



Published in final edited form as:

Mult Scler. 2008 July ; 14(6): 770–778. doi:10.1177/1352458507088104.

Magnetic resonance imaging as a surrogate outcome for multiple sclerosis relapses

J Petkau¹, SC Reingold², U Held³, GR Cutter⁴, TR Fleming⁵, MD Hughes⁶, DH Miller⁷, HF McFarland⁸, and JS Wolinsky⁹

¹Department of Statistics, University of British Columbia, Vancouver, BC, Canada

²Scientific and Clinical Review Associates, LLC and National Multiple Sclerosis Society, New York City, NY, USA

³Horton Centre for Patient-Oriented Research, University Hospital, Zurich, Switzerland (previously at The Sylvia Lawry Centre for Multiple Sclerosis Research, Munich, Germany)

⁴Department of Biostatistics, University of Alabama at Birmingham, AL, USA

⁵Biostatistics Department, University of Washington, Seattle, WA, USA

⁶Harvard School of Public Health, Boston, MA, USA

⁷NMR Research Unit, Institute of Neurology, London, UK

⁸Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

⁹Department of Neurology, University of Texas Health Sciences Center, Houston, TX, USA

Abstract

Background—Magnetic resonance imaging (MRI) of lesions in the brain may be the best current candidate for a surrogate biological marker of clinical outcomes in relapsing remitting multiple sclerosis (MS), based on its role as an objective indicator of disease pathology. No biological surrogate marker has yet been validated for MS clinical outcomes.

Objective—The objective of this study was to use a multi-phased study to determine if a valid surrogate relationship could be demonstrated between counts of contrast enhancing lesions (CELs) and occurrence of relapses in MS.

Methods—We examined correlations for the concurrent and predictive relationship between CELs over 6 months and MS relapses over the same 6 months and an additional 6 months (total: 12 months), using available data on untreated patients from a large clinical trial and natural history database.

Results—Concurrent and predictive correlations were inadequate to justify continuation of this study to the planned additional phases required to demonstrate a surrogate relationship between CELs and MS relapses.

Conclusions—Confidence intervals for correlations between CELs and MS relapses exclude the possibility that CELs can be a good surrogate for relapses over the time scales we investigated. Further exploration of surrogacy between MRI measures and MS clinical outcomes may require improved

datasets, the development of MRI techniques that couple better to clinical disease, and the ability to test a wide range of imaging- and clinically-based hypotheses for surrogacy.

Keywords

multiple sclerosis; magnetic resonance imaging; gadolinium enhanced lesions; surrogacy; prognosis; correlations

Introduction

Clinical outcomes from pivotal therapeutic trials of new agents for multiple sclerosis (MS) are the only measures accepted by regulatory agencies and, generally, by the MS clinical community. These have included rate of relapses and accumulation of disability, primarily assessed using the Expanded Disability Status Scale (EDSS). Statistical differences in these measures of clinical importance have led to marketing approval for six agents in four distinct therapeutic classes [1–6]. With more therapies available, it has become practically and ethically difficult to design and implement multiyear placebo-controlled randomized clinical trials. Shorter trials based on surrogate outcomes using fewer patients are of interest when determining a clinical outcome is impractical, inconvenient, uncomfortable, or intolerable for subjects [7–10]. Changes in MS brain lesion patterns determined by magnetic resonance imaging (MRI) reflect changes in underlying disease pathology [7–9,11], providing a strong theoretical basis for surrogacy.

Unvalidated MRI surrogate outcomes are accepted for earlier Phase II clinical trials in MS to demonstrate biological activity, provide proof of concept for a potential new treatment, and guide further protocol development [8,12]. Unvalidated surrogates might also find regulatory favor if treatment effects on these measures are reasonably likely to predict treatment effects on clinical outcomes and when the agent being tested is found to have ‘meaningful benefit over existing therapies’ [8]. However, MRI outcomes have not been validated as surrogates for clinical outcomes in MS. In the absence of a rigorously validated surrogate endpoint, the primary endpoint for any pivotal therapeutic trial should be clinical [13].

Formal validation of a surrogate is not an easy task. A surrogate outcome is considered valid for a specific intervention if the surrogate is correlated with the true clinical outcome *and* the impact of the intervention on the true clinical outcome is accurately captured by its effect on the surrogate outcome [14]. A change in a valid surrogate outcome must predict a change in the true clinical outcome. In a clinical trial, this is achieved if a therapy affects the true clinical outcome mediated through its effect on the surrogate outcome and there are not other important pathways of impact on the true outcome that bypasses the surrogate (Figure 1). While surrogacy might be demonstrated in a particular clinical trial setting with one class of agent, without appropriate prospective trials or retrospective meta-analyses, the overall validity of a proposed surrogate for any disease cannot be assumed for different clinical trials settings or for different classes of interventions.

To demonstrate that an imaging outcome is a valid surrogate for MS clinical trials, an adequate natural history and treatment database is needed that captures both clinical and imaging data at regular intervals and which can be subject to statistically rigorous meta-analyses [8]. Modeling of extant data can then lead to prospective studies where the concept of surrogacy can be evaluated in a trials setting. A large database resource resides at the Sylvia Lawry Centre for MS Research (SLCMSR). The authors developed a tiered multi-phased strategy for assessing the validity of using contrast enhancing lesions (CELs) as a surrogate for relapses in MS clinical trials using data from the SLCMSR resource. The periodic nature of both CELs

and relapses and their apparent temporal association made this MRI measure a reasonable candidate for which to begin a formal evaluation of surrogacy.

Methods

The dataset

When this work was carried out, the SLCMSR database included virtually all major past MS natural history studies and usable data from the placebo arms from 31 past clinical trials in MS. The datasets were provided under contractual agreements between the Centre and investigators and corporate sponsors, agreements that among other restrictions rigorously protect the anonymity of the data source. The SLCMSR randomly splits the dataset into an open fraction and a closed fraction that is held back to allow validation of the major findings from the analyses on the open part. This policy was established to avoid the potential bias that arises when a dataset is used to generate hypotheses and then formal statistical tests for those hypotheses are carried out using the same dataset [15–17].

For the analyses in this paper, the authors required access to patient information from the placebo arms of randomized controlled clinical trials and had access only to the open fraction of the database. An interactive on-line query system was used to search for relevant qualifying studies and patients. In all, 2714 patient-years of data from 1354 patients in placebo arms of such studies were potentially informative for this analysis. Only a subset of these data, from 208 patients, were relevant for the current study, as it required availability of specific time-sensitive information about MRI scans, the number of new CELs on each scan and information about relapses for each patient (Figure 2). The data pool was heterogeneous since usable subjects came from a variety of studies done at different times using different inclusion criteria, different but contemporary imaging measures and different imaging analysis centers.

Project design

Through an iterative consensus process that included a survey of needed imaging and clinical data resident in the SLCMSR database, the investigator team formulated a study with seven potential phases. At the conclusion of each study phase, an assessment of the continued feasibility of the project was to be made, considering the analysis outcomes at that phase and the availability of needed data for the next phase. The overall project would terminate if any of the phases were not successfully concluded.

We chose to begin with an assessment of correlations that would suggest surrogacy between the number of new CELs per scan over 6 months of follow-up and: (a) the annualized relapse rate over the same 6-month interval (to assess *concurrent validity*); and (b) the annualized relapse rate over 12 months of follow-up (to assess *predictive validity*). The project included the following phases:

Phase 1—A search of the open fraction of the SLCMSR database for patients who could contribute data needed for the analysis. We used an on-line analysis tool to search the database to determine how many studies and how many patients from each study fulfilled a set of study-relevant operator-specified criteria for MRI and clinical relapse data.

Phase 2—A feasibility assessment, from the available dataset identified in phase 1, to determine the strength of the relationship between annualized relapse rate and number of new CELs per scan determined using all available scans for each patient. Correlations were evaluated separately for each trial contributing a reasonable number of patients as well as for the overall dataset. In addition, a ‘combined’ estimate of the correlation was evaluated using a ‘fixed effects’ meta-analysis approach to combine the estimated correlations from the

different trials. We agreed that phase 3 would be initiated only if the correlation between CELs and relapse outcomes determined in phase 2 was sufficiently high ($r > 0.5$ was considered a reasonable guideline) but that we would not rely solely upon a numerical ‘cut-off’ to make that decision.

Statistical analysis—The statistical analysis in phase 2 consisted of correlation analyses (Spearman rank and Pearson correlation coefficients with corresponding 95% confidence intervals) to assess the concurrent and predictive relationships between relapse outcomes and putative CEL surrogates. The ‘combined’ estimate of the correlation is the inverse transform of the weighted average of the Fisher transforms of the different correlations with weights given by $n_i - 3$, where n_i is the number of patients contributing data from the i th trial. Provided the correlations do not exhibit evidence of heterogeneity across the trials, the ‘combined’ estimate provides a reasonable description of the level of correlation in the collection of trials. The heterogeneity of the correlations across trials was assessed with Cochran’s chi-squared test.

Phase 3—We planned to model the relationship between relapse outcomes and putative MRI surrogate outcomes of interest, using data from placebo patients in the open fraction of the SLCMSR database. The objective was to determine whether putative MRI surrogates continue to be predictive of relapse outcomes after adjustment for baseline covariates that may be important predictors for relapse outcomes. For these assessments, the annualized relapse rate would be the clinical outcome of primary interest and those of secondary interest would be time to first relapse and relapse-free status for patients. For CELs, the number of new lesions per scan and the proportion of active scans were considered to be potential surrogates. Trial identification, disease course, gender, disability as measured by the Expanded Disability Status Scale (EDSS), age, disease duration, and prior relapses were considered to be potentially important baseline covariates to be assessed.

Statistical analysis—A predictive (regression) model for the relapse outcome was to be developed using a carefully guided stepwise forward selection procedure (with Poisson, logistic and proportional hazards regression for relapse rate, relapse status and time to first relapse). Each step would examine whether the putative CEL surrogate outcome continued to add predictive capability beyond that provided by the included baseline covariates. Interactions among the predictors and non-linear relationships were also to be considered. Detailed diagnostic evaluation of the model at each step in this process was viewed as essential to enhance confidence in the overall analysis.

Phases 4–7—In the initial study design, only phases 1–3 were outlined in detail. Assuming successful completion of phases 1–3, subsequent phases were to include the following:

- Phase 4: Validation of major findings on the relationship between relapse outcomes and putative MRI surrogates in placebo patients from the closed fraction of the SLCMSR database.
- Phase 5: Acquisition of data on treated patients (in particular those who were part of trials from which placebo patients were used in phase 2 of the study), to attempt to replicate findings from placebo patients in similar treated patients. This would require the solicitation and cooperation of original data donors to the SLCMSR database, with a focus on obtaining data from trials of a variety of classes of medication.
- Phase 6: Replication of the model fit, to confirm the findings from prior phases on treated patients.

- Phase 7: Final assessment of surrogacy, to determine the extent to which treatment effects seen in the surrogate CEL outcomes are predictive of treatment effects observed in relapse outcomes. This would require analysis in trials for which both placebo and treatment data are available, possibly with assessment in prospective controlled clinical trials.

Results

Phase 1: Identifying the dataset

To undertake this analysis, an on-line analysis tool was adapted to search for data on placebo patients from clinical trials in the SLCMSR database with relapsing-remitting and secondary progressive MS or clinically isolated syndrome suggestive of MS. We identified trials and patients who contributed clinical data and MRI scans at baseline (time of entry into the clinical trial) and at 3 and 6 months of follow-up, and for whom counts were available of new CELs for 6 months of follow-up and relapse rate information for 12 months of follow-up. Annualized relapse rates over both 6 and 12 months of follow-up for patients with adequate imaging data could be calculated for 208 patients from 11 separate trials, including 132 with relapsing-remitting disease, 76 with secondary progressive disease, and none with clinically isolated syndrome suggestive of MS (Figure 2). All but one of these 208 patients had at least six MRI scans with new CEL counts available over the specified time period. This constituted the dataset to be explored in the next phase of assessment.

Phase 2: Correlation between CELs and relapses

For the analysis of the relationship between CELs and relapses, Spearman and Pearson correlations and corresponding 95% confidence intervals were calculated between the number of new CELs per scan over 6 months of follow-up and the annualized relapse rate over the same 6-month interval (to assess concurrent validity), and the annualized relapse rate over 12 months of follow-up (to assess predictive validity). The raw data for all $n = 208$ patients presented in Figure 3 illustrates the low strength of these concurrent and predictive relationships.

Figure 4 demonstrates the low correlation between these outcomes, in the individual trials and overall, for both concurrent and predictive validity. Overall, both concurrent and predictive correlations between numbers of CELs and relapses were close to zero in all analyses, using both Spearman and Pearson estimates of correlation, with some negative correlations (Table 1 and 2). The estimates exhibited no evidence of heterogeneity across the trials ($p > 0.50$ for both the concurrent and predictive relationships). Therefore, the level of correlation in these trials can be reasonably described by the 'combined' estimates. The confidence intervals based on the 'combined' Pearson estimates effectively exclude the possibility of strong linear associations between numbers of CELs and relapses. This was also the case for the smaller subset of patients identified with pure relapsing-remitting disease ($n = 132$, data not shown).

Discussion

This effort to demonstrate a surrogate relationship between new CELs and relapse rate in MS clinical trials did not proceed beyond the phase 2 feasibility step. The strength of the necessary correlations for both concurrent and predictive surrogacy was insufficient to sustain the originally planned subsequent five phases of analysis. As per prior agreement the project as initially designed was terminated at the conclusion of phase 2.

Correlations were initially assessed using the Spearman correlation coefficient, r_S , which showed weak but generally positive correlations (Table 1 and 2). Although r_S describes the

strength of association between two outcomes and can be used to assess whether an association exists, it is not an estimate of the usual population correlation (ρ = Pearson's measure of *linear association* for a population) [18]. The objective of our study was to estimate the magnitude of the usual population correlation rather than merely to assess whether the data provided convincing evidence of a non-zero association (i.e., to test the null hypothesis that the association was zero). Therefore, the Pearson correlation coefficient, r_p , was used as an additional estimator and for construction of 95% confidence intervals for the usual population correlation between new CELs per scan and annualized relapse rate for concurrent and predictive validity assessment. As illustrated in Figure 4, these confidence intervals included the possibility of there being no correlation. Despite the significance of the 'combined' Spearman correlations for concurrent relapses and CELs ($r_S = 0.19$, $p = 0.01$) and predictive relapses and CELs ($r_S = 0.23$, $p = 0.003$), the general pattern was of modest correlations in all the individual trials and the confidence intervals based on the 'combined' Pearson estimates effectively exclude the possibility of strong linear associations between numbers of CELs and relapses (Figure 4).

This outcome has many possible causes. The most obvious, but not necessarily correct, conclusion is that new CELs are not a surrogate for relapses in MS. However, there are other explanations, including the short duration of the follow-up period for relapses we chose. A more cautious conclusion is that, within the narrow imaging and relapse measures we considered and the limitations of data available for analysis, we could not demonstrate a hypothesized surrogate relationship. While the available sample for analysis was small, the confidence intervals based on the combined Pearson estimates indicate CELs cannot be a good surrogate for relapses, at least not for the time scales we investigated. Having access to a larger dataset (perhaps by using the closed fraction as well as the open fraction of the database, which we considered) would not likely have had any impact on this outcome: the complete database would not have been expected to substantially change the estimated size of the true correlation between CELs and relapses, but would mainly have shortened the length of the confidence intervals and reduced the p -values.

In formulating the steps for this project, the authors were aware that the clinical trials that contributed data to the SLCMSR dataset were done at different times and for different purposes (and for purposes in no way related directly to this project). The studies were conducted by different investigators using different MRI protocols and enrolled patients with different demographics, disease type, duration, and severity. Information was not available for our analysis about most of these potential covariant factors or for the potential effect of steroid treatment on CELs. Further division of the sample as a consequence of such variables would have further eroded the sample size, but these factors could possibly have impacted our analysis. Only placebo subjects from clinical trials were available for analysis, as is essential in any initial steps to evaluate surrogacy. Had the tested relationship been valid for this group of 'untreated' subjects, testing surrogacy in treated subjects would have been an important additional confirmatory step (as outlined in the description of phase 5 in the study plan).

The usefulness and limitations of the SLCMSR database with regard to study heterogeneity have recently been explored [19]. In our analyses, in step-wise fashion, a potential database of 1354 placebo arm patients from clinical trials housed in the open fraction of the SLCMSR database reduced to 208. This sample erosion was a consequence of the hypothesis that was tested; we required subjects for whom there was on-study relapse information at both 6 and 12 months (which alone would have reduced the number of available subjects from 1354 to 1020) and for whom information was available from CEL analysis at baseline, 3 and 6 months (which alone would have reduced the number of available subjects from 1345 to only 238). The need for subjects with *both* specified relapse and MRI information resulted in an overall attrition from 1345 to 208 subjects. Different hypotheses or analytic plans may have resulted in more

available subjects for analysis, but would not necessarily have resulted in a different outcome. Sample attrition is a problem inherent in any large collection of data where there are multiple data sources, no broadly applicable *a priori* unifying agreement about data field definitions, and where the scientific questions are, by necessity, usually developed ‘post hoc’ well after the stored data – which are usually collected for different reasons than the point of any given later analysis – are aggregated.

Although the current study did not demonstrate correlations that would signal surrogacy between a single easily-measured assessment of new CELs and clinical relapses in MS clinical trials, the requirements for establishing surrogacy for any biological marker are complex [10, 14] and require large, and carefully constructed, datasets. Such datasets may need to be established specifically for the purpose of assessing surrogacy. Previous work supports further investigation of this premise. Statistically significant but modest correlations were found between cumulative number of CELs and total number of relapses in both placebo and treatment arms of a clinical trial of glatiramer acetate for relapsing-remitting MS [20]. A European based clinical trial of interferon beta-1b for secondary progressive MS, also demonstrated significant correlations between change in T2 lesion volume and change in disability measured by EDSS; between cumulative number of active T2 lesions and both change in disability and relapse rate; and in a subgroup of patient undergoing frequent monthly scanning, a significant but modest correlation between number of active CELs over 6 months and total relapses over 24 months [21]. From the same study, a significant correlation between relapse rate and number of active T2 lesions over three years and between relapse rate and T2 lesion volume change over one year was found [22]. The relationship between imaging and relapses was stronger than between imaging and progression of disability in this analysis, leading the authors to suggest that using MRI measures alone to measure disability outcomes in studies of secondary progressive disease could be problematic. Finally, a meta-analysis of nine natural history and placebo controlled studies of both relapsing remitting and secondary progressive MS, showed that the mean number of CELs in monthly scans over the first 6 months of study moderately predicted relapse rate in the first full year, potentially indicating that our concurrent outcome period of 6 months (and perhaps even our predictive outcome period of 12 months) was too short to accurately estimate the patients’ relapse rates [23]. However, the data from the majority of the studies do not support investigation of longer periods of time. In the meta-analysis, frequent scanning gave a better predictive value for relapses than a single baseline scan but neither was valuable in predicting change in disability over one or two subsequent years.²⁴ None of these possibilities were pursued in a formal assessment of surrogacy.

A strength of our study was the development of a strict analysis plan before data analysis which should have reduced the probability to produce false positive findings [16]. However, it may have been limited by our initial hypothesis which focused solely on the controversial topic of MRI surrogacy for clinical relapses and which therefore placed restrictions on the measures of CELs and relapses that were considered for surrogacy evaluation. We did not explore any other potential surrogacy relationship like the value of clinical variables such as pre-study relapse rate, disability, or disease duration in predicting future relapses, which has been explored elsewhere [25]. While our chosen measures, the number of CELs per scan and the annualized relapse rate, are among the most simple to explore, it is possible that more sophisticated approaches to both imaging and relapse outcomes would reveal stronger correlations suggestive of surrogacy. For instance, the appearance of some threshold number of CELs on a single scan at any point might increase the relative risk for a clinical exacerbation over some limited epoch of time; or cumulative numbers of new CELs above a predetermined threshold level on serial imaging might better predict future relapses at various specified times post-scan. Other MRI lesion activity measures warrant investigation in future studies, such as lesion volume, counts of new T2 lesions which may be more sensitive than new CELs [6] and more

sophisticated measures such as magnetization transfer ratio imaging, which has recently been shown to have potential predictive value for accumulation of disability [24]. Finally, lesion location and morphology (for instance the presence of ring enhancement) and the relative clinical eloquence of the lesion may also be a factor in surrogacy for clinical outcomes. These hypotheses were not tested in our analysis. Further exploration of surrogacy between MRI measures and MS clinical outcomes might be enhanced by development and validation of MRI techniques that better reflect disease pathology, improved standardization across trials and imaging centers, composite MRI measures and enhanced automated image analysis.

Acknowledgements

The authors thank Fredrick Barkhof MD (Free University, Amsterdam, The Netherlands) for helpful commentary on the manuscript and Martin Daumer PhD and Anneke Neuhaus MSc (Sylvia Lawry Centre for MS Research, Munich, Germany) for participation in the project, data gathering and analysis. Data access and analysis at the Sylvia Lawry Centre and group meetings were supported by a research contract from the National Multiple Sclerosis Society (USA).

References

1. IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. 1. Clinical results from a multicenter, randomized, doubleblind, placebo-controlled trial. *Neurology* 1993;43:655–661. [PubMed: 8469318]
2. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results from a phase III multicenter, double-blind, placebo-controlled trial. *Neurology* 1995;45:1268–1276. [PubMed: 7617181]
3. Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, et al. The Multiple Sclerosis Collaborative Research Group (MSCRG). Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 1996;39:285–294. [PubMed: 8602746]
4. PRISMS. (Prevention of Relapses and Disability by Interferon-beta 1a Subsequently in Multiple Sclerosis) Study Group. Randomized, double-blind, placebo-controlled study of interferon-beta 1a in relapsing-remitting multiple sclerosis: clinical results. *Lancet* 1998;352:1498–1504. [PubMed: 9820297]
5. Hartung HP, Gonsette R, König N, Kwiecinski H, Guseo A, Morrissey SP, et al. Mitoxantrone in progressive multiple sclerosis: a placebo-controlled, double-blind, randomized, multicentre trial. *Lancet* 2002;360:2018–2025. [PubMed: 12504397]
6. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2006;354:899–910. [PubMed: 16510744]
7. Whitaker JN, McFarland HF, Rudge P, Reingold SC. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Mult Scler* 1995;1:37–47. [PubMed: 9345468]
8. McFarland, HF, Barkhof F, Antel J, Miller DH. The role of MRI as a surrogate outcome measure in multiple sclerosis. *Mult Scler* 2002;8:40–51. [PubMed: 11936488]
9. McFarland HF, Reingold SC. The future of multiple sclerosis therapies: redesigning multiple sclerosis clinical trials in a new therapeutic era. *Mult Scler* 2005;11:669–676. [PubMed: 16320726]
10. Fleming TR. Surrogate endpoints and the FDA's accelerated approval process. *Health Aff* 2005;24:67–78.
11. Miller DH, Grossman RI, Reingold SC, McFarland HF. The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain* 1998;121:3–24. [PubMed: 9549485]
12. Miller DH, Khan OA, Sheremata WA, Blumhardt LD, Rice GP, Libonati MA, et al. A controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2003;348:115–123.
13. Fleming TR, Demets RL. Surrogate endpoints in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605–613. [PubMed: 8815760]
14. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–440. [PubMed: 2727467]

15. Schach, S.; Daumer, M.; Neiss, A. Maintaining high quality of statistical evaluations based on the SLCMSR database. Validation procedure of the SLCMSR. 2007 [accessed 6 August 2007]. Available on line at: http://www.slcmr.net/download/publikationen/Validation_Policy.pdf
16. Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124. [PubMed: 16060722]
17. Young PJ, Lederer C, Eder K, Daumer M, Neiss A, Polman CH, et al. Relapses and subsequent worsening of disability in relapsing remitting multiple sclerosis. *Neurology* 2005;67:804–808. [PubMed: 16966541]
18. Lehmann, EL. Nonparametrics: statistical methods based on ranks. San Francisco: Holden-Day, Inc; 1975.
19. Schach S, Scholz M, Wolinsky JS, Kappos L. Pooled historical MRI data as a basis for research in multiple sclerosis - A statistical evaluation. *Mult Scler* 2007;13:509–516. [PubMed: 17463073]
20. Comi G, Filippi M, Wolinsky JS. European/Canadian Glatiramer Acetate Study Group. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging-measured disease activity and burden in patients with relapsing multiple sclerosis. *Ann Neurol* 2001;49:290–297. [PubMed: 11261502]
21. Molyneux PD, Barker GJ, Barkhof F, Beckmann K, Dahlke F, Filippi M, et al. Clinical-MRI correlations in a European trial of interferon beta-1b in secondary progressive MS. *Neurology* 2001;57:2191–2197. [PubMed: 11756596]
22. Sormani MP, Bruzzi P, Beckmann K, Wagner K, Miller DH, Kappos L, et al. MRI metrics as surrogate endpoints for EDSS progression in SPSM patients treated with IFN beta-1b. *Neurology* 2003;60:1462–1466. [PubMed: 12743232]
23. Kappos L, Moeri D, Radue EW, Schoetzau A, Schweikert K, Barkhof F, et al. Gadolinium MRI Meta-analysis Group. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. *Lancet* 1999;353:964–969. [PubMed: 10459905]
24. Agosta F, Rovaris M, Pagani E, Sormani MP, Comi G, Filippi M. Magnetization transfer MRI metrics predict the accumulation of disability 8 years later in patients with multiple sclerosis. *Brain* 2006;129:2620–2627. [PubMed: 16951409]
25. Held U, Heigenhauser L, Shang C, Kappos L, Polman C. Predictors of relapse rate in MS clinical trials. *Neurology* 2005;65:1769–1773. [PubMed: 16344520]

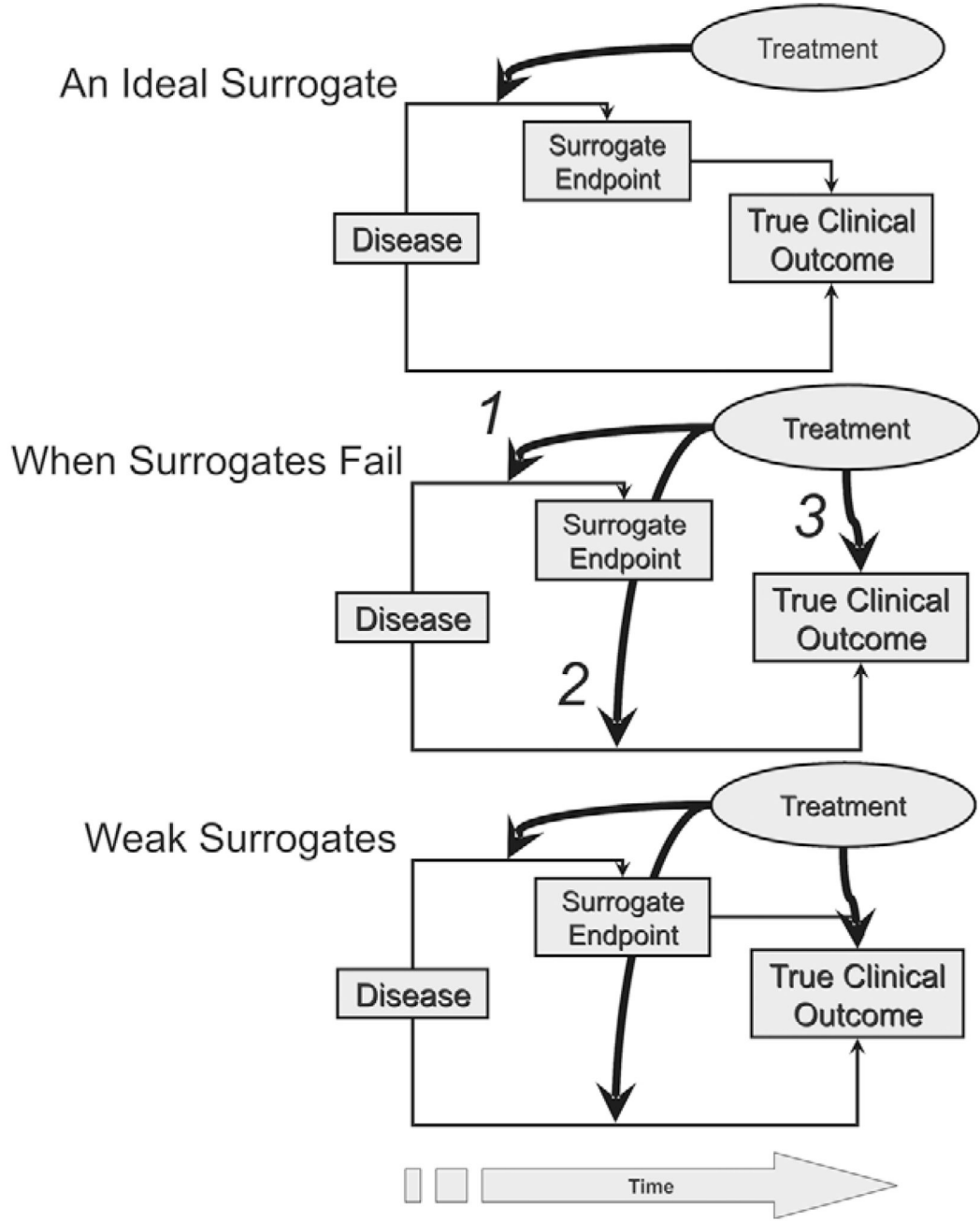


Figure 1. Surrogates of clinical outcomes. For an ideal surrogate the intervention must act through a surrogate in a manner that is directly expressed on the true clinical outcome (upper panel). Surrogates may fail to predict the clinical outcome even if modified by the intervention when the surrogate is not in the causal pathway of the disease process (middle panel). In the example of a treatment effective in only the pathway represented by arrow 1, the surrogate will be affected but the clinical outcome cannot be affected. If the therapy is active only on the pathway represented by arrow 2, the surrogate is NOT affected and yet the treatment has a true effect on the disease process and a beneficial outcome. If active only on the pathway represented by arrow 3, neither the surrogate nor the disease is affected by the treatment, but a symptomatic

effect might lead to a misinterpreted drug effect, as might occur for the use of baclofen for spasticity. Finally, the surrogate could reside in the causal pathway of the disease process, but multiple processes are active with different effects of the treatment on these (lower panel). In this example the surrogate will be affected by therapy but its relationship to the clinical outcome will be diluted by an effect of the treatment on a more important downstream pathway connected to the beneficial outcome, and an additional symptomatic effect of therapy will distort interpretation of the drug effect and further dilute the effect on the surrogate.

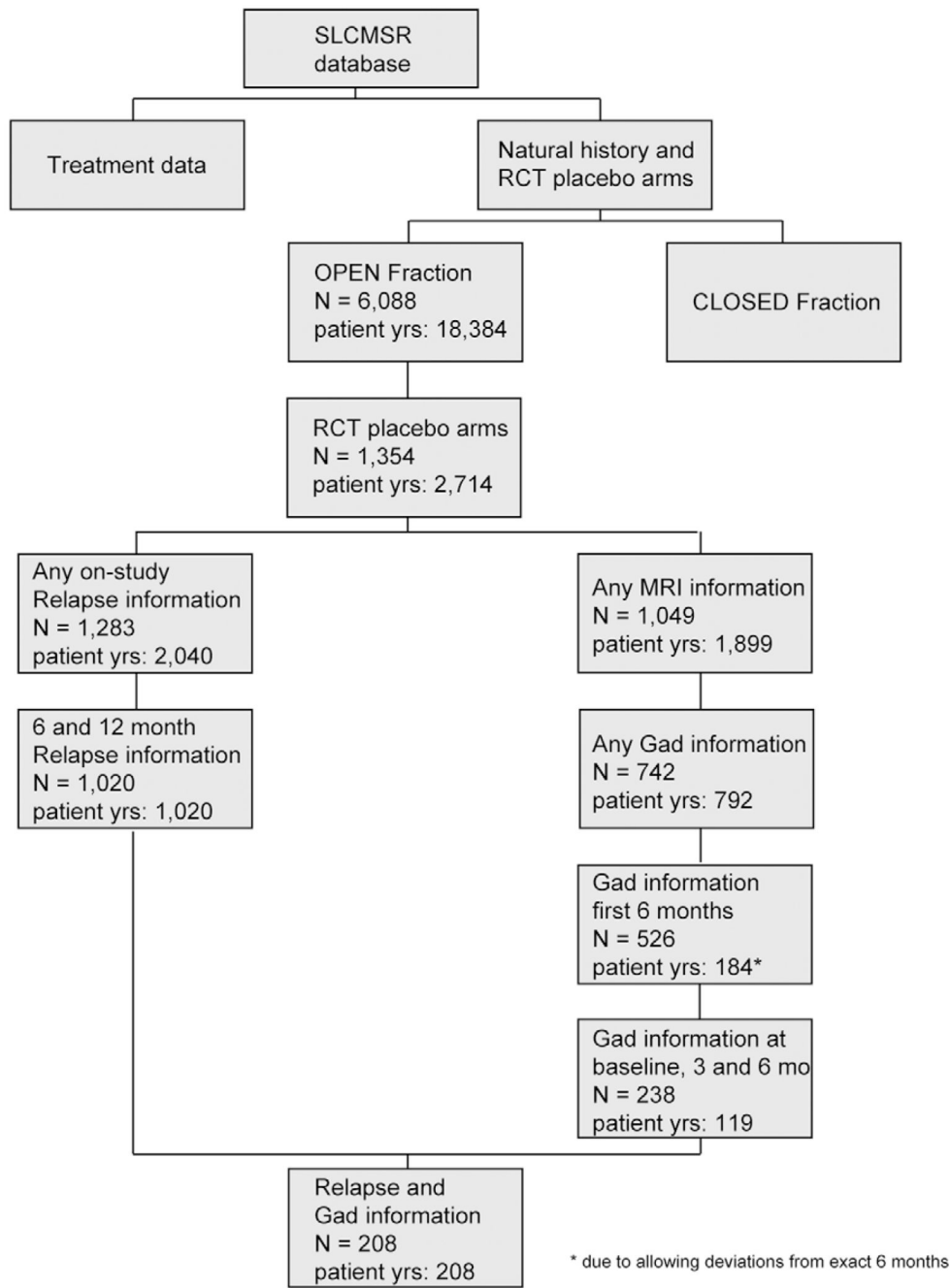


Figure 2.
Flow chart of data selection from the SLCMSR database.

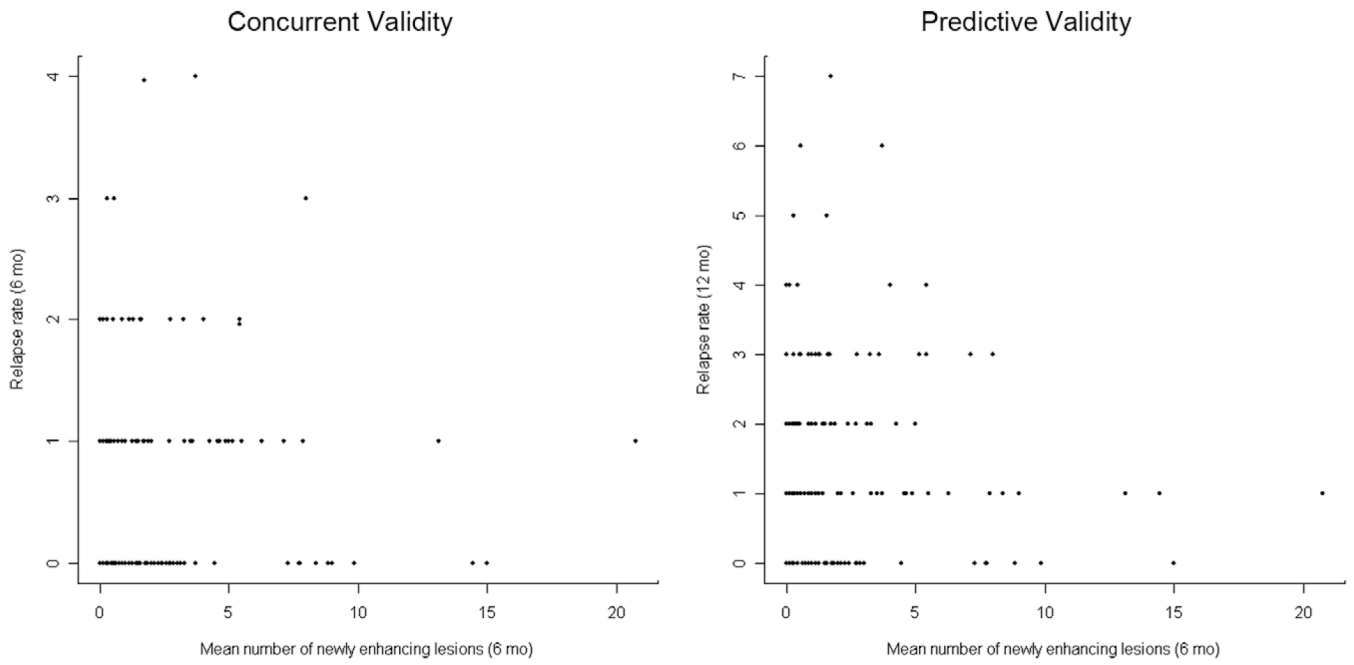


Figure 3. Number of new contrast enhancing lesions per scan over 6 months for $n = 208$ patients from 11 clinical trials versus 6-month relapse rate (concurrent validity) and versus 12-month relapse rate (predictive validity).

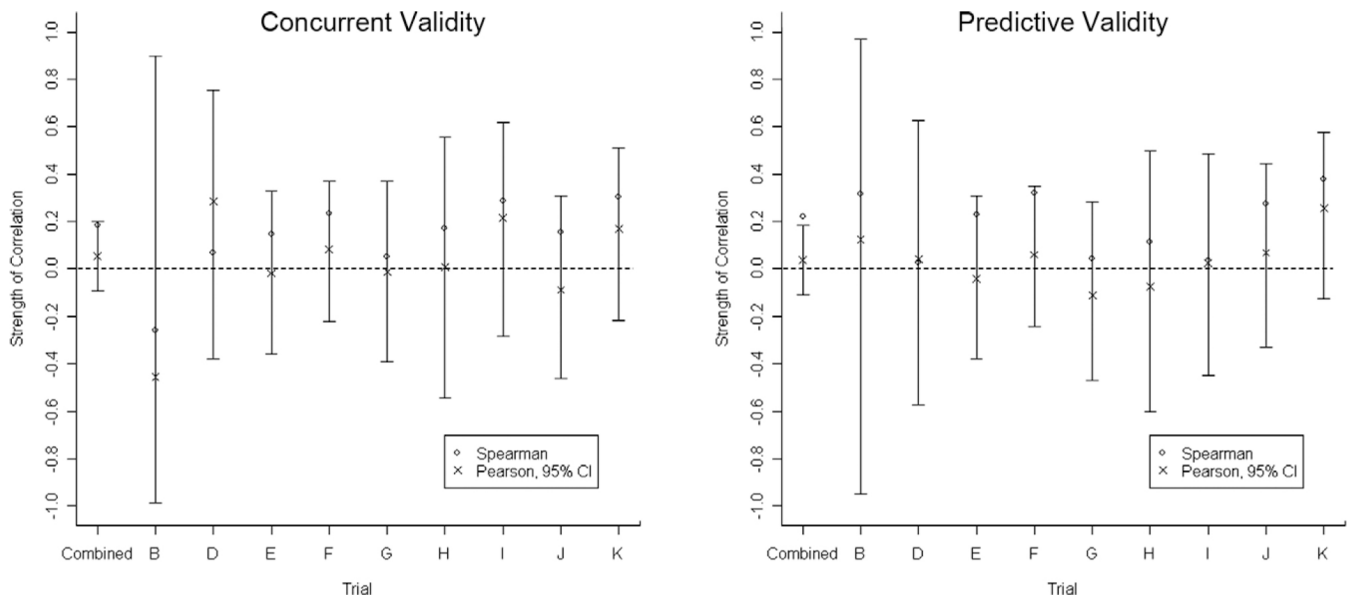


Figure 4. Graphical display of the correlation between new contrast enhancing lesions per scan and relapse rate for concurrent and predictive validity.

Table 1
Correlation between new contrast enhancing lesions per scan over 6 months and relapse rate over 6 months

Trial	n	Correlation (Spearman)	p-value (Spearman)	Correlation (Pearson)	95% CI (Pearson)
Overall	208	0.21	0.002	0.11	-0.03, 0.24
A	1	-	-	-	-
B	4	-0.26	> 0.20	-0.45	-0.99, 0.90
C	3	-	-	-	-
D	11	0.07	> 0.20	0.29	-0.38, 0.76
E	33	0.15	> 0.20	-0.02	-0.36, 0.33
F	44	0.25	0.10	0.09	-0.21, 0.38
G	27	0.05	> 0.20	-0.01	-0.39, 0.37
H	13	0.17	> 0.20	0.01	-0.54, 0.56
I	18	0.29	> 0.20	0.21	-0.28, 0.62
J	26	0.16	> 0.20	-0.09	-0.46, 0.31
K	28	0.30	0.12	0.17	-0.22, 0.51
Combined*	204	0.19	0.01	0.06	-0.09, 0.20

* Weighted mean correlation, trials A and C excluded (see text).

Table 2
Correlation between new contrast enhancing lesions per scan over 6 months and relapse rate over 12 months

Trial	n	Correlation (Spearman)	p-value (Spearman)	Correlation (Pearson)	95% CI (Pearson)
Overall	208	0.22	0.001	0.06	-0.08, 0.19
A	1	-	-	-	-
B	4	0.32	> 0.20	0.12	-0.95, 0.97
C	3	-	-	-	-
D	11	0.03	> 0.20	0.04	-0.57, 0.63
E	33	0.23	0.20	-0.04	-0.38, 0.31
F	44	0.33	0.03	0.06	-0.24, 0.35
G	27	0.04	> 0.20	-0.11	-0.47, 0.28
H	13	0.11	> 0.20	-0.07	-0.60, 0.50
I	18	0.04	> 0.20	0.03	-0.45, 0.49
J	26	0.28	0.17	0.07	-0.33, 0.44
K	28	0.38	0.05	0.26	-0.13, 0.58
Combined*	204	0.23	0.003	0.04	-0.11, 0.18

* Weighted mean correlation, trials A and C excluded (see text).