

Consensus definitions and application guidelines for control groups in cerebrospinal fluid biomarker studies in multiple sclerosis

Multiple Sclerosis Journal
19(13) 1802–1809
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1352458513488232
msj.sagepub.com


Charlotte Teunissen¹, Til Menge², Ayse Altintas³, José C Álvarez-Cermeño⁴, Antonio Bertolotto⁵, Frode S Berven^{6,7}, Lou Brundin⁸, Manuel Comabella⁹, Matilde Degn¹⁰, Florian Deisenhammer¹¹, Franz Fazekas¹², Diego Franciotta¹³, Jette L Frederiksen¹⁰, Daniela Galimberti¹⁴, Sharmilee Gnanapavan¹⁵, Harald Hegen¹¹, Bernhard Hemmer¹⁶, Rogier Hintzen¹⁷, Steve Hughes¹⁸, Ellen Iacobaeus¹⁹, Ann C Kroksveen²⁰, Jens Kuhle²¹, John Richert²², Hayrettin Tumani²³, Luisa M Villar²⁴, Jelena Drulovic²⁵, Irena Dujmovic²⁵, Michael Khalil¹² and Ales Bartos^{26,27}

Abstract

The choice of appropriate control group(s) is critical in cerebrospinal fluid (CSF) biomarker research in multiple sclerosis (MS). There is a lack of definitions and nomenclature of different control groups and a rationalized application of different control groups. We here propose consensus definitions and nomenclature for the following groups: healthy controls (HCs), spinal anesthesia subjects (SASs), inflammatory neurological disease controls (INDCs), peripheral inflammatory neurological disease controls (PINDCs), non-inflammatory neurological controls (NINDCs), symptomatic controls (SCs). Furthermore, we discuss the application of these control groups in specific study designs, such as for diagnostic biomarker studies, prognostic biomarker studies and therapeutic response studies. Application of these uniform definitions will lead to better comparability of biomarker studies and optimal use of available resources. This will lead to improved quality of CSF biomarker research in MS and related disorders.

Keywords

Biomarkers, cerebrospinal fluid, control groups, study design

Date received: 14 November 2012; accepted: 26 March 2013

¹Neurological Laboratory, Innogentecis and Roche, VU University Medical Center, Netherlands.

²Department of Neurology, Heinrich Heine-University, Germany.

³Neurology Department, Cerrahpasa Medical School, Istanbul University, Turkey.

⁴Department of Neurology, MS Unit, Hospital Ramón y Cajal, IRYCYS, REEM, Spain.

⁵Neurologia 2, Ospedale S. Luigi, Orbassano, Italy.

⁶Department of Biomedicine, Proteomics Unit (PROBE), University of Bergen, Norway.

⁷The Norwegian Multiple Sclerosis Competence Centre, Haukeland University Hospital, Norway.

⁸Neurology Clinic, Karolinska University Hospital, Sweden.

⁹Centre d'Esclerosi Múltiple de Catalunya, Hospital Universitari Vall d'Hebron, Spain.

¹⁰Department of Neurology, Glostrup University Hospital, University of Copenhagen, Denmark.

¹¹Innsbruck Medical University, Department of Neurology, Austria.

¹²Department of Neurology, Medical University of Graz, Austria.

¹³Laboratory of Neuroimmunology, IRCCS - C. Mondino National Neurological Institute, Italy.

¹⁴Neurology Unit, Dept. of Pathophysiology and Trasplantation, University of Milan, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Italy.

¹⁵Neuroscience Centre, Barts and The London School of Medicine and Dentistry, UK.

¹⁶Department of Neurology, Klinikum rechts der Isar, Technische Universität München/ Munich Cluster for Systems Neurology (SyNergy), Munich/German Competence Network Multiple Sclerosis (KKNMS), Munich, Germany.

¹⁷Department of Neurology, Erasmus MC, The Netherlands.

¹⁸Ixis Pharmaceuticals Inc., USA.

¹⁹Department of Clinical Neuroscience, Karolinska Institute, Sweden.

²⁰The KG Jebsen Centre for MS-research, University of Bergen, Norway.

²¹Department of Neurology and Clinical Immunology, University Hospital Basel, Switzerland.

²²Global Medical Affairs Department, Biogen Idec Inc., USA.

²³University Hospital of Ulm, Germany.

²⁴Department of Immunology, MS Unit, Hospital Ramón y Cajal, IRYCYS, REEM, Spain

²⁵Clinic of Neurology, University of Belgrade, Serbia.

²⁶Prague Psychiatric Center, Czech Republic.

²⁷Charles University in Prague, Third Faculty of Medicine, Czech Republic.

Corresponding author:

Charlotte Teunissen, Neurological Laboratory, Department of Clinical Chemistry, VU University Medical Center Amsterdam, PO Box 7057, Amsterdam, 1007 MB, The Netherlands.
Email: c.teunissen@vumc.nl

Introduction

The choice of right control group(s) is an important issue for every clinical study and therefore also critical in cerebrospinal fluid (CSF) biomarker research. Since study outcomes are almost inevitably influenced by the characteristics of the control groups, control groups should be well-defined and be of sufficient size to enable adequate comparison with the investigated disease group. However, control groups are often not comparable between CSF biomarker studies, having variable sizes and containing different syndromes and disease entities. CSF from healthy individuals, probably the ideal control group, is difficult to obtain for various reasons. Consequently, CSF samples of completely healthy persons are rare and disease controls are usually applied in CSF biomarker studies. Ultimately, disease controls are also clinically more relevant than healthy controls (HCs). However, there is a lack of consistent and rationalized design of control groups for specific biomarker research studies.

The fact that study outcomes can be influenced by the choice of control groups is illustrated in recent literature. A search of the literature on 'cerebrospinal fluid', 'multiple sclerosis (MS)' and 'control(s)' in 2010 and 2011 has shown that out of 66 studies, only 10 studies included more than one control group. The studies that included more than one control group showed that biomarker levels can be different among the several control groups. For example, neural cell adhesion molecule (NCAM) levels were lower in meningitis patients compared to HCs and MS patients, while levels in viral encephalitis patients were higher than in every other group.¹ Another example is that CSF CXCL13 levels are higher in MS patients compared to several control groups, including inflammatory controls, though not higher than in patients with a viral and bacterial infection, as these were extremely high.² These examples not only clearly illustrate that biomarker concentrations can be extreme in a single type of disease entity, but also show that merging of these groups into one inflammatory control-group would have led to different conclusions.

As already introduced, another problem is that different definitions and names are used for control groups in the biomarker literature. Most of the studies refer to the use of 'non-inflammatory' or 'inflammatory' disease controls, but clear guidelines are not yet available as to how these are characterized and selected.

To obtain comparable results between biomarker studies, clear consensus definitions are a prerequisite. In the current paper, the BioMS-eu network has developed definitions for control group stratification through several extensive discussions within the network. As a start, we established the current 'state of the art' of definitions used for control groups and the disease entities or symptoms included. We set out an inventory to generate detailed information on these definitions and the characteristics of healthy, inflammatory and non-inflammatory controls

among 12 laboratories of the BioMS-eu network. We discovered that classifications are extremely diverse in terms of condition labeling (from symptoms through syndromes to diagnoses), and regarding categorization. For example, for the control group named 'healthy controls', two out of 12 BioMS-eu centers included patients undergoing spinal anesthesia for leg operations in this category, while two other centers included strictly healthy volunteers within this category. Yet other centers consider patients with subjective complaints, often headache or vertigo, for whom no clear explanation could be found after a clinically detailed work-up, as 'healthy controls'.

The consortium next elaborated, discussed and fine-tuned the definitions which we will present below. The definitions are based on both clinical and laboratory criteria. The presence of these definitions will provide the opportunity to intensify national and international collaborations to obtain the appropriate material serving as controls. Furthermore it will lead to more clear answers in biomarker studies. The current paper provides not only the consensus definitions and uniform nomenclature, but also typical example-disorders for each suggested subgroup. Next, we provide guidelines for uniform and rationalized employment of one or several of these controls in the design of CSF biomarker studies.

The recommendations focus on CSF biomarker studies, as the difficulty of getting HCs is most clear for this body fluid and therefore appropriate inclusion and interpretation of alternative controls is important. However, the definitions must be the same for biomarker studies using other body fluids, such as blood, saliva, and urine. We started our discussions using MS as a model disease. Nevertheless, we realized that these recommendations will be applicable in a variety of other neurological diseases, especially neuroinflammatory diseases, and will be a good basis for development of recommendations for other disorders related to MS. Consequently, MS is included in the tables with typical examples.

Terminology and definitions of control groups in CSF biomarker studies

Fundamental requirements

Basic CSF examination must be performed for every individual, including immunoglobulin measurements and evaluations in the quotient diagram, albumin quotient (QAlb), i.e. the albumin CSF/serum ratio, and CSF/serum glucose ratio.³ We recommend use of QAlb rather than total CSF protein in order to characterize protein content in CSF since QAlb analysis is more reliable (better coefficient of variation than total protein).³ In addition, QAlb serves as reference measure for evaluation of systemic vs intrathecal origin of candidate protein biomarkers in CSF, as it has no intrathecal origin.⁴

Demographic and clinical information, and the use of medication for every control must be collected. This allows us to exclude patient- or study-confounding effects on specific biomarkers under study, e.g. effects of anti-inflammatory drugs on an inflammatory biomarker. If brain magnetic resonance imaging (MRI) data could be available for controls as well, confidence in the definition would increase. However, an MRI is expensive and often confronted with ethical constraints and it is not realistic for this to be conducted for every possible control. Furthermore, it is imperative that control samples are collected and stored under similar conditions as the patient samples, i.e. in agreement with published guidelines for CSF collection and biobanking.⁵ This remark may be redundant, but applies if control samples are remnant CSF from diagnostic procedures, or taken during an operation (SAS group) and thereby collected under different conditions than samples from the target disease population.

1. Healthy controls (HCs)

Clinical criteria: The HC are volunteers without specific medical complaints (also not recurrent and disabling headache, back pain etc.) invited for research purposes only. The HC individual is actively approached and invited by the researcher, and lumbar puncture is undertaken to obtain CSF for research in the first place, not to exclude any unknown disease. Patients with either focal or diffuse current or historical neurologic deficits, examined by a neurologist, should be excluded. In the ideal situation, these persons are clinically followed for 1–2 years to ensure absence of pre-symptomatic neurological disease. *Laboratory criteria:* Basic CSF analysis should be normal including the laboratory criteria absence of oligoclonal IgG bands (OCBs) and a normal QAlb to exclude a preclinical neurological disease.

While the inclusion of this control group in studies and number of individuals in this control group is often restricted for obvious reasons, their inclusion as controls is important to define whether a biomarker is related to central nervous system (CNS) pathology anyway. In fact, pathogenic processes within the CNS of other neurological diseases can follow similar pathobiological mechanisms, in spite of their underlying different causes, and thus a biomarker can be altered in any CNS disease and consequently all possible disease control groups.

On the other hand this HC group stays problematic for several reasons. Unlike cohorts of neurological patients they do not usually have other auxiliary investigations (neuroimaging, electrophysiological investigations etc.). Other obstacles include ethical considerations, the relative invasiveness of a lumbar puncture, expensive insurance or cost of additional instrumental investigations. Sometimes the HCs are relatives, but this is a less ideal control group when

genetic interference is suspected and thus spouses may be better controls where, on the other hand, environmental factors may introduce a bias. To form an optimal HC group, first or second-degree relatives to patients with the target disease or, in the case of MS, other autoimmune diseases should be excluded.

2. Spinal anesthesia subjects (SASs)

Clinical criteria: Normal neurological examination, preferably performed by a neurologist, and no previous history of any neurological deficit are mandatory. With this information it can be assumed that the contribution of pre- or oligosymptomatic individuals remains negligible. Usually less information on the history and neurological examination is available for SASs than in neurological controls. Data from auxiliary methods are similarly difficult to obtain as in the HCs.

Laboratory criteria: Normal basic CSF analysis is a minimal requirement, including absence of OCBs and normal QAlb to exclude preclinical neurological disease.

The lumbar puncture in this situation is usually not performed by a neurologist, but by an anesthesiologist. Therefore, special attention should be given to the fact that collection procedures are in agreement with the consensus guidelines for collection and sampling of CSF similar to the other patient groups (sufficient sample volume, polypropylene tubes, blood collected in parallel to CSF etc.).⁵

3. Inflammatory neurological disease controls (INDCs)

Clinical criteria: This group includes all neurological diseases that are inflammatory in nature. The determination of the exact causative agent is preferable, i.e. etiopathogenetic diagnosis, although it is not necessary for the biomarker research. The most common examples are listed in Table 1.

Laboratory criteria: Abnormal CSF findings, such as pleocytosis, elevated QAlb, or both, should be present. However, this is not obligatory if other findings clearly support the diagnosis of an inflammatory disorder, e.g. neurolyupus.

4. Peripheral inflammatory neurological disease controls (PINDCs)

Clinical criteria: This group includes autoimmune diseases of the peripheral nervous system that are inflammatory, but differ from INDCs. The most common examples are listed in Table 2.

Laboratory criteria: Elevated QAlb is expected, but not a necessary condition.

5. Non-inflammatory neurological disease controls (NINDCs)

Clinical criteria: This group is defined mainly by exclusion of INDCs. It involves patients with defined neurological diseases that have no clear inflammatory aspect. The most common examples are listed in Table 3.

Laboratory criteria: The CSF cell count must be normal. QAlb can be normal or elevated.

6. Symptomatic controls (SCs) (the name is given to distinguish this group from HCs)

Clinical criteria: These are patients with neurological symptoms, but have no objective clinical or paraclinical findings to define a specific neurological disease at the time of sampling. They are without putative structural damage. These patients may have a somatoform (psychogenic) disorder that cannot be explained by any somatic neurological pathology. Thus, this category is also defined by exclusion, i.e. not fulfilling the NINDC and NINDC and listed in Table 1–3. The most common examples of SCs are listed in Table 4. Importantly, these individuals cannot be categorized as HCs as is often done in publications.

Laboratory criteria: CSF is required to be negative for OCBs, and to have a normal QAlb, and a normal cell count.^{3,6}

Table 1. Study classification and examples of inflammatory neurological disease controls (INDCs).

Main biomarker purpose: differential diagnosis:

Infectious diseases: neuroborreliosis, neuroleues

Autoimmune neurological diseases: MS, neurolupus, Sjögren's disease with CNS involvement, antiphospholipid antibody syndrome, other types of systemic vasculitides, isolated CNS vasculitis, neurosarcoidosis, Behcet's disease, acute/ multiphasic disseminated encephalomyelitis, neuromyelitis optica (NMO).

Paraneoplastic CNS neurological syndromes

Rare diseases: Hashimoto encephalopathy, Fisher syndrome

Main biomarker purpose: other research issues (e.g. unraveling pathological specificity):

Aseptic encephalitis or myelitis

Aseptic meningitis

Bacterial meningitis

CNS: central nervous system; MS: multiple sclerosis.

Table 2. Study classification and examples of controls with inflammatory neurological disease of the peripheral nervous system (PINDCs).

Peripheral inflammatory neurological disease controls

Main biomarker purpose: other research issues (e.g. unraveling pathological specificity):

Inflammatory demyelinating neuropathies: acute inflammatory demyelinating polyneuropathy/Guillain-Barré syndrome, chronic inflammatory demyelinating polyneuropathy, multifocal motor neuropathy

Inflammatory neuritis including facial palsy, plexopathies, radiculitis, neuropathies associated with inflammatory systemic diseases

Table 3. Study classification and examples of non-inflammatory neurological disease controls (NINDCs).

Main biomarker purpose: differential diagnosis:

CNS expansions: tumor of hypophysis, cavernoma, CNS lymphoma, glioblastoma, astrocytoma, syringomyelia

Vascular diseases: transient ischemic attacks, minor strokes, strokes

Neurodegenerative diseases: motor neuron disease – amyotrophic lateral sclerosis, primary lateral sclerosis, spinal muscle atrophy, hereditary spastic paraparesis, cerebellar neurodegeneration

Hereditary/metabolic encephalopathies: cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), metabolic leukodystrophies, mitochondrial diseases

Main biomarker purpose: other research issues (e.g. unraveling pathological specificity):

Vascular diseases: transient ischemic attacks, minor strokes, strokes

Unconsciousness: syncope, epilepsy of non-inflammatory cause

Non-inflammatory PNS involvement: polyneuropathy, plexopathy, radiculopathy (e.g. disc herniation), mononeuropathy (e.g. n. III, peroneal palsy)

CSF flow abnormalities: benign intracranial hypertension, normotensive hydrocephalus, congenital hydrocephalus)

Dementia disorders and Parkinson syndromes:

Alzheimer's dementia, vascular dementia, frontotemporal dementia, Lewy Body dementia, Parkinson's disease. (However, the typical higher age of these patients makes comparison with MS patients usually difficult.)

CNS: central nervous system; CSF: cerebrospinal fluid; MS: multiple sclerosis. PNS: peripheral nervous system; n. III: oculomotor nerve.

Table 4. Study classification and examples of symptomatic controls (SCs).

Main biomarker purpose: differential diagnosis:

Sensory disturbances: paresthesias, hypesthesias in any part of the human body (brain and/or spinal cord MRI and CSF oligoclonal bands need to be performed for judgement of 'normal')

Dizziness: only if neurological reasons for dizziness are excluded and brain MRI and CSF including oligoclonal bands are normal, i.e. to exclude e.g. vascular cause

Polymorphic complaints: different symptoms simultaneously – e.g. paresthesias, vertigo, headache, visual disturbances, fatigue etc.

Main biomarker purpose: other research issues (e.g. unraveling pathological specificity):

Headache: acute, chronic, tension headache, other headache syndromes, migraine, when an underlying organic cause is excluded

Idiopathic Bell's facial palsy: normal basic CSF findings

Vertebrogenic syndromes, disc herniation, spinal stenosis, spondylosis

CSF: cerebrospinal fluid; MRI: magnetic resonance imaging; MS: multiple sclerosis.

Further subdivisions of control groups

In addition to dividing different diseases into main categories like INDC, there is also an option to further divide these main categories into subcategories. This subdivision could be particularly useful for the INDC group since it contains very heterogeneous diseases. By (current) definition the INDC group may consist of patients with bacterial, viral and autoimmune encephalitis, meningitis, or polyradiculoneuritis. This group could be subdivided into several categories: (a) encephalitis, (b) diseases with meningeal involvement and (c) peripheral neuroinflammatory diseases. The example of NCAM levels, where high levels were seen in viral encephalitis but not in meningitis, shows that it could be useful to differentiate between bacterial and viral meningitis patients.¹ A further subdivision will certainly lead to more information on the specificity of the biomarker. However, to include a sufficient number of patients in each category, limited availability of patients within these subgroups will hamper the ability to perform these studies, even in collaborative effects, and as we provide guidelines for novel biomarkers studies, it is difficult to make a generalized recommendation for further subdivision. We, therefore, do not provide more subdivisions in our recommendations. If the different diagnoses are well balanced and the sample size is sufficiently large, heterogeneity will not matter and the influence of only a few outliers will be reduced. The research question asked should determine if a further subdivision is preferred or not. If the experimental results show a wide range of outcomes, then one can zoom into the details of the diseases to find a common denominator that can be used to include or exclude specific disease entities, e.g. 'autoimmune or peripheral'. If that is not possible, larger and more heterogeneous control groups will still be needed in subsequent studies. If a priori knowledge of a biomarker is present, then specific disease groups can be selected.

Which control group should be included?

The choice of a specific control group depends on the research question. It is therefore not possible to provide general guidelines. It is also not imperative to include all control groups into a study, as the kind of control groups depends on the research question, e.g. this could concern the pathobiological mechanism, issues for differential diagnosis, or may even address etiological factors for several diseases, whereby some controls even become the target disease. In practice, the INDC, NINDC and SC groups will most frequently be used. For several specific research questions that are a frequent focus of biomarker studies, we here provide considerations and recommendations for the optimal application of control groups.

Search for a diagnostic marker. Diagnostic markers for e.g. MS could have value if they were able to (a) differentiate MS patients, preferably at an early stage, from all other control groups, (b) differentiate early MS from the

disease entities that are similar in appearance and currently most difficult to differentiate from MS, such as patients with Sjogren's syndrome, neuromyelitis optica (NMO), acute disseminated encephalomyelitis, or (c) differentiate other autoimmune neurological diseases. Examples are shown under each subheading of 'differential diagnosis' in Table 1–4. If a marker is novel, then as many different diagnoses as possible should be included, starting with the diseases most commonly seen in clinical practice. Moreover, there may also be questions where one control group would be sufficient, e.g. looking for a biomarker specific for NMO vs MS. It is impossible to define in advance what the levels of a new biomarker in specific control groups will be. For example, the high levels of NCAM, which was assumed to be a neuronal specific marker, in viral encephalitis but not in meningitis could not have been predicted.¹

Search for a prognostic marker. The optimal study design for prognostic markers includes a comparison between patients of different disease stages (e.g. early MS vs severely progressed MS), or the longitudinal evaluation of patients over different disease phases. Here, a biomarker does not need to be different in the control group compared to the patient group under study, but evaluation of the prognostic value in a control group could be used to reveal the specificity of the biomarker for the pathology of the target disease. However, a biomarker can have prognostic value for a particular disease even though it is not specific at all, such as for neurofilament proteins being predictive for conversion from clinically isolated syndrome (CIS) to MS.^{7–9} With respect to the nature of the control group, this can be more diverse than in studies aimed at defining diagnostic markers.

Search for a therapeutic marker. The placebo-treated group is the gold standard in such studies. Alternatively, the patients may be their own control in follow-up studies, requiring repeated lumbar puncture. This may be the best situation in multifactorial diseases such as MS, since a number of possible interferences (genetic, environmental, phenotypic expression) are completely avoided. In this instance, the sample sizes should be calculated and planned very conservatively anticipating a higher than usual dropout rate after the first lumbar puncture. On the other hand, a statistical analysis of dependent variables can be applied (such as the paired *t* test), that provide an increase in statistical power.

Search for a biomarker reflecting pathology. Here, the same applies as for a prognostic marker. Controls will reveal the specificity of the biomarker and point to the mechanism of the pathology of the target disease. And, just as for the diagnostic marker, the more different diseases and control groups that are included the better as this will give more accurate information about the specific involvement of the protein in the pathology of the target disease.

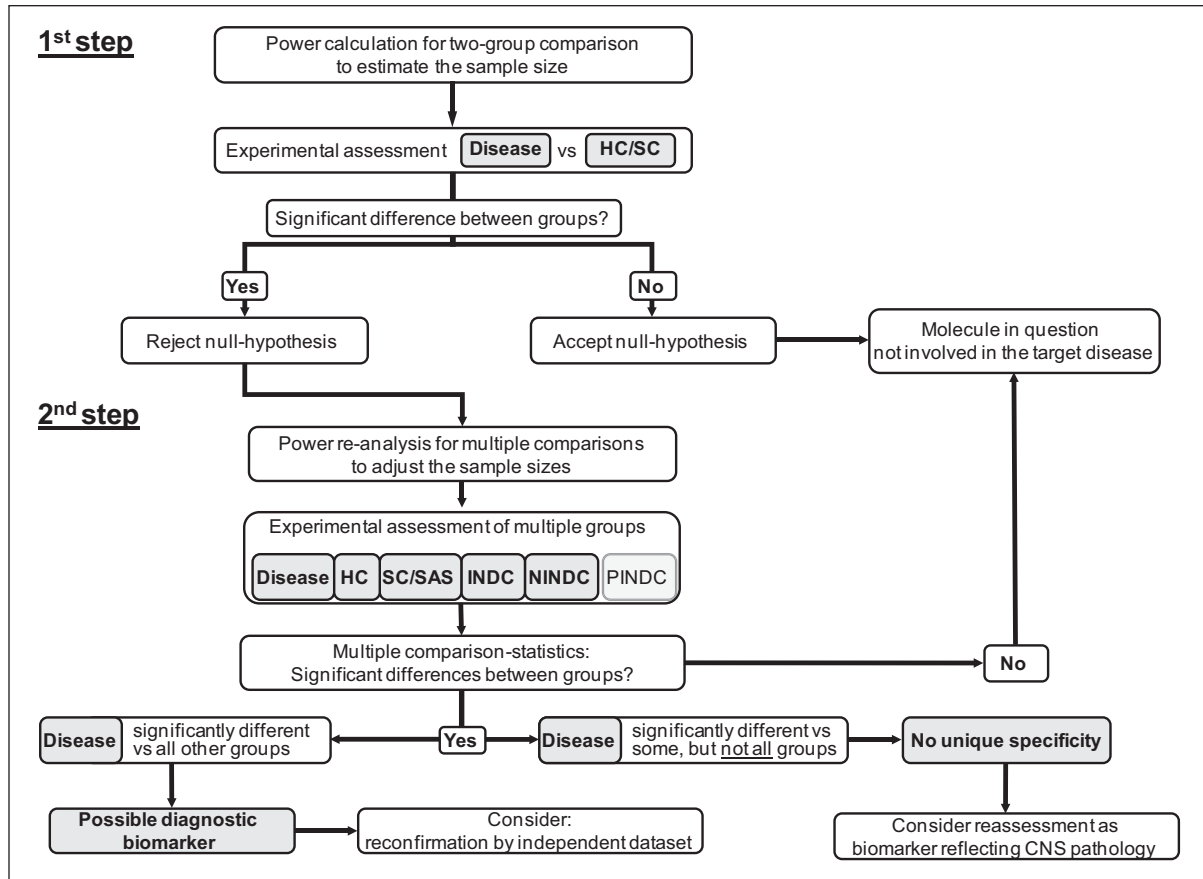


Figure 1. Two-step approach for the assessment of diagnostic use of candidate cerebrospinal fluid (CSF) biomarkers.

Flowchart visualizing the two consecutive experimental steps necessary to reliably assess the potency of a CSF target molecule to serve as a diagnostic biomarker for a target disease, indicated by 'disease'. First, CSF samples from patients with the target disease are analyzed against a single control group consisting of healthy controls (HCs), and/or symptomatic controls (SCs). Thus statistical analysis can be done by unpaired two-tailed t test or a Mann-Whitney U test avoiding loss of statistical power due to multiple comparisons. For the a priori sample size estimation, the most conservative theoretical discriminatory cut-off for the CSF molecule should be assumed, if this is not known. If a significant difference is not detected, specificity for the CSF molecule discerning the target disease from controls, may not be assumed and the experiment/project should be terminated. Otherwise in the second step, additional (disease) control groups should be tested, namely inflammatory neurological disease controls (INDCs), non-inflammatory neurological controls (NINDCs), and possibly peripheral inflammatory neurological disease controls (PINDCs). Spinal anesthesia subjects (SASs) may serve as another non-diseased control group. Due to the multiple comparisons approach in this second step, the sample sizes should be re-calculated using the observed discriminatory power of the first experiment. Thus, additional CSF samples of the patients and/or HCs/SCs may have to be tested in order to adjust the sample sizes. Only if the CSF samples of the target disease are significantly different from all other groups tested and sensitivity and specificity are sufficiently high, can a possible diagnostic potency of the target molecule be assumed and should be validated in a second independent dataset. CNS: central nervous system.

Number of patients per control group

With respect to the numbers of patients, general and basic rules of biostatistics have to be applied. If possible, a formal power analysis should be performed to determine the minimal sample size needed to correctly reject the null hypothesis that has been defined (i.e. no difference for the biomarker in question between the targeted CNS disease and the controls). The sample sizes should be sufficiently large: however, there is no need to over-exceed the sample sizes as no additional information is gained, rather CSF material and procedural material (antigens, reagents etc) is wasted. The more control groups that are included in the experiment, the smaller is the statistical power of the analysis due to the post hoc multiple comparisons of any statistical

test. Thus, with increasing numbers of control groups the sample sizes have to be increased as well. Care should be taken to match for gender and age in between groups, since e.g. QAlb is age dependent and so are many biomarkers.^{10,11} Pooling of samples should be avoided, as dilution may be introduced as a confounding factor, e.g. if one CSF sample which is highly positive for a given biomarker is pooled with four negative CSF samples from the control group, the concentration of the biomarker in the pooled analysis is decreased to 20% and may become negative.

If a CSF molecule is assessed as a diagnostic biomarker specific for the targeted CNS diseases, the experiment could be planned with a two-step approach in order to save material and workload (see Figure 1). First, CSF samples of patients with the target disease are

tested against non-diseased control groups, i.e. either HCs or SCs. If there are no differences detected between these two groups, the hypothesis has to be rejected and no further experiments are needed. Otherwise, in a second experiment the CSF of the targeted disease should be tested against a number of control groups to specifically validate the results. Thus, while homogeneity is desirable when defining a control group (e.g. inflammatory vs non-inflammatory), within any control group the spectrum of disorders should be somewhat heterogeneous and clusters of only a few entities avoided, unless the project is hypothesis-driven. In this case, a recalculation of the power analysis is advisable obeying the magnitude of the effect observed in the first-step of the experiment. This two-step approach may not be suitable for every study, but should be considered during project planning as it poses the advantage of increasing the statistical power in the first step (or conversely providing the option to decrease the sample sizes) and of specifically validating the results in the second step.

Final remarks

By defining groups, we realize that there will be patients that will not fall within one of the categories, and thereby will generally be excluded from biomarker studies. These are patients with multiple diagnoses and thus do not fit into one single control group as defined above. An example is cerebral hemorrhage, as blood contamination will confound the results. One may argue that CSF samples from patients with infectious hazards (hepatitis B or C, purulent meningococcal meningitis) are also unsuitable due to safety reasons.

There is biological heterogeneity within nosologically homogeneous groups (especially in those that need the application of polythetic criteria for diagnosis). The approach of increasing homogeneity of control groups by application of clinical and biological criteria as we propose here is challenging, but should not be discarded a priori.

It is worthwhile emphasizing that the ideal control group may not exist. Even HCs may not be completely healthy after all, or may possess known or unknown factors that confound the research outcome. Furthermore, even individuals considered as healthy may suffer from disorders which remain undetected if they are not questioned and examined in a formal manner. It is likely that certain conditions such as sports activities (jogger or marathon runner), eating habits (vegetarian or frequent fast-food intake), and disturbed sleep rhythm (shift worker or mild depression) in otherwise healthy persons may influence levels of CSF biomarkers. This may be overcome by applying a sufficiently large control cohort.

It is expected that the guidelines given in this paper will be instructive for harmonizing CSF biomarker studies. Replication of outcomes obtained in single studies will become easier and their interpretation more

straightforward. Ideally organizations such as the BioMS-eu network will, at some stage, even be able to provide a set of well-defined, homogeneous controls to validate results and make them globally comparable. Providing adequate infrastructure and information networks will allow multi-centre studies and inclusion of sufficient numbers of controls in each group. The use of these definitions is another important step forward to facilitate multi-center collaboration, which is essential in CSF biomarker studies in MS and related disorders and will ultimately lead to well- and independently-validated biomarkers for clinical applications.

Author contributions

Charlotte Teunissen has organized and prepared the consensus discussions, made the design of the paper, and had the lead in writing and editing the paper. Til Menge has contributed to the preparation of the manuscript and its discussion. Ayse Altintas participated in the consensus discussions and in editing the paper. José C Álvarez-Cermeño participated in the consensus discussions and in editing the paper. Antonio Bertolotto contributed to the consensus discussions and in editing the paper. Frode S Berven participated in the consensus discussions and in editing the paper. Lou Brundin participated in the consensus discussions and in editing the paper. Manuel Comabella participated in the consensus discussions and in editing the paper. Matilde D Vinther participated in the consensus discussions and in editing the paper. Florian Deisenhammer has provided critical comments and suggestions and reviewed and approved the final version of the manuscript. Franz Fazekas has provided critical comments and suggestions and reviewed and approved the final version of the manuscript. Diego Franciotta participated in the consensus discussions and in editing the paper. Jette Frederikson participated in the consensus discussions and in editing the paper. Daniela Galimberti participated in the consensus discussions and in editing the paper. Sharmilee Gnanapavan has contributed in discussions and provided critical comments and suggestions for the manuscript. Harald Hegen participated in the consensus discussions and in editing the paper. Bernhard Hemmer has contributed in discussions, writing and editing the paper. Rogier Hintzen participated in the consensus discussions and in editing the paper. Steve Hughes participated in the consensus discussions and in editing the paper. Ellen Iacobaeus has critically reviewed the manuscript contents and tables and provided suggestions for improvement. Ann C Kroksveen has critically reviewed the manuscript contents and tables and provided suggestions for improvement. Jens Kuhle participated in the consensus discussions and in writing and editing the paper. John Richert participated in the consensus discussions and in editing the paper. Hayrettin Tumani participated in the consensus discussions and in editing the paper. Luisa M Villar participated in the consensus discussions and in editing the paper. Jelena Drulovic participated in the consensus discussions and in editing the paper. Irena Dujmovic participated in the consensus discussions and in editing the paper. Michael Khalil participated in the consensus discussions and in editing the paper. Ales Bartos designed and wrote initial versions of the manuscript, formulated first definitions of control groups, created the very first tables and repeatedly changed them in a

well-balanced way according to different opinions of others and repeatedly revised and edited the text of the manuscript.

Conflict of interest

Charlotte Teunissen is member of the International Advisory Board of Innogenetics and Roche. José C Álvarez-Cermeño received speaker honoraria from Biogen Idec, Merck Serono, TEVA, Bayer Schering, Sanofi-Aventis and Novartis; have been a member of scientific boards supported and received research support from Biogen Idec, Bayer Schering, Merck Serono, Sanofi-Aventis and Novartis. Lou Brundin has received speaker honorariums from Biogen Idec, Schering, Novartis and SanofiAventis. Dr. Deisenhammer has participated in meetings sponsored by or received honoraria for acting as an advisor for Bayer, Biogen Idec, Genzyme-Sanofi, Merck Serono, Novartis and Teva-Ratiopharm. His institution has received financial support for participation in randomized controlled trials of INFb-1b (Betaferon, Bayer Schering Pharma), INFb-1a (Avonex, Biogen Idec; Rebif, Merck Serono), glatiramer acetate (Copaxone, Teva Pharmaceuticals) and Natalizumab (Tysabri, Biogen Idec) in multiple sclerosis. Jette L Frederiksen has served on scientific advisory boards for and received funding for travel related to these activities and honoraria from Biogen Idec, Sanofi-Aventis, Genzyme, Teva, Novartis and Almirall. Jette Frederiksen has received speaker honoraria from Biogen Idec, Merck Serono and Teva. She has served as advisor on preclinical development for Takeda. Harald Hegen received honoraria (lectures) from pharmaceutical companies marketing treatment for multiple sclerosis (Bayer Schering, Biogen Idec, Merck Serono).

H. Tumani serves on a scientific advisory board, is a consultant for, and/or has received funding from the following companies: Bayer Healthcare, Biogen Idec, Genzyme Virotec, Merck Serono, Novartis, Roche and Teva. Luisa M Villar has received speaker honoraria from Biogen Idec, Merck Serono, TEVA, Bayer Schering, and Novartis. Michael Khalil has received research support from The Austrian Science Fund (FWF) [J2992-B09].

Antonio Bertolotto has been on steering committees in clinical trials sponsored by Biogen Idec, Roche; has received speaker honoraria from Biogen Idec, Merck Serono, TEVA, Bayer Schering, Sanofi-Aventis and Novartis; has been a member of scientific boards supported by Allergan, Almirall; has received research support from Biogen Idec, Biogen-Dompè, Bayer Schering, Merck Serono, Sanofi-Aventis and from the Italian Multiple Sclerosis Society and from Associazione Ricerca Biomedica Onlus.

John Richert is employed by Biogen Idec Inc. The remaining authors have no conflicts of interest to report.

Funding

The work of Ales Bartos on this manuscript was partly supported by grant IGA No. NT 13183 from the Czech Ministry of Health

and by grant Research Project Charles University in Prague, PRVOUK 34/LF3. Ann C Kroksveen and Frode Berven were supported by the Kristian Gerhard Jebsen Foundation. Jelena Drulovic has received research grant support from the Ministry of Education and Science, Republic of Serbia (project no. 175031), and has received speaker honoraria and travel support from Merck Serono SA, Bayer, Richter Gedeon and Novartis. Irena Dujmovic has received research grant support from the Ministry of Education and Science, Republic of Serbia (project no. 175031), and has received travel support from Merck Serono SA, Bayer and Boehringer Ingelheim. Rogier Q Hintzen received support from the Dutch Foundation of MS Research. The funding sources had no role in study design, definitions or writing of this paper.

References

- Gnanapavan S, Grant D, Illes-Toth E, et al. Neural cell adhesion molecule—description of a CSF ELISA method and evidence of reduced levels in selected neurological disorders. *J Neuroimmunol* 2010; 225: 118–122.
- Khademi M, Kockum I, Andersson ML, et al. Cerebrospinal fluid CXCL13 in multiple sclerosis: A suggestive prognostic marker for the disease course. *Mult Scler* 2011; 17: 335–343.
- Deisenhammer F, Bartos A, Egg R, et al. Guidelines on routine cerebrospinal fluid analysis. Report from an EFNS task force. *Eur J Neurol* 2006; 13: 913–922.
- Brettschneider J, Hartmann N, Lehmsiek V, et al. Cerebrospinal fluid markers of idiopathic intracranial hypertension: Is the renin-angiotensinogen system involved? *Cephalalgia* 2011; 31: 116–121.
- Teunissen CE, Petzold A, Bennett JL, et al. A consensus protocol for the standardization of cerebrospinal fluid collection and biobanking. *Neurology* 2009; 73: 1914–1922.
- Andersson M, Alvarez-Cermeño J, Bernardi G, et al. Cerebrospinal fluid in the diagnosis of multiple sclerosis: A consensus report. *J Neurol Neurosurg Psychiatry* 1994; 57: 897–902.
- Lim ET, Sellebjerg F, Jensen CV, et al. Acute axonal damage predicts clinical outcome in patients with multiple sclerosis. *Mult Scler* 2005; 11: 532–536.
- Norgren N, Rosengren L and Stigbrand T. Elevated neurofilament levels in neurological diseases. *Brain Res* 2003; 987: 25–31.
- Teunissen CE, Iacobaeus E, Khademi M, et al. Combination of CSF N-acetylaspartate and neurofilaments in multiple sclerosis. *Neurology* 2009; 72: 1322–1329.
- Deisenhammer F, Egg R, Giovannoni G, et al. EFNS guidelines on disease-specific CSF investigations. *Eur J Neurol* 2009; 16: 760–770.
- Kuhle J, Leppert D, Petzold A, et al. Neurofilament heavy chain in CSF correlates with relapses and disability in multiple sclerosis. *Neurology* 2011; 76: 1206–1213.