DOI: 10.1093/brain/awh176 *Brain* (2004), **127,** 1463–1478

REVIEW ARTICLE

Development of biomarkers in multiple sclerosis

Bibiana Bielekova and Roland Martin

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

Correspondence to: Dr Bibiana Bielekova, Neuroimmunology Branch, NINDS, National Institutes of Health, Bldg. 10, Room 5B-16, 10 Center DR MSC 1400, Bethesda, MD 20892-1400, USA
E-mail: bielekob@ninds.nih.gov

Summary

Multiple sclerosis is a complex disease, as several pathophysiological processes (including inflammation, demyelination, axonal damage and repair mechanisms) participate in the disease process. Furthermore, as new pathological evidence reveals, these processes are not uniformly represented across patient populations but can selectively predominate in individual patients, thus contributing to the heterogeneity in phenotypic expression of the disease, its prognosis and response to therapies. While the armamentarium of available therapies for multiple sclerosis broadens, little is known about factors that predict treatment response in individual patients to a specific drug. More importantly, we are beginning to understand that, analogous to cancer therapy, the successful therapeutic strategy in multiple sclerosis might ultimately involve the combination of different therapeutics targeting several dominant pathophysiological processes. The development of these process-specific therapies will be impossible without the use of biomarkers that reflect the targeted process, can select patient population in which the targeted process is prevailing and can aid during the more rapid screening of therapeutic agents in the early phase of their development. This review summarizes the general concepts of biomarkers and their potential use as surrogate endpoints and tailors these concepts to specific applications in multiple sclerosis research.

Keywords: multiple sclerosis; surrogate markers; autoimmunity; immunological biomarkers

Abbreviations: MBP = myelin basic protein; PCR = polymerase chain reaction; RT-PCR = reverse-transcription polymerase chain reaction; WMLL = white matter lesion load

Received January 16, 2004. Revised February 19, 2004. Accepted February 26, 2004. Advance Access publication June 4, 2004

Contents

Definitions of 'biomarker' and criteria for the use of biomarkers as surrogate endpoints

In recent years we have observed a vast expansion of the biomedical scientific literature in which the terms 'biomarker' and 'surrogate marker' are used (1998). These terms have often been used loosely, at times interchangeably, creating controversy surrounding the use of biomarkers as surrogate endpoints. Scientists and regulatory agencies have made joint attempts to clarifying this ambiguous terminology (2001; Lesko and Atkinson, 2001; Rolan *et al.*, 2003) and we adopt the following working definitions in this review.

A *biomarker* is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention.

A *type 0 biomarker* is a marker of the natural history of a disease and correlates longitudinally with known clinical indices.

A *type I biomarker* captures the effects of a therapeutic intervention in accordance with its mechanism of action.

A *surrogate endpoint* is a biomarker that is intended to serve as a substitute for a clinically meaningful endpoint and is expected to predict the effect of a therapeutic intervention. The use of the term 'surrogate marker' is discouraged.

A *clinical endpoint* is a clinically meaningful measure of how a patient feels, functions or survives. Clinical endpoints may be further classified as *intermediate endpoints*, which are clinical endpoints that are not the ultimate outcome but are nonetheless of real clinical usefulness [e.g. exacerbation rate in relapsing–remitting MS (RR-MS)]; and *ultimate clinical outcomes*, which are clinical endpoints reflective of the accumulation of irreversible morbidity and survival (e.g. accumulation of irreversible disability in multiple sclerosis).

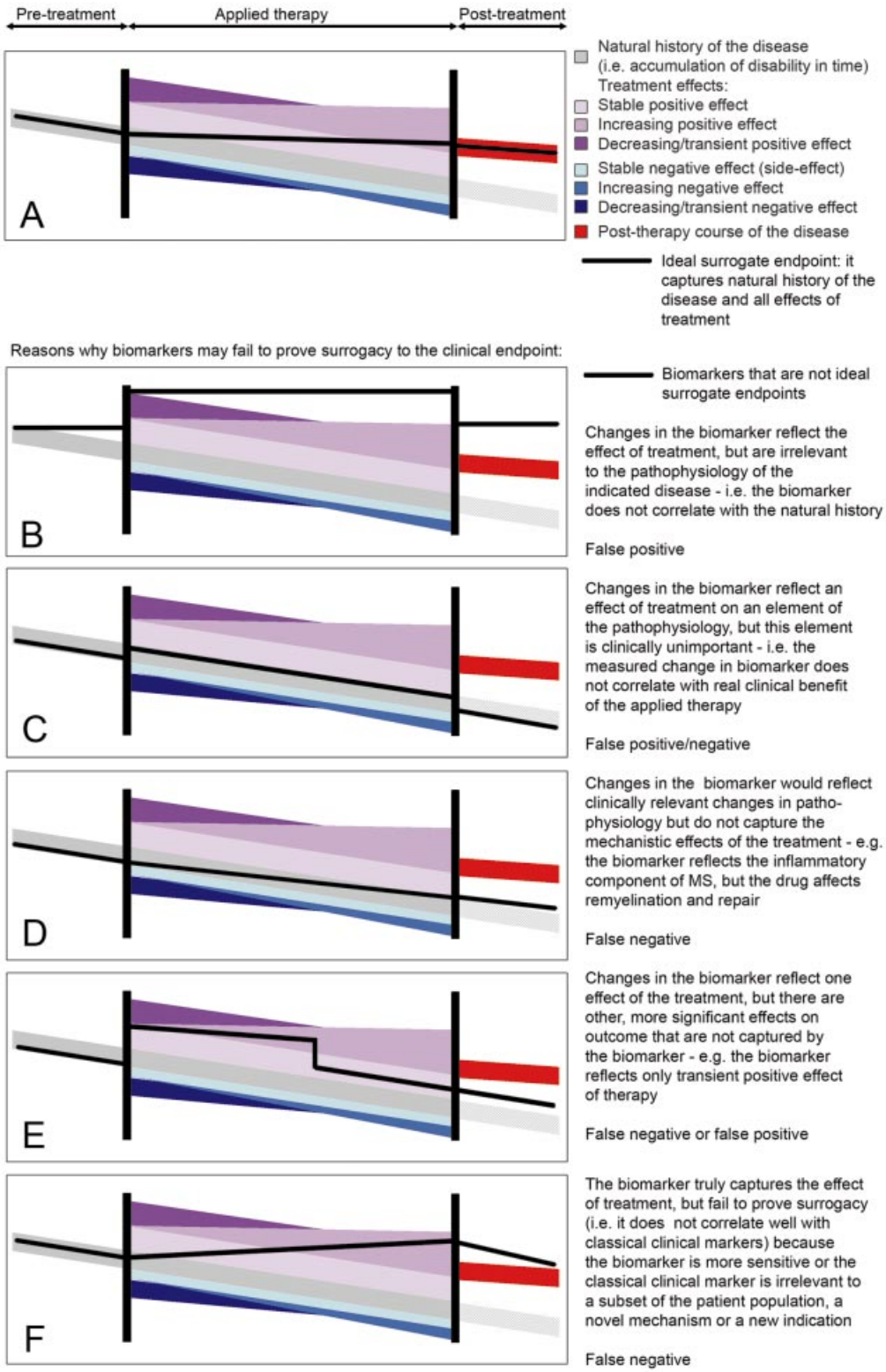
These definitions indicate a clear hierarchical distinction between biomarkers and surrogate endpoints. While numerous laboratory markers may be associated with a particular disease state, the term 'surrogate' indicates the ability of a biomarker to provide information about the clinical prognosis or efficacy of a therapy. The word 'surrogate' implies a strong correlation with a clinical endpoint, but in order to be clinically useful a surrogate must provide information about prognosis or therapeutic efficacy in a significantly shorter time than would be needed by following the clinical endpoint.

Numerous criteria for the validation of the surrogacy of a biomarker have been proposed, both by statisticians and by clinicians/regulatory agencies (De Gruttola *et al.*, 1997; Buyse *et al.*, 2000; Hughes, 2002; Molenberghs *et al.*, 2002). Prentice defined two conditions that together are sufficient to ensure the surrogacy of a biomarker (Prentice, 1989). The first requirement is a strong and significant correlation between the biomarker and the clinical endpoint. The second condition requires that the biomarker fully capture the net effect of the treatment on the true clinical endpoint. In this context, 'net effect' is defined as the cumulative effect accounting for all mechanisms of action, both beneficial and harmful. As depicted schematically in Fig. 1A, it becomes obvious that the second of Prentice's conditions is extremely restrictive, and it represents the main reason for the failure of biomarkers to prove their surrogacy. Even if the biomarker correlates extremely well with clinical markers in natural history studies (depicted as a thick grey line showing the accumulation of clinical disability over time), in order to prove its surrogacy it would have to reflect collectively all positive (beneficial) and negative (adverse) affects of the applied therapy. Clearly, this is virtually impossible, because some of the adverse effects of therapy may be completely unrelated to the pathophysiology of the disease and yet may negatively influence the ultimate clinical endpoint [e.g. cardiotoxicity of linomide or mitoxantrone in multiple sclerosis (Noseworthy *et al.*, 2000; Ghalie *et al.*, 2002)]. In multiple sclerosis the situation is further complicated by the fact that the disease pathophysiology is complex and the applied therapy may positively influence only one of the contributing processes (e.g. effect of immunosuppressive therapies on inflammation) and have no effect, or potentially have even negative influence on others. Understanding this limitation, the Food and Drug Administration (FDA) proposed a somewhat less restrictive criterion for considering evidence based on the use of biomarkers for drug approval purposes, where the second of Prentice's conditions is not reinforced if the biomarker is reasonably likely, on epidemiological, therapeutic, pathophysiological or other evidence, to predict clinical benefit (1997). However, approval under this section is subject to the requirement that the applicant studies the drug further to verify and describe its clinical benefit in adequate and well-controlled post-marketing studies (1997).

There are other reasons why biomarkers may fail to prove surrogacy to the clinical endpoint and these have been described previously (De Gruttola *et al.*, 1997; Frank and Hargreaves, 2003) (Fig. 1B–F). These reasons, leading to either false-negative or false-positive results, should always be kept in mind and tested for (if feasible) before the biomarker is selected for testing of surrogacy in large clinical trials. The two most obvious false-positive results are depicted in Fig. 1B and C. In both cases the changes in the biomarker reflect the effect of treatment, but these effects are either irrelevant to the pathophysiology of the disease (the

Fig. 1 Relationship between biomarker and surrogate endpoint in therapeutic trials. (**A**) This scheme depicts the development of clinical disability during the natural history of the disease (the grey bar reflects the accumulation of clinical disability over time). The progression of disability may be altered during a therapeutic trial, depending on the ratio of all positive (beneficial) and negative (side-effects or deleterious) effects of the applied therapy on the clinical parameters. Both positive (shades of purple) and negative (shades of blue) treatment effects may be stable during the whole therapy period, may be transient (i.e. decreasing during therapy) or may increase with the duration of treatment. The resulting relationship between all positive and negative effects of treatment will determine the extent of therapeutic benefit, depicted as a divergence between the natural history of the disease and the new course of the disease (red bar) and its duration (slope of the red bar). The ideal surrogate endpoint correlates well with accumulation of disability in the natural history of the disease and in addition captures all (positive and negative) effects of treatment. Furthermore, the sensitivity of the biomarker measure must be significantly greater than the sensitivity of clinical measurement (depicted as a thinner line for biomarkers and a thicker line for the clinical end-point), so that the biomarker measurement allows evaluation of the treatment effect in a significantly shorter time than would be required by following the clinical endpoint. (**B–F**) These five schemes depict the instances in which the evaluation of a biomarker (thick black line) in relation to clinical marker would lead to erroneous conclusions about the surrogacy of the biomarker.

biomarker does not correlate with clinical outcome in natural history; Fig. 1B) or are clinically unimportant (the measurable change in the biomarker is not paralleled by the measurable change in the clinical endpoint; Fig. 1C). Such false-positive surrogate markers are likely to emerge from the molecular/immunological studies supplementing therapeutic trials, which identify biomarkers reflective of the therapeutic intervention and possibly even of therapeutic efficacy.



However, their correlation with the clinical outcome that reflects the disease state still needs to be verified. Examples of such biomarkers in multiple sclerosis is the use of MxA expression for determining the biological effect of interferon-β (IFN-β) *in vivo* (Wandinger *et al.*, 2001), and a recently published study of using TNF-related apoptosis inducing ligand (TRAIL) as a biomarker of IFN-β therapeutic efficacy (Wandinger *et al.*, 2003). Regardless how promising these results may appear, we cannot assume surrogacy of any of these biomarkers without first demonstrating their strong correlation with the clinical outcome in an untreated multiple sclerosis population. Another potential false-positive result can emerge from a situation in which the changes in the biomarker reflect one effect of the treatment, but there are other, more significant effects on outcome that are not captured by the biomarker (Fig. 1E). Actually, this situation can cause both false-positive and false-negative results depending on whether the more profound effects on outcome (not captured by the biomarker) are predominantly negative or positive. In multiple sclerosis, the use of MRI measures in the linomide (Roquinimex) trial (Noseworthy *et al.*, 2000) serves as an example of a false-positive result: while linomide therapy had a positive effect on brain contrast-enhancing lesions and on a composite MRI measure (Wolinsky *et al.*, 2000), this marker did not reflect the unanticipated cardiopulmonary toxicities that resulted in a poor clinical outcome in the treatment group and eventually led to premature termination of the trial (Noseworthy *et al.*, 2000). Similarly, we envisage that the putative biomarker could be significantly influenced by a transient negative effect of the therapy; the net therapeutic effect on the biomarker may thus be negative, despite the fact that the long-term clinical outcome (after cessation of the transient side-effect that was not clinically relevant) may be favourable (false-negative result). Another two causes of false-negative outcomes in the use of biomarkers as surrogate endpoints (Fig. 1D and F) are very pertinent to multiple sclerosis, because it is a complex disease with multiple contributing pathophysiological mechanisms. A biomarker can lead to false-negative results in surrogacy testing if it reflects clinically relevant changes in pathophysiology but would not capture the mechanistic effects of the treatment applied (Fig. 1D). For example, MRI measures of brain inflammation and accumulation of white matter lesion load (WMLL) cannot serve as useful biomarkers in measuring the therapeutic effect of a drug that has positive effects on remyelination and repair in multiple sclerosis. Finally, the biomarker might not correlate well with classical clinical markers if the biomarker is more sensitive than the clinical marker, or if the clinical marker is irrelevant to a subset of the patient population, to a novel mechanism of action of the drug or to its new indication (Fig. 1F). Again, we need to ask how sensitive are our currently used clinical markers in depicting changes in the extent of remyelination and repair/scar formation after clinical attacks, especially if used short term. It is more than likely that a molecular/immunological biomarker reflecting the mechanism of action may be much more sensitive in this regard.

Although we conclude that, due to the complexity of multiple sclerosis pathophysiology, it is highly unlikely that we will ever find a single biomarker in multiple sclerosis that completely substitutes for a meaningful clinical outcome, we believe that it is exactly for this reason (the complexity of this disease) that the development of biomarkers is essential for future drug development in multiple sclerosis and a better understanding of disease pathogenesis.

Use of the biomarkers for drug development in multiple sclerosis

The current costs of drug development are extraordinary, especially in the final stages of clinical testing. Drug companies as well as regulatory agencies realize that faster screening of prospective therapies and earlier elimination of unpromising agents can substantially speed up drug development while limiting the associated cost. Moreover, the mechanism of action of many of the approved therapies is not fully established at the time when they go into clinical testing, or even at the time of their approval and widespread clinical use. This precludes the rational selection and design of combination therapies, and it renders the choice of the most efficacious therapy for a particular patient a guessing game for the treating physician. Wider employment of candidate surrogate markers during Phase I/II of clinical testing would lead to a better understanding of the mechanism of action of the tested drug, may help to stratify patients in terms of the potential effectiveness of an applied therapy, and ultimately will teach us important lessons about disease pathogenesis and patient heterogeneity.

In diseases with a complex pathogenesis, such as multiple sclerosis, an individual biomarker is likely to reflect only one of many ongoing pathogenic processes. As we pointed out above, this process-specificity of a biomarker is a reason why it cannot serve as a surrogate for a clinical outcome in a complex disease. However, this relationship is mutual; it also implies that the clinical outcome, which is reflective of all contributory pathophysiological processes, is too insensitive to capture the full effect of any process-specific therapeutic application. The vast majority, if not all, currently approved therapies for multiple sclerosis target only the inflammatory component of multiple sclerosis pathophysiology (Bielekova and Martin, 1999); but it is becoming exceedingly clear that, if we want to successfully treat multiple sclerosis at later stages we will need to add treatments that affect other mechanisms, especially remyelination and repair. Furthermore, elegant pathological studies imply that, for some multiple sclerosis patients, the inflammatory component of the disease is less prominent and other factors, such as excitotoxicity or neurodegeneration might be the primary events for the accumulation of disability (Lucchinetti *et al.*, 2000; Pitt *et al.*, 2000). Subsequently, there is a strong

Table 1 *The role of biomarkers in the development of new treatments for multiple sclerosis*

Screening candidate agents for the desired mechanism of action (preclinical, *in vitro* and animal studies)
Facilitation of the development of novel therapeutic compounds (by using techniques of combinatorial chemistry) (preclinical phases)
Screening candidate agents for drug toxicity (preclinical and clinical phases)
Guide to dose selection and escalation (Phase I/II trials) and for addressing regulatory concerns related to dose exposure–response relationship
Use of pharmacokinetic–pharmacodynamic relationship as a strategy to bridge different patient populations (e.g. paediatric, geriatric, patients with liver or renal diseases) or different formulations (new dosing intervals, new routes of administration) (Phase II/III and post-marketing)
Earlier elimination of unpromising drugs (Phase I/II trials)
Better and earlier differentiation of superior agents (Phase II trials)
Faster and more compelling confirmation of efficacy (Phase III trials)
Better understanding of the mechanism of action of the tested drug by confirmation of the presumed mechanisms that were established during preclinical studies
Identification of patients who will respond to the specific intervention in clinical practice
Providing a rationale for combination therapies and allowing the testing of therapeutic compatibilities between the two agents (preclinical and Phase I/II)

need for developing new therapies in multiple sclerosis that are more process-specific and can be used in specific patient subpopulations and possibly as sequential or simultaneous combination therapies. This important new direction in multiple sclerosis research will be unfeasible without the development of process-specific biomarkers. We performed a systematic analysis of all studies published within the last 20 years that examined biomarkers in multiple sclerosis, and concluded that none of the proposed biomarkers can serve as surrogate for clinical outcomes. This extensive work is clearly beyond the scope of the present review. However, perhaps even more importantly, more than 95% of the proposed biomarkers merely reflect different components of the inflammatory phase of the disease and there is a remarkable lack of suitable biomarker candidates reflective of other pathophysiological mechanisms. The reason for this lack of process-specific biomarkers is perhaps not too surprising: it is impossible to tease out specific pathogenetic processes in an unseparated patient population unless one studies very large cohorts of patients, which is not feasible in a disease as rare as multiple sclerosis at any single centre. Therefore, there are two possible ways to proceed. On the one hand, new candidate biomarkers in multiple sclerosis could emerge from the application of novel unbiased discovery tools (such as gene expression profiling, proteomics, metabolomics and biochemical profiling) in large patient cohorts from multicentre trials; on the other hand, hypothesis-based candidate biomarkers could be tested for their process-specificity in carefully stratified patient populations (Lassmann *et al.*, 2003). In both instances this effort will require close cooperation among scientists from all areas of multiple sclerosis research (i.e. clinicians with pathologists, MRI specialists and molecular biologists) to ensure that the patient stratifications from one area of research, such as pathology (Lucchinetti *et al.*, 2000), can be correlated with and enhanced by all other contributing specialities, such as MRI–pathological correlations (Barkhof and van Walderveen, 1999). In this regard, we have recently proposed a MRI-based classification of multiple sclerosis patients that is stable over time, has been validated in an independent cohort, and subdivides multiple sclerosis patients into four clinically meaningful subgroups (Bielekova *et al.*, submitted). We believe that this classification may be the first step in this multifaceted approach to the examination of disease heterogeneity in multiple sclerosis and will subsequently facilitate the development and testing of process-specific therapies.

The wide spectrum of potential roles that biomarkers will likely play in the future development of therapeutic applications in multiple sclerosis is summarized in Table 1.

Proposed classification of process-specific biomarkers in multiple sclerosis

The classification of process-specific biomarkers in multiple sclerosis has to be based on careful evaluation of all contributing pathophysiological processes. On the basis of analysis of the published studies examining pathophysiological mechanisms in multiple sclerosis (Trapp *et al.*, 1998; Genain *et al.*, 1999; Pitt *et al.*, 2000; John *et al.*, 2002; Lassmann, 2002, 1997; Lucchinetti *et al.*, 2000) we were able to classify all biomarkers proposed so far in multiple sclerosis into one of these seven categories:

(i) Biomarkers reflecting alteration of the immune system:
   (a) Cytokines and their receptors
   (b) Chemokines and their receptors
   (c) Antibodies
   (d) Complement-related biomarkers
   (e) Adhesion molecules

(f) Biomarkers reflective of antigen processing and presentation
(g) Other activation markers
(h) Cell cycle and apoptosis-related biomarkers
(i) Markers reflective of immune-mediated neuroprotection
(j) Changes in cellular subpopulations
(k) Functional assays for immunological reactivity

(ii) Biomarkers of blood–brain barrier (BBB) disruption
(iii) Biomarkers of demyelination
(iv) Biomarkers of oxidative stress and excitotoxicity
(v) Biomarkers of axonal/neuronal damage
(vi) Biomarkers of gliosis
(vii) Biomarkers of remyelination and repair

We will use this classification when referring to methodological considerations in the measurement of immunological biomarkers and to the decision-making process for the selection of biomarkers for further development in multiple sclerosis.

Methodological considerations in the measurement of immunological biomarkers

The requirements for a biomarker, i.e. validity, specificity and ease of measurement, will be quite different depending on its intended use [e.g. preclinical versus early clinical (Phase I/II) and late clinical (Phase III) stages of drug development]. Here, we briefly review the methodological considerations for the usefulness and applicability of a specific biomarker in multiple sclerosis.

(i) Markers indicative of different pathophysiological processes in multiple sclerosis

(a) *Markers of immunological activation* are likely to fluctuate with the relapsing–remitting course of the disease (similar to contrast-enhancing MRI activity), and therefore frequent serial measurements and the use of period averages (baseline versus treatment) may be necessary for the documentation of co-fluctuation with disease activity and a treatment effect on this type of biomarker. Moreover, markers of immunological activation are significantly influenced by ongoing infectious processes and by circadian and menstrual cycles, and many are also affected by age and sex. Therefore, every effort should be made to standardize specimen collection in terms of collection time and time to processing, and to screen out samples that may have been influenced by, for example, an infectious process. Due to multiple intertwined regulatory loops that are effective *in vivo*, the question of the clinical relevance of a change in a biomarker always emerges. This notion applies particularly to biomarkers that measure the effect of *in vivo* therapy, which will have much broader effects than those anticipated from *in vitro* testing. From a practical point of view, only biomarkers that can be obtained by non-invasive or minimally invasive procedures and evaluated by relatively simple and reliable assays may be considered as useful biomarkers from this group.

(b) *Markers of demyelination, axonal damage, oxidative stress, gliosis and remyelination* would be extremely valuable, since, based on experience with MRI markers of axonal damage (Barkhof and van Walderveen, 1999), they may correlate better with the development of long-term disability, may have higher prognostic value and may significantly enhance our understanding of the mechanism of action of tested therapies. These markers can be conceptually divided into positive and negative biomarkers. *Positive biomarkers* are biomarkers that are not produced (or occur only at very low levels) under physiological conditions, and are induced by the specific pathophysiological process in question. An example of such a marker would be glial fibrillary acidic protein (GFAP) in the CSF. These markers will be less desirable as surrogate endpoints because they are likely to fluctuate with disease activity and may not reliably represent the accumulated damage. *Negative biomarkers*, on the other hand, are produced under physiological conditions by intact tissues that are targeted by the disease process (e.g. insulin or pro-insulin in diabetes mellitus). Therefore, their decline will correspond to the accumulated damage in the target structures (oligodendrocytes, myelin and neurons in multiple sclerosis). These biomarkers are ideal, as they should correlate with the overall disability and can be assayed at two time-points: before and after therapy. Even biomarkers that must be obtained by invasive procedures only (like lumbar puncture in multiple sclerosis) should be considered as useful surrogate endpoints from this category, especially for use in Phase I/II clinical trials.

(ii) Type of specimen collected

(a) Urine

Advantages. The main advantage lies in its non-invasive collection and ability to be stored at the patient's home (–20°C). Collection of several hours' urine production (the first morning urine collects a substance excreted during the night) can to some extent eliminate diurnal variation. The use of the biomarker/creatinine ratio may eliminate the need for 24 h urine collection (the biomarker/albumin ratio can be used to study the excretion of proteins and peptides that undergo significant tubular reabsorption). Due to lower creatinine excretion in females, a correction factor (1.2) was proposed by some investigators (Giovannoni and Thompson, 1998).

Disadvantages. chronic urinary tract infections and asymptomatic bacterial colonization of the bladder, which are both common in more-disabled multiple sclerosis patients, can negatively affect the results. In addition, patients with bladder instability may regulate their fluid intake with resulting artificial hypohydration.

Currently measurable markers. Neopterin, nitrate and nitrite, prostaglandin metabolites, β_2-microglobulin, immunoglobulin (Ig) light chains, interleukin (IL)-1, IL-2, sIL-2R

(sCD25), IL-6, IL-8 and myelin basic protein (MBP)-like material

(b) Blood

Advantages. Blood is relatively simple to collect. The amount collected depends on the assay used and ranges from a few millilitres of whole blood/serum for PCR, ELISA and immunofluorescent studies to up to 120 ml, or even a leucocytapheresis, for extensive functional immunological studies.

Disadvantages. The main disadvantage is the diurnal variation of many soluble markers (e.g. IL-6 levels may vary as much as 350% during 24 h) and invasiveness of the collection. Levels of measured biomarkers are often affected by biological degradation in the liver or by excretion in the kidney. Separation techniques and storage (freezing) may affect the level of the studied marker, and therefore these have to be strictly standardized for assay reproducibility. Moreover, the time from collection to processing is absolutely critical, as delay in processing can significantly alter the level of measured biomarkers, both by degradation of some of them (e.g. by activating serum proteases) and by artificial induction of the expression of others (unpublished personal observations).

Currently measurable markers. Unseparated blood can be used for flow cytometry analysis of cellular subpopulations and their functional phenotypization, and for PCR studies. After coagulation, serum can be used for the measurement of soluble markers such as antibodies and cytokines. After separation procedures of uncoagulated samples, different cellular populations can be used for functional studies.

(c) Cerebrospinal fluid

Advantages. CSF may better reflect the relevant inflammatory process due to its proximity to inflammatory lesions in the CNS, although this notion remains controversial. Due to the flow pattern of CSF, it is unlikely that the CSF in the lumbar cistern accurately reflects the production of the inflammatory markers in the supratentorial region, where most of the multiple sclerosis-related inflammation occurs. In addition, the intraparenchymal extracellular space may not necessarily communicate with the free CSF space (Giovannoni *et al.*, 1998*a*). However, CSF collection does prevent biological degradation of excreted markers by the liver or by renal excretion.

Disadvantages. CSF collection is an invasive, although relatively benign, procedure, and therefore can be used only in the experimental setting for a selected patient population and sampled only for a limited number of time-points. CSF production is not homogeneous; maximum CSF production occurs around 2 a.m. (42 $\pm$ 2 ml/h) and minimum production around 6 p.m. (12 $\pm$ 7 ml/h) (Giovannoni *et al.*, 1998*a*), and therefore the time of collection should be standardized. It remains unclear whether CSF markers possess definite advantages over markers collected from blood and/or urine.

Currently measurable markers. Cell populations in the CSF sample can be used for flow cytometry analysis, PCR studies and cell functional studies. Separated CSF can be used for measurements of soluble markers.

(d) Tears

An interesting pivotal study suggested the detection of oligoclonal bands in the tears of multiple sclerosis patients (Liedtke *et al.*, 1992) and a follow-up study concluded that the specificity and sensitivity of such detection is similar to that of CSF (Devos *et al.*, 2001). Due to the uninvasiveness of such approach, it should be investigated further whether tears may serve as a valuable biological material for other biomarker measurements.

Types of assay used for the measurement of surrogate marker

(a) Enzyme-linked immunosorbent assay (ELISA)

ELISA is the most widely used assay for detecting soluble proteins in plasma. It has almost completely replaced the radioimmunoassay due to its lower cost and better safety profile. There are numerous commercially available kits for the detection of various proteins of interest and, following standardization protocols, the assays are easily reproducible in different laboratories. The other major advantage is the ability to use properly frozen and stored samples. The disadvantages are relatively low sensitivity and high cost. ELISA assays are useful for measurements of biomarkers in both Phase I/II and Phase III of clinical trials.

(b) Fluorescence immunoassays

Fluorescence-based immunoassays have, overall, a much broader potential than ELISA. They can detect proteins, peptides, haptens and antibodies with much higher sensitivity. Time-resolved fluorescence immunoassays are among the most sensitive assays available today with sensitivities in the range of 10^{-17} M. Fluorescence polarization immunoassays can reliably detect small peptides or haptens and fibre-optic fluoroimmunosensors enable multiple or continuous measurements of a biomarker *in vivo*. These assays are currently used widely in high-throughput screening in drug discovery, and it is very likely that they will replace the ELISA in clinical settings due to their significantly higher sensitivity and very good reproducibility. Frozen samples can be used for analysis and, like ELISA, these assays should be useful in both early and later phases of clinical testing.

(c) Flow cytometry-based analyses

Flow cytometry is a relatively simple and reproducible means of studying cell subpopulations in whole or separated blood samples and to study the cell surface expression of different markers of interest. Newer approaches, such as vital fluorescein dyes [e.g. 5- and 6-carboxyfluorescein diacetate, succinimidyl ester (CFSE)] and intracellular cytokine staining, enable us to obtain also functional data about specific cell populations. However, in our experience the handling of the sample (time from sample collection to processing, separation procedures, freezing, acquiring fresh or paraformaldehyde-fixed samples) can significantly influence the results, especially if one attempts to compare the mean fluorescence intensities (MFI) of measured markers. Mathematical modifications based on comparison of the MFIs of simultaneously acquired negative and positive controls can minimize at least some of these problems (Kraus *et al.*, 1998). Furthermore, these assays require specialized and expensive equipment. We believe that these assays are currently technically too tedious for widespread use in Phase III trials, but may be extremely useful for preclinical and Phase I/II trials.

(d) Polymerase chain reaction (PCR)

PCR-based techniques can reliably detect a single copy of genetic information. However, in order to be useful for the detection of surrogate markers, techniques need to be quantifiable. Quantitative PCR methods include non-competitive PCR, competitive PCR and real-time PCR (e.g. Taqman®). Of these, we believe that only real-time PCR has the potential for widespread use in clinical trials, due to its relative methodological simplicity, reproducibility and specificity. It is, however, very expensive. The greatest advantage of reverse transcription–PCR (RT-PCR) is the reliable and accurate analysis of minute amounts of mRNA and its ability to use properly frozen and stored samples. However, as with flow cytometry techniques, sample handling and delay in processing can have a significant influence on the results. Moreover, PCR-based methods exclude the detection of the post-translational modifications and regulatory mechanisms that may lead to decreased surface expression and/or function of the studied marker, and therefore the detected change may not be always translatable into functional relevance *in vivo*. As a result it is recommended that the functional relevance of the observed changes in PCR-based biomarker measurements should be assessed by different means (e.g. flow cytometry or functional assays). PCR techniques can be widely used across all stages of drug development.

(e) Cellular assays

Primary proliferation, IL-7-modified primary proliferation, split-well/limiting dilution analysis, ELISPOT. Most of these assays require relatively large quantities of blood and/or lymphocytapheresis, are time-consuming and are methodologically very challenging. The handling of the sample (delay in processing, separation techniques, freezing) may have significant effects on the results. Many of these techniques require stimulation with specific antigen, and the selection of the stimulating antigen/s, its concentration and the stimulation conditions all affect the data. Depending on the antigen used, the responses may or may not have pathophysiological relevance in a given patient. Although some of these methods have been used as biomarkers in previous therapeutic trials (Bielekova *et al.*, 2000) and although we believe that they have a firm place in the development of all antigen-specific immune interventions, they are limited to Phase I/II trials.

(f) New unbiased biomarker discovery techniques

As mentioned in the previous section, we believe that unbiased discovery tools, such as gene expression profiling, proteomics, metabolomics and biochemical profiling, which allow simultaneous analysis of the expression of thousands of genes, proteins or metabolites, will play a significant role in the development of novel candidate biomarkers in every disease process, including multiple sclerosis. Thorough description of these techniques is beyond the scope of this review. However, the discovery-type studies published so far in the field of multiple sclerosis (Wandinger *et al.*, 2001; Whitney *et al.*, 2001; Lock *et al.*, 2002; Maas *et al.*, 2002; Bomprezzi *et al.*, 2003; Mycko *et al.*, 2003) have revealed several problems that will need to be resolved before these techniques can yield more clinically applicable results. The first and most important drawback of currently published studies is the enormous number of candidate genes that have been selected by each of these studies; unfortunately, these are largely non-overlapping. Additionally, very few of the studies have attempted to distinguish between the disease state and the disease phenotype (Maas *et al.*, 2002). We believe that progress in this field is inevitable, but it will require the application of discovery tools in significantly larger cohorts of patients and appropriate controls in order to increase the specificity of selected candidate biomarkers and in order to start selecting biomarkers reflective of disease heterogeneity or of the treatment response (Sturzebecher *et al.*, 2003).

Decision-making in the selection of biomarkers for further development in multiple sclerosis

We have mentioned previously that we have undertaken a critical analysis of published studies in the field of multiple sclerosis in the past 20 years (between 1982 and 2002; based on a Medline search) that propose multiple sclerosis-related laboratory biomarkers. We deliberately omitted MRI and other imaging biomarkers from this review because of previously established consensus on their use in multiple

Table 2 *Simplified summary of reviewed biomarkers in multiple sclerosis*

Biomarker categories List of all evaluated biomarkers	Biomarkers with potential for further development	Biological rationale	Correlation with disease activity	Correlation with disability progression	Correlation with treatment effect	Notes[1]
Biomarkers reflecting alteration of the immune system						Unlikely candidates for surrogate endpoints; may prove useful in studying disease heterogeneity and in development of new therapies
(a) Cytokines and their receptors						The most extensively studied biomarkers in MS
IL-1, IL-2, IL-6, IL-10, IL-12, IL-18, TNF-α, LT-α/β, TGF-β, CD25	IL-6 (+sIL-6R and sgp130)	+++[2]	++/–	+	+/+	Candidate cytokine system linking innate immune system with both arms of adaptive immune responses (T and B cells)
	IL-10	++	++/–	+	+/+	Candidate immunoregulatory cytokine
	IL-12 (p70)/IL-23	+++	++	++	+/+/+	Suggested as biomarker that can differentiate between RR and SP-MS stages
(b) Chemokines and their receptors						Biomarkers that may aid in studying disease heterogeneity and proof-of-principle therapy trials
CCR5, CXCR3, CXCL10, CCR2/CCL2	CCR5	++		nd[3]	+/–	Suggested as a candidate biomarker of Th1 T cells
	CXCR3/CXCL10	++	++	nd	–	Marker of activated T cells
(c) Antibodies						The least systematically studied category with some interesting novel markers; e.g. diagnostic relevance of Ab in neuromyelitis optica. These biomarkers need systematic development and standardization of techniques
CSF IgG index, κ light chains, oligoclonal bands, anti-MBP Ab, anti-MOG Ab,	Anti-MBP and anti-MOG Ab	+++	nd	nd	nd	Suggested as a possible diagnostic tool for prediction of development of definite MS after first clinical symptom (CIDS)
(d) Complement-related biomarkers						Biomarkers needed for assessment of disease heterogeneity (based on pathological classification of MS lesions) and for development of novel therapies
C3, C4, activated neo-C9, Regulators of complement activation (CD35, CD59)	Activated neo-C9	++	+	+	nd	Biomarker reflecting formation of membrane-attack complex (MAC) that is expected to contribute to demyelination at least in a subgroup of MS patients
(e) Adhesion molecules						It is unlikely that these biomarkers would become more useful than MRI-based markers of BBB dysfunction
E-selectin, L-selectin, ICAM-1, VCAM-1, CD31, surface expression of LFA-1 and VLA-4						

Table 2 *Continued*

Biomarker categories		Biological rationale	Correlation with disease activity	Correlation with disability progression	Correlation with treatment effect	Notes[1]
List of all evaluated biomarkers	Biomarkers with potential for further development					
(f) Biomarkers reflective of antigen-processing and presentation						Very important category, little explored; needs further development in MS
CD40/CD40L, CD80, CD86, Heat shock proteins (hsp)	CD40/CD40L	++	+	nd	+	Suggested as candidate biomarker that can differentiate between RR and SP-MS stages
	hsp	+	nd	nd	nd	Dysregulation in the hsp system is the most prominent and consistent result of gene expression studies in MS and other autoimmune diseases
(g) Other activation markers						Markers reflecting activation of the innate immune system would contribute to studies of disease heterogeneity and aid in selection and screening of prospective novel immunomodulatory agents
CD26, CD30, CD71, perforin, OX-40 (CD134), osteopontin, macrophage-related proteins MRP-8 and MRP-16, neopterin, amyloid A protein, somatostatin	Neopterin	++	++	nd	+/–	
(h) Cell-cycle and apoptosis-related biomarkers						Very important category of biomarkers because they may reflect both defect in regulation of immune cells as well as proapoptotic properties of CNS components
Fas (CD95) and Fas-L, FLIP, Bcl-2, TRAIL	FLIP	++	+	+	+	Anti-apoptotic protein overexpressed in MS
	TRAIL	+/?	nd	nd	+	Suggested as biomarker reflective of clinical response to IFN-β therapy in MS
(i) Biomarkers reflective of immune-mediated neuroprotection						Potentially very interesting biomarkers that need to be developed further; would contribute to disease heterogeneity studies and to development of process-specific therapies
BDNF expression						
(j) Changes in cellular subpopulations						Markers studied predominantly in the past; many should be reassessed by new more precise techniques
NK cells, Vα24+ NKT cells, CD4+/CD25bright and IL-10-producing immunoregulatory T cells, CSF cells, CD45RA–/RO+/CD4+ (memory) T cells	CD4+/CD25bright T cells and IL-10-producing regulatory T cells, regulatory NK cells and NKT cells	+++	+	nd	+/nd	These cellular subpopulations were shown to have important immunoregulatory role in animal models and other human autoimmune disorders and they merit careful evaluation in MS

Table 2 *Continued*

Biomarker categories		Biological rationale	Correlation with disease activity	Correlation with disability progression	Correlation with treatment effect	Notes[1]
List of all evaluated biomarkers	Biomarkers with potential for further development					
(k) Functional assays for immunological reactivity						Although potentially very interesting, these assays are very tedious and therefore are likely to remain restricted to early phases of drug development and to proof-of-principle clinical trials
Proliferation assays (Ag-specific and polyclonal), cytokine-secretion assays, cytotoxic assays						
Biomarkers of BBB disruption						It is unlikely that these biomarkers would become more useful than MRI-based markers of BBB dysfunction
MMPs and their inhibitors (TIMP), platelet activating factor, thrombomodulin						
Biomarkers of demyelination						Would greatly enhance the understanding of MRI/pathological correlations and have a potential for partial surrogacy
MBP and MBP-like material, proteolytic enzymes, endogenous pentapeptide QYNAD, gliotoxin	QYNAD – endogenous peptide with Na-channel blocking properties	++	nd	nd	nd	QYNAD is an endogenous substance in CSF that probably originates from proteolytic cleavage during inflammatory process; deserves further evaluation
Biomarkers of oxidative stress and excitotoxicity						Very important biomarkers from the standpoint of disease heterogeneity and potentially for development of novel therapies; need to be developed further
NO and its stable metabolites (nitrite NO2– and nitrate NO3–), uric acid, isoprostane, marker for hypoxia-like tissue damage in MS	NO (+ NO2– and NO3–)	++	+/–	–/nd	nd	May help in disease heterogeneity studies
	Uric acid	++	++	+	+/–	Strong natural peroxynitrate scavenger
	Isoprostane	++	nd	+	+	Interesting candidate marker that merits further studies
	Marker for hypoxia-like tissue damage in MS	+	nd	nd	nd	Described in pivotal study as an endogenous epitope that is cross-recognized by mAb against canine distemper virus and may become a diagnostic tool to identify specific MS subtype
Biomarkers of axonal/neuronal damage						Most likely category of biomarkers with surrogate potential in MS
Cytoskeletal proteins (actin, tubulin and neurofilaments), Tau protein	Neurofilaments: light subunit (NF-L)	+++	+	+	nd	Might be the most likely candidate for surrogacy and its development warrants further efforts
	Tau protein	++	++	nd	nd	Also potentially very useful biomarker that needs further development

Table 2 *Continued*

Biomarker categories List of all evaluated biomarkers	Biomarkers with potential for further development	Biological rationale	Correlation with disease activity	Correlation with disability progression	Correlation with treatment effect	Notes[1]
Biomarkers of gliosis						May be useful for disease heterogeneity studies but with unpredictable surrogacy potential
GFAP, S-100 proteins						
Biomarkers of remyelination and repair						Much-needed biomarkers that would guide development of repair-promoting strategies in MS and aid in disease heterogeneity studies
NCAM, CNTF, microtubule-associated protein-2 exon 13 (MAP-2 + -13), protein 14-3-3, CPK-BB, peptidylglycine α-amidating monooxygenase (PAM), neural-specific enolase (NSE)	NCAM – neural cell adhesion molecule	+	+	nd	nd	Very sparse data on both biomarkers. However, because these are so far the only potential candidates, their evaluation warrants further effort
	CNTF – ciliary neurotrophic factor	++	nd	nd	nd	

[1]Brief opinion of the authors about the potential use of the biomarkers and the need for further developments. [2]We attempted to grade the strength of supportive evidence for each characteristic of the biomarker as low (+), medium (++) and high (+++). In treatment effects, +/– implies positive correlation with one type of therapy and negative with another. [3]No reliable data. Ab = antibody; Ag = antigen; APC = antigen-presenting cell; BBB = blood–brain barrier; CNTF = ciliary neurotrophic factor; CTL = cytotoxic T lymphocyte; EAE = experimental autoimmune encephalomyelitis; EDSS = Expanded Disability Status Scale; ELISA = enzyme-linked immunosorbent assay; FLIP = Fas-associated death domain-like interleukin-1β-converting enzyme inhibitory protein; GFAP = glial fibrillary acidic protein; hsp = heat shock protein; ICAM-1 = intracellular adhesion molecule-1; IFN = interferon; Ig = immunoglobulin; IL = interleukin; LAK cells = lymphocyte-activated killer cells; LP = lumbar puncture; LT-α/β = lymphotoxin α/β; MAC = membrane-attack complex; MBP = myelin basic protein; MOG = myelin oligodendrocyte glycoprotein; MMP = matrix metalloproteinase; MP = methylprednisolone; NCAM = neural cell adhesion molecule; NF-L = neurofilament light subunit; NK cells = natural killer cells; NKT cells = NK-like T-cells; NSE = neuron-specific enolase; OCB = oligoclonal bands; PAM = peptidylglycine α-amidating monooxygenase; PLP = proteolipid protein; PP-MS = primary progressive multiple sclerosis; RCA = regulators of complement activation; RIA = radioimmunoassay; RR-MS = relapsing–remitting multiple sclerosis; Scripps NRS = Scripps Neurological Status scale; SP-MS = secondary progressive multiple sclerosis; TGF-β = transforming growth factor β; TIMP = tissue inhibitor of matrix metalloproteinases; TNF-α = tumour necrosis factor α; VCAM-1 = vascular cell adhesion molecule-1; WBC = white blood cells.

sclerosis (for review see McFarland, 2002). In this section, we will summarize the criteria we used for the rating of published studies, the criteria that define the characteristics of biomarkers that underline their clinical utility, and the criteria required for biomarker validation; and in one simplified table (Table 2) we summarize the biomarkers that we considered in our review and choose those that, based on these extensive criteria, would in our opinion merit further development in multiple sclerosis.

Grading of the reviewed studies on the basis of recommendations of an evidence-based medicine working group

We graded each study that we reviewed according to the recommendations of the evidence-based medicine working group (Hayward *et al.*, 1995; Jaeschke *et al.*, 1994*a*, *b*). For each study we asked five questions:

(i) Are complete (raw) data provided?

(ii) Was there an independent comparison to a reference standard/ age- and sex-matched reference group?

(iii) Was an appropriate spectrum of patients included (clinically definite multiple sclerosis, different clinical subtypes, sample size)?

(iv) Were the methods used valid? (type of detection method, frozen versus fresh samples, intra-assay variation, detection limits, standardization of specimen collection, processing and storage, use of the appropriate statistical analysis)?

(v) Was there a processing and/or work-up bias (blinded processing and analysis)?

Studies that fulfilled all five criteria were considered grade A and those that fulfilled at least three criteria were considered grade B. All grade A and B studies were reviewed. All other studies were considered separately and were included only as preliminary study if they supported the biological rationale of the studied biomarker as a process-specific biomarker or as a potential surrogate endpoint.

Characteristics of a biomarker that underlie its potential clinical utility

The initial evaluation of a biomarker for its potential clinical utility can be based on determining how many of the characteristics of an ideal biomarker are met relative to the context of its use (Lesko and Atkinson, 2001). We considered the following criteria for the evaluation of biomarkers in multiple sclerosis:

(i) Biological rationale

The biomarker should have a rational association with a particular pathogenic aspect in multiple sclerosis. For example, GFAP, which is a structural protein of the glial intermediate filament and in its soluble form is secreted into the CSF, is considered a biomarker of astrogliosis in neurodegenerative disorders. It is also a main protein constituent of chronic multiple sclerosis plaques. Therefore, it is a biologically reasonable candidate biomarker of astrogliosis in multiple sclerosis.

(ii) Clinical relevance

If a biomarker is to serve as surrogate endpoint (or even as a partial surrogate endpoint) it has to have, in addition to the biological rationale, the potential for clinical relevance; i.e. the biomarker has to be positioned in the causal chain of pathological events leading to a meaningful clinical endpoint. For example, it is not known whether gliosis (proliferation and migration of glial cells in damaged areas of the CNS) observed in multiple sclerosis has overall beneficial (e.g. neuroprotective) or negative (e.g. preventing remyelination) effects on disease progression. Therefore, while GFAP may have a solid biological rationale, its clinical relevance in multiple sclerosis is currently unpredictable.

(iii) Practicality

The practical aspects of measurement of the biomarker (i.e. invasiveness of collection, need for serial analyses, the reproducibility, ease and cost of the assay) will to a large degree determine the context of its use (i.e. only exploratory/restricted use versus widespread clinical application).

(iv) Correlation with disease activity

Many of the published studies, especially those that investigated biomarkers of immune activation, related the studied biomarker to clinical or paraclinical measures of disease activity in multiple sclerosis. There are currently two types of accepted measures of disease activity in multiple sclerosis. The first is clinical: the number of multiple sclerosis relapses per patient/year or the extent of progression in one of the disability scales over a period of time (progression index). The second is paraclinical: the number of contrast-enhancing lesions on MRI of the brain and accumulation of disease burden measured as WMLL, accumulation of T1 hypointensities (black hole volume, BHV) and increases in brain atrophy measures. However, these MRI measures of disease activity do not necessarily correlate with accumulated disability (the ultimate clinical measure in multiple sclerosis) (Li *et al.*, 2001; Paolillo *et al.*, 1999), most likely because of the confounding effects of the locations and sizes of lesions, the severity of associated axonal damage, and the ability to remyelinate, which may be substantially different among individual patients. Several studies of correlation between immunological markers and measurements of MRI activity suggested that improving the quantification of the MRI disease burden, both inflammatory and total (Rieckmann *et al.*, 1997), considering the temporal profile of contrast-enhancing MRI lesions (Giovannoni *et al.*, 1997) and taking

into account anatomical relations of MRI lesions (Rieckmann *et al.*, 1997) may significantly enhance the correlation with immunological markers and possibly also with clinical endpoints (Giovannoni *et al.*, 1998*b*). Nevertheless, despite the current limitation of clinical and MRI measures of disease activity, they are widely used as outcomes in clinical trials, and the therapies that affect these measures may also demonstrate an effect on the long-term progression of disability (Jacobs *et al.*, 1996; Durelli *et al.*, 2002).

(v) Correlation with disability/prognosis

Ultimately, the correlation of a biomarker with the accumulation of clinical disability in time is the most important characteristic for determining biomarker surrogacy.

(vi) Correlation with treatment effect

Therapies influencing a biomarker should predictably influence a clinical endpoint. Additionally, therapies with the same mechanism of action should influence the biomarker congruently and comparatively with respect to their clinical effect. We considered reports of effects of biomarkers on all available therapies in multiple sclerosis (i.e. all formulations of IFN-β, glatiramer acetate, all formulations of glucocorticoids, and immunosuppressive agents, including mitoxantrone).

Evaluation criteria to define the clinical usefulness of a biomarker

The evaluation of a biomarker is a complex process and the criteria and the strength of supporting evidence will differ according to the intended use of the biomarker. Obviously, the highest level of stringency is required when a biomarker is evaluated for its surrogacy to a clinical endpoint. However, biomarkers intended for other purposes (e.g. as a diagnostic tool, for staging of the disease or as an indicator of disease prognosis, facilitating understanding of the disease mechanisms, selection for process-specific application, or probing for mechanisms of the treatment effect in proof-of-principle studies) do not need to be evaluated as rigorously and may still play very important roles. The different evaluation criteria for biomarkers (2001; Lesko and Atkinson, 2001) are summarized below.

(i) Sensitivity/specificity

These are standard statistical evaluations comparing a new marker (i.e. biomarker) with a gold standard marker (i.e. a clinical endpoint). Sensitivity refers to the ability to measure the biomarker with the magnitude of change that would be sufficient to reflect a clinically relevant endpoint (the incorrectly predicted values would be described as 'false negative'; i.e. the biomarker would not be able to distinguish a small but still meaningful change in clinical endpoint). Specificity refers to the accuracy with which the biomarker is able to distinguish correctly between the two clinically relevant endpoints (incorrectly predicted values would be described as 'false positive'; i.e. the biomarker would distinguish changes that are not reflected in a clinically meaningful endpoint).

(ii) Reliability

The reliability of a biomarker can be assessed in several ways: as the consistency of a measurement in time (biological variation), the consistency of a measurement dependent on an assay method (intra- and inter-assay variability) and the consistency of measurement based on the probability of false-positive and false-negative results.

(iii) Evaluation of a biomarker in epidemiological studies or natural history cohorts

These studies are essential for establishing the statistical relationship between the biomarker and the clinical endpoint under basal conditions. Such data may emerge from large-scale cross-sectional studies or, ideally, from longitudinal natural history studies. However, in view of the available partially effective therapies, the prospective use of long-term natural history cohorts has become ethically unsupportable at the present time in multiple sclerosis.

(iv) Evaluation of a biomarker in proof-of-principle clinical trials

Biomarkers that reflect the anticipated mechanism of action of the tested drug should in our opinion always accompany Phase I/II proof-of-principle trials, because their use can significantly enhance our understanding not only of drug action but also of the disease process. By using the precursor frequency of MBP(83–99)-specific T cells and altered-peptide ligand (APL)-specific T cells in a proof-of-principle therapeutic trial of an APL based on the MBP(83–99) sequence, we obtained invaluable information about the potential for cross-reactivity between the two epitopes that was completely unforeseen from *in vitro* data, and also established that MBP(83–99)-specific T cells indeed have encephalitogenic potential in humans under certain circumstances (Bielekova *et al.*, 2000). Without these data, we could today only hypothesize on the reason for the occurrence of atypical multiple sclerosis exacerbations in some patients during APL therapy (Bielekova *et al.*, 2000), and we would not be able to formulate the criteria for improving risk to efficacy ratio in the design of possible future APL-based therapies (Bielekova and Martin, 2001).

Evaluation of a biomarker in large multicentre double-blinded therapeutic trials

According to Hughes, only randomized controlled clinical trials can distinguish a surrogate endpoint from a mere marker of drug activity (Hughes *et al.*, 1995). In addition, such studies can estimate the proportion of a treatment effect that is accounted for by the biomarker/surrogate endpoint, and the evaluation of placebo arm can provide invaluable information about the surrogacy of the biomarker in an untreated population.

Evaluation of a biomarker in meta-analyses and mathematical modelling of the biomarker–clinical endpoint relationship

The ultimate estimation of the proportion of a treatment effect that is reflected by the biomarker comes from meta-analyses of treatment effects on the same biomarker in a series of trials in which different classes of therapies are evaluated using the same clinical endpoint (2001; Hughes *et al.*, 1995). Therefore, in order to validate the surrogacy of a biomarker in multiple sclerosis, investigators must collaborate in order to design large trials that incorporate marker evaluation and at the same time adhere to standardized methods that would allow potential surrogate endpoints to be evaluated in a meta-analysis.

Summary and conclusion

In the moment there is no biomarker available that fulfils the criteria of a surrogate endpoint in multiple sclerosis. Disappointing as this conclusion may be, it is also becoming clear that biomarkers will play a very important role in multiple sclerosis research and clinical practice in the future.

Due to their fluctuating nature, biomarkers of immune activation are unlikely to show surrogacy even after extensive additional studies. However, some of them may prove extremely useful in defining disease heterogeneity in multiple sclerosis and thus in stratifying patients into distinct subgroups. As suggested by pathological data, dysfunctional or imbalanced immune processes (e.g. complement activation, autoantibodies, cell-mediated lysis) may predominate in some patient subgroups. Here, the establishment of biomarkers that reflect these processes *in vivo* would be essential in studying disease heterogeneity and assessing how different phenotypes correlate with or allow the prediction of the response to novel process-specific therapies. Additionally, there is a great need to identify and develop novel candidate biomarkers in multiple sclerosis that would reflect other contributing pathophysiological mechanisms, *viz.* oxidative stress and excitotoxicity, demyelination and remyelination, gliosis, neurodegeneration and repair. These candidate biomarkers will probably evolve from the unbiased discovery techniques of gene expression profiling, proteomics and pharmacogenomics. These approaches offer the opportunity to progress from the examination of single genes or proteins to the study of entire systems, such as cell metabolism, expression and the regulation of proinflammatory proteins, apoptosis. Obviously, while such studies have only begun, they are probably appropriate in addressing the complex immune or tissue alterations in multifaceted diseases such as multiple sclerosis.

Clearly, more unified efforts are needed in the standardization of collection techniques and measurement methods, in order to allow comparison of the results from multiple studies. With the ethically imposed limitation on the study of untreated multiple sclerosis cohorts in the long term and the overall limit on patient resources, further biomarker research should be pursued in a coordinated, ideally international, effort similar to the one that has greatly advanced the use of MRI in multiple sclerosis. In our opinion, clinical trials should always be accompanied by biomarker studies oriented towards mechanisms of action. Investigators and sponsors should both understand that large multicentre clinical trials are becoming the only source of patient cohorts that are sufficiently large and defined to enable the evaluation of biomarkers for their surrogacy in multiple sclerosis, and that extra resources need to be allocated to biomarker discovery, since their evaluation will be crucial for the advancement of clinical and basic multiple sclerosis research.

References

Barkhof F, van Walderveen M. Characterization of tissue damage in multiple sclerosis by nuclear magnetic resonance. Philos Trans R Soc Lond B Biol Sci 1999; 354: 1675–86.

Bielekova B, Martin R. Multiple sclerosis: immunotherapy. Curr Treat Options Neurol 1999; 1: 201–20.

Bielekova B, Martin R. Antigen-specific immunomodulation via altered peptide ligands. J Mol Med 2001; 79: 552–65.

Bielekova B, Goodwin B, Richert N, Cortese I, Kondo T, Afshar G, et al. Encephalitogenic potential of the myelin basic protein peptide (amino acids 83–99) in multiple sclerosis: results of a phase II clinical trial with an altered peptide ligand. Nat Med 2000; 6: 1167–75.

Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001; 69: 89–95.

Bomprezzi R, Ringner M, Kim S, Bittner ML, Khan J, Chen Y, et al. Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. Hum Mol Genet 2003; 12: 2191–9.

Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000; 1: 49–67.

De Gruttola V, Fleming T, Lin DY, Coombs R. Perspective: validating surrogate markers—are we being naive? J Infect Dis 1997; 175: 237–46.

Devos D, Forzy G, de Seze J, Caillez S, Louchart P, Gallois P, et al. Silver stained isoelectrophoresis of tears and cerebrospinal fluid in multiple sclerosis. J Neurol 2001; 248: 672–5.

Durelli L, Verdun E, Barbero P, Bergui M, Versino E, Ghezzi A, et al. Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). Lancet 2002; 359: 1453–60.

Food and Drug Administration. Modernization Act of 1997. Title 21 Code of Federal Regulations Part 314 Subpart H Section 314.500, 1997.

Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. Nat Rev Drug Discov 2003; 2: 566–80.

Genain CP, Cannella B, Hauser SL, Raine CS. Identification of autoantibodies associated with myelin damage in multiple sclerosis. Nat Med 1999; 5: 170–5.

Ghalie RG, Edan G, Laurent M, Mauch E, Eisenman S, Hartung HP, et al. Cardiac adverse effects associated with mitoxantrone (Novantrone) therapy in patients with MS. Neurology 2002; 59: 909–13.

Giovannoni G, Thompson EJ. Urinary markers of disease activity in multiple sclerosis. Mult Scler 1998; 4: 247–53.

Giovannoni G, Lai M, Thorpe J, Kidd D, Chamoun V, Thompson AJ, et al. Longitudinal study of soluble adhesion molecules in multiple sclerosis: correlation with gadolinium enhanced magnetic resonance imaging. Neurology 1997; 48: 1557–65.

Giovannoni G, Green AJ, Thompson EJ. Are there any body fluid markers of brain atrophy in multiple sclerosis? Mult Scler 1998a; 4: 138–42.

Giovannoni G, Kieseier B, Hartung HP. Correlating immunological and magnetic resonance imaging markers of disease activity in multiple sclerosis. J Neurol Neurosurg Psychiatry 1998b; 64 Suppl 1: S31–6.

Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. JAMA 1995; 274: 570–4.

Hughes MD. Evaluating surrogate endpoints. Control Clin Trials 2002; 23: 703–7.

Hughes MD, De Gruttola V, Welles SL. Evaluating surrogate markers. J Acquir Immune Defic Syndr Hum Retrovirol 1995; 10 Suppl 2: S1–8.

Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. Ann Neurol 1996; 39: 285–94.

Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1994a; 271: 389–91.

Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994b; 271: 703–7.

John GR, Shankar SL, Shafit-Zagardo B, Massimi A, Lee SC, Raine CS, et al. Multiple sclerosis: re-expression of a developmental pathway that restricts oligodendrocyte maturation. Nat Med 2002; 8: 1115–21.

Kraus J, Oschmann P, Engelhardt B, Schiel C, Hornig C, Bauer R, et al. Soluble and cell surface ICAM-1 as markers for disease activity in multiple sclerosis. Acta Neurol Scand 1998; 98: 102–9.

Lassmann H. Mechanisms of demyelination and tissue destruction in multiple sclerosis. Clin Neurol Neurosurg 2002; 104: 168–71.

Lassmann H, Bruck W, Lucchinetti C, Rodriguez M. Remyelination in multiple sclerosis. Mult Scler 1997; 3: 133–6.

Lassmann H, Reindl M, Rauschka H, Berger J, Aboul-Enein F, Berger T, et al. A new paraclinical CSF marker for hypoxia-like tissue damage in multiple sclerosis lesions. Brain 2003; 126: 1347–57.

Lesko LJ, Atkinson AJ Jr. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol 2001; 41: 347–66.

Li DK, Zhao GJ, Paty DW. Randomized controlled trial of interferon-beta-1a in secondary progressive MS: MRI results. Neurology 2001; 56: 1505–13.

Liedtke W, Weller M, Wietholter H, Dichgans J. Immunological abnormalities in the tears of multiple sclerosis patients. Acta Neurol Scand 1992; 85: 228–30.

Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, et al. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. Nat Med 2002; 8: 500–8.

Lucchinetti C, Bruck W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. Ann Neurol 2000; 47: 707–17.

Maas K, Chan S, Parker J, Slater A, Moore J, Olsen N, et al. Cutting edge: molecular portrait of human autoimmune disease. J Immunol 2002; 169: 5–9.

McFarland HF, Barkhof F, Antel J, Miller DH. The role of MRI as a surrogate outcome measure in multiple sclerosis. Mult Scler 2002; 8: 40–50.

Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. Control Clin Trials 2002; 23: 607–25.

Mycko MP, Papoian R, Boschert U, Raine CS, Selmaj KW. cDNA microarray analysis in multiple sclerosis lesions: detection of genes associated with disease activity. Brain 2003; 126: 1048–57.

NIH-FDA Conference. Biomarkers and surrogate endpoints: advancing clinical research and applications. Abstracts. Dis Markers 1998; 14: 187–334.

Noseworthy JH, Wolinsky JS, Lublin FD, Whitaker JN, Linde A, Gjorstrup P, et al. Linomide in relapsing and secondary progressive MS: part I: trial design and clinical results. North American Linomide Investigators. Neurology 2000; 54: 1726–33.

Paolillo A, Coles AJ, Molyneux PD, Gawne-Cain M, MacManus D, Barker GJ, et al. Quantitative MRI in patients with secondary progressive MS treated with monoclonal antibody Campath 1H. Neurology 1999; 53: 751–7.

Pitt D, Werner P, Raine CS. Glutamate excitotoxicity in a model of multiple sclerosis. Nat Med 2000; 6: 67–70.

Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989; 8: 431–40.

Rieckmann P, Altenhofen B, Riegel A, Baudewig J, Felgenhauer K. Soluble adhesion molecules (sVCAM-1 and sICAM-1) in cerebrospinal fluid and serum correlate with MRI activity in multiple sclerosis. Ann Neurol 1997; 41: 326–33.

Rolan P, Atkinson AJ Jr, Lesko LJ. Use of biomarkers from drug discovery through clinical practice: report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development. Clin Pharmacol Ther 2003; 73: 284–91.

Sturzebecher S, Wandinger KP, Rosenwald A, Sathyamoorthy M, Tzou A, Mattar P, et al. Expression profiling identifies responder and non-responder phenotypes to interferon-beta in multiple sclerosis. Brain 2003; 126: 1419–29.

Trapp BD, Peterson J, Ransohoff RM, Rudick RM, Mork S, Bo L. Axonal transection in the lesions of multiple sclerosis. New Engl J Med 1998; 338: 278–85.

Wandinger KP, Sturzebecher CS, Bielekova B, Detore G, Rosenwald A, Staudt LM, et al. Complex immunomodulatory effects of interferon-beta in multiple sclerosis include the upregulation of T helper 1-associated marker genes. Ann Neurol 2001; 50: 349–57.

Wandinger KP, Lunemann JD, Wengert O, Bellmann-Strobl J, Aktas O, Weber A, et al. TNF-related apoptosis inducing ligand (TRAIL) as a potential response marker for interferon-beta treatment in multiple sclerosis. Lancet 2003; 361: 2036–43.

Whitney LW, Ludwin SK, McFarland HF, Biddison WE. Microarray analysis of gene expression in multiple sclerosis and EAE identifies 5-lipoxygenase as a component of inflammatory lesions. J Neuroimmunol 2001; 121: 40–8.

Wolinsky JS, Narayana PA, Noseworthy JH, Lublin FD, Whitaker JN, Linde A, et al. Linomide in relapsing and secondary progressive MS: part II: MRI results. MRI Analysis Center of the University of Texas–Houston, Health Science Center, and the North American Linomide Investigators. Neurology 2000; 54: 1734–41.