Introduction

Therapeutic trials in multiple sclerosis (MS) may be particularly vulnerable to negative results because the disease process is not fully understood and animal models may not adequately mimic MS [4, 21]. Many treatments applied for MS patients also have poorly defined mechanisms and are restricted to the immune system, which may lead to efficacy or toxicity in unpredictable ways. It is important to publish these failed trials so that they are available for the planning of future studies. This reinforces the need for thorough testing of new agents in an intelligent way before approving them for clinical use. But failed clinical studies can teach other lessons as well, in so far as unexpected results can promote new hypotheses about disease pathophysiology or prompt questions about an experimental agent’s proposed mechanism of action. In addition, negative trials can provide valuable information about study design and outcome measures for future trials.

A statistical approach: “Absence of evidence is not evidence of absence” [3]

The purpose of a clinical trial is to compare the estimate of the true response rate for a treatment regime under study with the estimate of response rate from a competing control therapy (in the MS field still often placebo treatment). The difference between response rates gives us an estimate of the effectiveness of the investigated therapy. Then, a statistical test is applied looking for sufficient reasons to reject the null hypothesis (“There is no treatment difference.”). The p value of this test gives us the probability that a difference as large as the one actually observed could occur by chance alone if the null hypothesis was true. The arbitrary level of significance which determines acceptance or rejection of the null hypothesis is conventionally set at 5%; for example, given a trial outcome for a beneficial new MS treatment with a p value of 0.05, there is a 1 in 20 risk that we would erroneously claim a real treatment difference. This level at which we accept or reject the null hypothesis is known as the type I (or α) error and is the probability of falsely rejecting a true null hypothesis.
claiming a real treatment effect when in fact there is none.

The converse of type I error is another equally important, but less understood risk – that of falsely claiming no treatment difference of a specified magnitude, when in fact a real treatment difference exists. The probability of such a false negative conclusion is known as type II (or β) error. The probability of type II error is not a single value treatment effect; very few studies especially in the MS area run a high risk of missing very large treatment differences such as 100%, but only very large studies have an acceptable risk (e.g. less than 10%) of missing a treatment difference of 10%. Many clinical trials are insufficiently powered so that they cannot detect any difference even if a real difference exists. In statistics jargon, the trial is “insufficiently powered”. The power of a trial is the probability that it can detect the difference if the difference really exists. The relationship of the null hypothesis to type I and type II error is summarized in Table 1.

In this setting, it is necessary to distinguish between studies that have meaningful but negative results, and those that are unable to be interpreted because of methodological flaws. Failure to report all meaningful trials devalues the participation of patients and investigators, and ignores the many lessons such trials can teach. Even though negative trials may not be the final step toward a new treatment, each and every carefully performed clinical trial brings us closer to a better understanding of the disease process, potential treatments, and ways to evaluate them.

In addition, negative trials can provide valuable information about study design and outcome measures for future trials. For example, a multi-center trial of 4-aminopyridine, which can improve conduction of action potentials through partially demyelinated axons, failed to demonstrate an effect in patients using the standard measure of MS impairment, the Expanded Disability Status Scale (EDSS). Subsequent studies have suggested that more sensitive measures of neurologic function may be more useful in this setting [18]. A pivotal trial of cladribine, in which the placebo group failed to worsen during the 3-year-study period, demonstrated that the EDSS may also be unacceptably insensitive to progression of MS impairment in trials of similar duration [17].

Many negative trials are not published either having been rejected or the authors have abandoned them [9, 10]. This gives rise to serious publication bias in meta-analysis. Some of the important findings can be buried under the unpublished negative trials.

### Experimental autoimmune encephalomyelitis (EAE): The animal model of MS, but not quite MS itself

To date, treatments for MS have been developed with the intention of intervening parts of certain autoimmune responses in the periphery. This approach is hampered by limited knowledge of the pathogenic cascade in human MS, which compromises the development of rationally designed immune treatments. Because of the limited pathogenetic data in MS, nearly all studies of agents targeting the disease have been performed not in the human disease but in animal models. In particular, the experimental autoimmune encephalomyelitis (EAE) model has been used widely to evaluate immunomodulatory treatments for multiple sclerosis as well as to explore the pathophysiology of the disease [8]. This model involves the generation of an autoimmune response in the immunological periphery by immunizing different animals with myelin proteins such as myelin basic protein (MBP) [25]. These animals develop a clinical and pathological pattern which is quite different to human MS (spinal cord lesions in EAE) although several histopathological hallmarks of MS, including focal inflammatory lesions in the nervous system can be found in theses models (Table 2).

When obtaining more and more scientific data from animal experiments, it is important to keep in mind that EAE is not human multiple sclerosis [24]. In fact, each EAE experiment only represents a small part of the still unknown pathogenetic cascade of autoimmune demyelination. It is perhaps for this reason that divergent results have been observed for a number of treatments assessed in both the EAE model and in clinical trials in multiple sclerosis (Table 2). Despite its shortcomings, EAE has been invaluable in elucidating some aspects of human MS and in the investigation of novel therapeutic approaches. Not only are all approved MS medications effective in EAE, but this model has been crucial in the

---

**Table 1** “The Truth”: type I and type II errors in clinical studies

<table>
<thead>
<tr>
<th>Actual result of significance test</th>
<th>Null hypothesis true, i.e., no difference between treatments</th>
<th>Null hypothesis false, i.e., difference between treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do not reject null hypothesis, i.e., “not statistically significant P &gt; 0.05”</td>
<td>Correct conclusion</td>
<td>Type II error, i.e., falsely claiming equivalence of treatments</td>
</tr>
<tr>
<td>Reject null hypothesis, i.e., “a significant treatment difference P &lt; 0.05”</td>
<td>Type I error, i.e., falsely claiming a real treatment difference</td>
<td>Correct conclusion</td>
</tr>
</tbody>
</table>