agents, the MRI end point of choice often is the cumulative number of new contrast enhancing lesions (CEL). This is conceivable, since contrast enhancement reflects disruption of the blood brain barrier and is strongly associated with the local inflammatory proces [1,2,3]. However, when the aim of a treatment trial is to demonstrate a reduction of the amount of brain tissue damage, counting the number of new CELs is a less appropriate outcome parameter since it is a marker of inflammation rather than brain tissue damage, and might disregard the neuroprotective effects of a drug or treatment regime. An MRI endpoint more suitable to serve this purpose is the number of hypointense lesions on T1 weighted images, so called Black Holes (BH), which is a known MRI-pathological correlate of axonal loss and loss of myelin [4,5,6]. Unfortunately, active inflammation is known to decrease signal intensity due to increased free extracellular water, resulting in the majority of CELs appearing as a hypointense lesion on the corresponding unenhanced T1 image. Once contrast enhancement subsides, approximately 45% of the initial T1 hypointense lesions gradually return to isointensity, reflecting both remyelination and the loss of extracellular oedema, and around 35% of the acute T1 hypointense lesions will remain hypointense, reflecting axonal loss and loss of myelin [7]. Therefore in particular lesions with persisting T1-hypointensity, or persisting black holes (PBH), are an interesting marker of axonal damage and demyelination.

An important aspect of designing a clinical trial is performing a statistical power analysis. Statistical power is the probability of getting a statistically significant result, given that there is a real biological treatment effect in the population being studied. Although there are no formal standards, a power of 80% is generally regarded as acceptable. Since statistical power is highly influenced by the number of patients enrolled in the study, a power analysis is mainly focussed on determining the required sample size for a given treatment effect. If the sample size is too small, the experiment will lack the precision to demonstrate an anticipated treatment effect. If the sample size is too large, time and resources will be unnecessary wasted, often for a minimal gain. In the present study, the feasibility of using the number of CELs evolving into PBH as a primary outcome measure in phase II clinical trials in MS is assessed. First, we propose a statistical model for the distribution of the number of CELs evolving into

PBH. Second, a sample size calculation derived from a parametric re-sampling and simulation procedure based on the chosen model will be performed. Ultimately, the calculations serve as a basis for the efficient design of future clinical trials.

MATERIALS AND METHODS

Patients

Our analyses were performed with data derived from the oral interferon beta-Ia (IFNB-1a) study [8]. In this study, 173 RRMS patients received 0.06, 0.6, or 6 million international units of IFNB-1a or placebo orally every other day for six months. No clinical or MRI effect of any dose of oral IFNB-1a could be observed (including the absence of an effect on the analyzed outcome parameter of this study, the number of CELs evolving into a PBH (X^2(df=3)=5.19, P=0.16) and combined with low neopterin levels in a subgroup of 21 patients, as well as the absence of neutralizing antibodies in all patients analyzed, oral IFNB-1a was assumed to be biologically inactive. Therefore the whole cohort was regarded as a natural history cohort. For inclusion into the original study at least two relapses within 24 months prior to study entry plus seven T2 lesions on the screening MRI, or at least one relapse prior to study entry in conjunction with at least one CEL and at least another three T2 lesions on the screening MRI were necessary. In this way active RRMS patients were selected, creating a cohort commonly tested in MS proof of concept studies. Of the original cohort of 173 patients, 169 patients were included in our analyses (124 women, 45 men). Patients had a mean age of 35.4 years (SD=8.4), a mean disease duration of 6.4 years (SD=5.3) and a mean baseline expanded disability status scale (EDSS) score of 2.5 (IQR=2.0).

MRI analysis

One baseline scan and six monthly follow-up scans were performed, including a dual-echo, T2-weighted, spin-echo or turbo/fast spin-echo (TR/TE of 2000-3000 / 20-40 & 60-100 ms) and a T1-weighted spin-echo (TR/TE of 400-700 / 5-25 ms), both after administration of 0.1 mmol/kg gadolinium-DTPA intravenously, with a field of view of 25 cm and a 256x256 matrix resulting in roughly 1x1 mm pixel size. Images were acquired in 2x23 interleaved sections with a 3 mm thickness and a 3 mm gap. A complete re-analysis of all available images was done by a single designated reader

following standard operating protocols in the VU medical centre in Amsterdam. For 169 patients, the cumulative number of new CELs developing into a PBH was counted per patient. BHs were considered as PBHs if they did not become isointense over the study period, and could be observed for a minimum of 3 months after initial onset of the respective CEL. Therefore, only CELs observed from baseline up to month three were assessed for PBH evolution. This minimum follow-up period is based on the observation that the majority of BHs that return to isointensity will do so within a period of 3 months [7]. BHs that were observed at baseline, occurring without any evidence of a corresponding Gd-enhancing lesion or that were separated by a gap after the cessation of enhancement, were excluded. All CELs observed at baseline were considered as new.

Statistical methods

The cumulative number of CELs evolving into a PBH after 3 months, was considered the primary outcome variable for the sample size analysis. The simplest statistical distribution to describe a count variable is the Poisson distribution. In case of MS lesion counts, the probability of observing κ lesions is given by

$$P(\kappa)=\frac{e^{-\mu}\,\mu^{\kappa}}{\kappa!}$$

where μ is the mean number of lesions in the study population. It is well known that the variance of the number of lesions is $Var(\kappa)=\mu$.

Sormani *et al*. [9] investigated the distribution of the number of new CELs across patients, and found that the Poisson distribution does not reflect the large variability of the number of new CELs, and that the assumptions it requires are too restrictive. A statistical model that was able to fit the number of CELs more accurately was the negative binomial (NB) model. In this model, the probability of observing κ lesions is described by

$$P(\kappa)=\frac{\Gamma(\kappa+\theta)}{\kappa!\Gamma(\theta)}\times\frac{\mu^{\kappa}\,\theta^{\theta}}{(\mu+\theta)^{\kappa+\theta}}$$

The parameter μ is the mean number of lesions per patient, θ^{-1} is the overdispersion parameter and $\Gamma(\bullet)$ is the gamma function. It can be verified that $Var(\kappa)=\mu+\mu^2/\theta$ [10].

In the present study, both the Poisson model and the NB model were fitted to the PBH evolution dataset, using the statistical software program SAS™, procedure GENMOD. To determine the model best describing the distribution of the number of CELs evolving into a PBH, goodness of fit analyses were performed in the program Stata™, comparing the observed values with the values predicted by the Poisson and the NB distribution. Next, the chosen model was implemented in a statistical re-sampling procedure using SAS™, simulating parallel-group designed MS treatment trials. For each simulated trial, a placebo group was obtained by randomly sampling a number of lesions for each patient from the chosen distribution, with the parameters derived from the oral IFNB-1a study population. A treatment group was obtained by randomly sampling a number of lesions for each patient from the chosen distribution, with parameters $\mu_{treated}=\mu_{placebo}$ X (1 – treatment effect) and $\theta_{treated}=\theta_{placebo}$. In this way, the treatment effect was expressed as the percentage difference in the mean number of counted CELs evolving into a PBH, between a group of treated patients and a group of placebo patients. A total of 100.000 trials were generated for treatment effects ranging from 50% to 90%, and the power was calculated as the proportion of trials yielding a significant result (significance determined by a Wilcoxon Rank sum test at a two-sided test level of 5%).

RESULTS

Descriptives

Table 1 displays the characteristics of the investigated lesions. 127 of 169 patients developed a total of 1216 CELs from baseline up to month three (mean 9.6 CELs / patient, range 1-63); 42 patients developed none. From the 1216 CELs, 258 CELs developed into a PBH. Most CELs were located juxtacortically (42.8%) and in the deep white matter (32.2%), while PBHs were mostly localized juxtacortically (39.9%) and periventricularly (27.5%). 132 of the 1216 analyzed CELs were ring enhancing at first appearance, and 1078 of 1216 CELs nodular enhancing. At subsequent scans, 43 of the 1078 initial nodular enhancing lesions appeared as a ring enhancing lesion. CELs located periventricularly, with a size larger than 10 mm and ring enhancing lesions most frequently evolved into a PBH (30.6%, 57.4% and 59.4% respectively).

Table 1 | Distribution of the number of CELs, PBHs and % of CELs evolving into a PBH, for location, size, and shape.

		Lesion type		% of CELs evolving into a PBH
		CEL	PBH	
Total		1216	258	21.2%
Location	Juxtacortical	520 (42.8%)	103 (39.9%)	19.8%
	Deep White Matter	392 (32.2%)	65 (25.2%)	16.6%
	Periventricular	232 (19.1%)	71 (27.5%)	30.6%
	Infratentorial	72 (5.9%)	19 (7.4%)	26.4%
Size	<5 mm	876 (72.0%)	121 (46.9%)	13.8%
	5 – 10 mm	272 (22.4%)	98 (38.0%)	36.0%
	>10 mm	68 (5.6%)	39 (15.1%)	57.4%
Shape	Nodular (at first appearance)	1078* (89.1%)	179 (69.4%)	16.6%
	Ring (at first appearance)	132 (10.9%)	79 (30.6%)	59.8%
	Ring (after initial nodular appearance)	43 (4.0%**)	25 (14.0%**)	58.1%
	Ring (total)	175 (14.4%)	104 (40.3%)	59.4%

* For 6 CELs no shape data was available

** From nodular shaped lesions at first appearance

Data modelling

In Table 2 and Figure 1, the observed and fitted PBH data according to the Poisson and NB distribution are presented. The estimated Poisson parameter was $\hat{\mu}$=1.527 (95%CI 1.340, 1.713). The estimated NB parameters were $\hat{\mu}$=1.527 (95%CI 1.167, 1.997) and $\hat{\theta}$=0.398 (95%CI 0.280, 0.566). (For comparison: the parameters of the NB model fitting the distribution of the 1216 new CELs is $\hat{\mu}$= 7.195 and $\hat{\theta}$=0.49. Thus, the distribution of new CELs has a different shape than the distribution of PBHs which might indicate that the new CELs evolving into a PBH are not a random subgroup of new CELs). The Poisson distribution proved a poor fit for the number of CELs evolving into a PBH with predicted frequencies deviating markedly from those observed in the data. In comparison, the NB distribution fitted the observed values very well, particularly in the lower categories. The goodness of fit test confirmed both observations with a X^2 value of 177.96 (df=4, P<0.001) for the Poisson model compared to a X^2 of 1.38 (df=8, P=0.99) for the NB model, favouring a choice for the NB model.

Table 2 | Goodness of fit tests and number of patients observed and expected under the Poisson distribution and NB distribution for each category of lesion counts.

Lesion count (k)	Observed	Expected Poisson	Expected NB
0	90	36.718	90.268
1	29	56.055	28.487
2	17	42.788	15.794
3	9	21.774	10.014
4	6	8.310	6.748
5	4	2.537	4.708
6	5	0.646	3.360
7	2	0.141	2.436
8	2	0.027	1.787
9	1	0.005	1.323
≥10	4	0.001	4.074
Sum	169	169	169
Chi square (g.o.f)		177.96*	1.38
df		4	8
p		<0.001	0.99

* Chi square is based on 6 cells by combining the cells of ≥5 lesion counts.

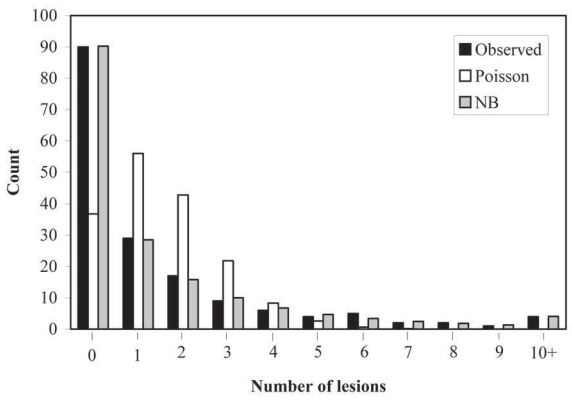

Figure 1 | Histogram of the observed distribution of the cumulative number of CELs developing into PBHs on three subsequent monthly MRI scans (black), and estimated by the Poisson distribution (white) and NB distribution (grey).

Power analysis

Table 3 presents the estimated number of patients in each treatment arm of a parallel group designed trial, necessary to obtain statistical powers of 80% or 90%, levels which are commonly required in MS treatment trials. The presented point estimates are based on simulations with the "real" estimators ($\hat{\mu}$=1.527 and θ=0.398), while the accompanying confidence limits are estimated by means of simulations ran with the lower and upper limits of the 95% confidence interval of the estimated NB parameters.

Roughly, the power analysis provided a range from 30 subjects per arm for a treatment effect of 90% to 200 subjects per arm for a treatment effect of 50%. Figure 2 illustrates the trend of the relation between effect size and sample size. It shows that treatment effects lower than 50% require a dramatic increase in sample size to obtain sufficient power. Since phase II clinical trials of more then 200 subjects are unfeasible in terms of time, costs and effort, the cut off point for the simulated treatment effect was set to 50%.

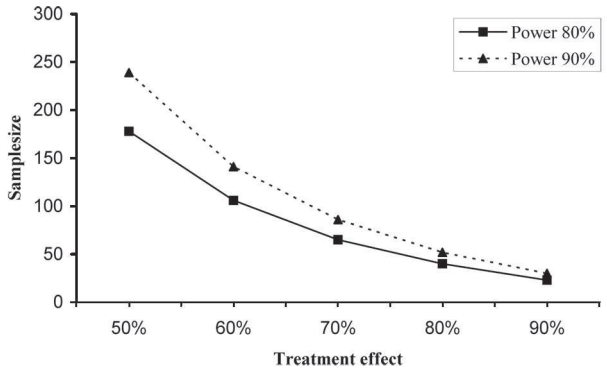

Figure 2 | Relation between the number of patients needed in each arm and treatment effect, for a power of 80% (solid line) and 90% (dotted line).

Table 3 | Number of patients per treatment arm (Point estimate and confidence interval based on lower and upper limits of the 95%CI of NB parameters) necessary to perform parallel group designed trials with statistical powers of 80% or 90%, to detect treatment effects ranging from 50% to 90%.

	Statistical power			
	80%		90%	
Treatment effect	n	CI	n	CI
50%	178	125 - 261	239	162 - 347
60%	106	72 - 152	141	97 - 204
70%	65	45 - 92	86	60 - 125
80%	40	28 - 55	52	37 - 73
90%	23	17 - 31	30	22 - 41

DISCUSSION

This study is based on a less common used MRI outcome parameter in phase II MS treatment trials, the number of CELs evolving into a PBH. Although the Poisson and NB model have previously been proposed to describe the distribution of the number of new CELs across patients [9], this is the first time both models are proposed for the number of CELs evolving into a PBH. We showed that this outcome parameter, assessed over a three month period, is adequately fitted by the NB model, and that the fit is superior to the Poisson distribution. The main aim of the study was to assess the feasibility of using the evolution of CELs into PBHs as the primary MRI outcome parameter in phase II clinical trials in MS. The estimated sample sizes prove to be larger than those required for new CELs per se [11], but are still realistic and acceptable for a phase II clinical trial. One must take into account that a potent drug, capable of reducing the number of CELs evolving into a PBH with at least 50%, is required.

A reduction of the number of CELs evolving into a PBH is either realized by decreasing the tendency of CELs to develop into a PBH, by suppressing the initial number of CELs, or a combination. In a clinical trial with Interferon beta-1b in 125 SPMS patients, Brex *et al*. [12] showed that 70/500 CELs evolved into a PBH in the placebo group whereas 24/179 CELs did so in the treatment group. Although this result demonstrated that IFN-beta 1b was unable to alter the process of CELs (once occurring) to evolve into a PBH, the absolute number of PBHs ultimately occurring had

been lowered dramatically (70 *vs*. 24) mainly as an effect of reduced inflammatory activity. When the aim of a study is to demonstrate the specific effect of a drug to down regulate the process of CELs to evolve into a PBH, the *proportion* of CELs evolving into a PBH is the outcome measure of choice. However, if one would like to determine the overall amount of neuro-axonal destruction in a given patient under treatment, the *absolute number* of PBHs is a more relevant measure since both neuroprotective effects as well as anti-inflammatory effects of a therapy are of influence on the eventual residual deficit. Moreover, many (immunomodulatory) treatments share anti-inflammatory and neuroprotective properties, and a formal separation of those processes may be artificial in most circumstances.

We based the sample size estimations on a unique sample of RRMS patients since it represents the natural evolution of CELs evolving into PBHs while patients were originally selected to receive actual treatment. Together with a natural evolution of 21.2% of the CELs evolving into a PBH, a proportion fitting in with the range of results from previous studies (14-38%) [12,13,14], the sample is highly representative for future research populations. However, the main prerequisite for valid sample size estimation is that the number of CELs evolving into a PBH is NB distributed. Although this is true for our data, a definite choice for the NB model can only be validated when the fit is able to show consistent results in other datasets. Minneboo *et al*. [15] showed that the within patient variance for MRI parameters used to describe lesion evolution is considerably lower then the between-patients-variance, indicating that lesion evolution on MRI is a patient-specific phenomenon. If this observation is taken into account, a clustered analysis would be a more appropriate method to analyze the data, possibly influencing the eventual statistical modelling. Furthermore, it remains to be seen if there is an effect on the distribution of the number of CELs evolving into a PBH when the follow up period for BH evolution is extended to 6 or 9 months, just as it is not unlikely that the distribution could deviate from the NB model for other MS phenotypes. When selecting patients for the presence of a CEL at baseline will enrich a study with active patients and subsequently increase the number of PBHs, it is also plausible that a baseline selection criterion for PBH development, for example the presence of a PBH or a T1/T2 lesion load ratio, will increase the number of PBHs even further. Although the current dataset lacked the MRI outcomes to assess both selection criteria, the effects on the distribution of the number of CELs evolving into a PBH and the subsequent sample size estimates is worth the exploration in future research.

With treatment strategies focussing on prevention of brain tissue damage, this study aimed to provide a starting point for the efficient design of future phase II clinical trials and concludes that the number of CELs evolving into a PBH is a feasible outcome parameter to measure the extent of neuro-axonal damage, provided that a potent drug is administered.

ACKNOWLEDGEMENTS

We are very grateful to Bayer-Schering Pharma, Berlin, Germany, for providing us with the data of the oral Interferon study used in our analyses. The MS Center Amsterdam is supported by the Dutch MS Research Foundation (Grant. no. 05-358c).

REFERENCES

1. **Miller DH, Barkhof F, Nauta JJP.** Gadolinium enhancement increased the sensitivity of MRI in detecting disease activity in MS. *Brain* 1993; 116: 1077-94.
2. **McFarland HF, Frank JA, Albert PS, Smith ME, Martin R, Harris JO et al.** Using gadolinium enhanced magnetic resonance imaging to monitor disease activity in Multiple Sclerosis. *Ann Neurol* 1992; 32: 758-66.
3. **Bruck W, Bitsch A, Kolenda H, Bruck Y, Stiefel M, Lassmann H.** Inflammatory central nervous system demyelination: correlation of magnetic resonance imaging findings with lesion pathology. *Ann Neurol* 1997; 42: 783-793.
4. **Van Walderveen MA, Kamphorst W, Scheltens P, van Waesberghe JH, Ravid R, Valk J et al.** Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. *Neurology* 1998; 50: 1282-1288.
5. **van Waesberghe JH, Kamphorst W, De Groot CJ, van Walderveen MA, Castelijns JA, Ravid R et al.** Axonal Loss in Multiple Sclerosis Lesions: Magnetic Resonance Imaging insights into substrates of disability. *Ann Neurol* 1999; 46: 747–754.
6. **Bitsch A, Kuhlmann T, Stadelmann C, Lassmann H, Lucchinetti C, Bruck W. A** longitudinal MRI study of histopathologically defined hypointense Multiple Sclerosis lesions. *Ann Neurol* 2001; 49: 793–796.
7. **van Waesberghe JH, van Walderveen MA, Castelijns JA, Scheltens P, Lycklama a Nijeholt GJ, Polman CH et al.** Patterns of lesion development in multiple sclerosis: longitudinal observations with T1-weighted spin echo and magnetization transfer MR. *AJNR Am J Neuroradiol* 1998; 19: 675-683.
8. **Polman C, Barkhof F, Kappos L, Pozzilli C, Sandbrink R, Dahlke F et al.** and the European Oral Interferon Beta-1a in Relapsing-Remitting MS Study Group. Oral interferon beta-1a in relapsing-remitting multiple sclerosis: a double-blind randomized study. *Mult Scler* 2003; 9: 342-348.
9. **Sormani MP, Bruzzi P, Rovaris M, Barkhof F, Comi G, Miller DH et al.** Modelling new enhancing MRI lesion counts in multiple sclerosis. *Mult Scler* 2001; 7 : 298-304.
10. **Agresti A.** Categorical data analysis. 2nd edition 2002, Wiley New York; 559-560.
11. **Sormani MP, Miller DH, Comi G, Barkhof F, Rovaris M, Bruzzi P et al.** Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large datasets. *J Neurol Neurosurg Psychiatry* 2001; 70: 494-499.
12. **Brex PA, Molyneux PD, Smiddy P, Barkhof F, Filippi M, Yousry TA et al.** and the European study group on Interferon beta-1b in secondary progressive MS. The effect of IFNß-1b on the evolution of enhancing lesions in secondary progressive MS. *Neurology* 2001; 57: 2185-2190.
13. **Ciccarelli O, Giugni E, Paolillo A, Mainero C, Gasperini C, Bastianello S et al.** Magnetic resonance outcome of new enhancing lesions in patients with relapsing-remitting multiple sclerosis. *Eur J Neurol* 1999; 6: 455-459.
14. **Bagnato F, Jeffries N, Richert ND, Stone RD, Ohayon JM, McFarland HF et al.** Evolution of T_1 black holes in patients with multiple sclerosis imaged monthly for 4 years. *Brain* 2003; 126: 1782-1789.

15. **Minneboo A, Uitdehaag BMJ, Ader HJ, Barkhof F, Polman CH, Castelijns JA.** Patterns of enhancing lesion evolution in multiple sclerosis are uniform within patients. *Neurology* 2005; 65: 56-60.