Neuronal and glial CSF biomarkers in multiple sclerosis: a systematic review and meta-analysis

Sara Momtazmanesh, Parnian Shobeiri, Amene Saghazadeh, Charlotte E. Teunissen, Joachim Burman, Levente Szalardy, Peter Klivenyi, Ales Bartos, Adelaide Fernandes and Nima Rezaei*

Abstract: Multiple sclerosis (MS) is a neurodegenerative disease associated with inflammatory demyelination and astroglial activation, with neuronal and axonal damage as the leading factors of disability. We aimed to perform a meta-analysis to determine changes in CSF levels of neuronal and glial biomarkers, including neurofilament light chain (NFL), total tau (t-tau), chitinase-3-like protein 1 (CHI3L1), glial fibrillary acidic protein (GFAP), and S100B in various groups of MS (MS versus controls, clinically isolated syndrome (CIS) versus controls, CIS versus MS, relapsing-remitting MS (RRMS) versus progressive MS (PMS), and MS in relapse versus remission. According to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, we included 64 articles in the meta-analysis, including 4071 subjects. For investigation of sources of heterogeneity, subgroup analysis, meta-regression, and sensitivity analysis were conducted. Meta-analyses were performed for comparisons including at least three individual datasets. NFL, GFAP, t-tau, CHI3L1, and S100B were higher in MS and NFL, t-tau, and CHI3L1 were also elevated in CIS patients than controls. CHI3L1 was the only marker with higher levels in MS than CIS. GFAP levels were higher in PMS versus RRMS, and NFL, t-tau, and CHI3L1 did not differ between different subtypes. Only levels of NFL were higher in patients in relapse than remission. Meta-regression showed influence of sex and disease severity on NFL and t-tau levels, respectively and disease duration on both. Added to the role of these biomarkers in determining prognosis and treatment response, to conclude, they may serve in diagnosis of MS and distinguishing different subtypes.

Keywords: chitinase-3-like protein 1; diagnosis; glial fibrillary acidic protein; neurofilament protein light; tau protein.

Introduction

Multiple sclerosis (MS) is an autoimmune, neurodegenerative disease associated with inflammatory demyelination and astroglial activation, which affects more than two million people worldwide (Filippi et al. 2018; Reich et al. 2018). MS can have heterogeneous disease courses and pathological changes. Therefore, there is a substantial need for tools to discriminate different subtypes and to define the stage of the disease activity.

https://doi.org/10.1515/revneuro-2020-0145
Received December 8, 2020; accepted January 15, 2021; published online February 17, 2021

*N Corresponding author: Nima Rezaei, Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children’s Medical Center, Tehran University of Medical Sciences, Tehran, Iran; Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; and Network of Immunity in Infection, Malignancy and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Tehran, Iran, E-mail: rezaei_nima@tums.ac.ir

Sara Momtazmanesh, School of Medicine, Tehran University of Medical Sciences (TUMS), Children’s Medical Center Hospital, Dr. Qarib St., Keshavarz Blvd, Tehran 14194, Iran; Systematic Review and Meta-Analysis Expert Group (SRMEG), Universal Scientific Education and Research Network (USERN), Tehran, Iran; and Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children’s Medical Center, Tehran University of Medical Sciences, Tehran, Iran. https://orcid.org/0000-0002-3946-1854

Parnian Shobeiri, School of Medicine, Tehran University of Medical Sciences (TUMS), Children’s Medical Center Hospital, Dr. Qarib St., Keshavarz Blvd, Tehran 14194, Iran; and Systematic Review and Meta-Analysis Expert Group (SRMEG), Universal Scientific Education and Research Network (USERN), Tehran, Iran.

Amene Saghazadeh, Systematic Review and Meta-Analysis Expert Group (SRMEG), Universal Scientific Education and Research Network (USERN), Tehran, Iran; and Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children’s Medical Center Hospital, Tehran University of Medical Sciences, Tehran, Iran.

Charlotte E. Teunissen, Neurochemistry Laboratory, Department of Clinical Chemistry, Location VUMC, PK 2 BR 141, Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands

Joachim Burman, Department of Neuroscience, Uppsala University Hospital, 75185 Uppsala, Sweden

Levente Szalardy and Peter Klivenyi, Department of Neurology, Faculty of Medicine, Albert Szent-Györgyi Clinical Center, University of Szeged, Semmelweis u. 6, 6725 Szeged, Hungary

Ales Bartos, Department of Neurology, Third Faculty of Medicine, Charles University, Ruska 87, 100 00 Prague 10, Czech Republic

Adelaide Fernandes, Department of Pharmacological Sciences and Medicines, Faculty of Pharmacy, Universidade de Lisboa, Avenida Professor Gama Pinto, 1649-003 Lisbon, Portugal
Brain and spinal cord magnetic resonance imaging (MRI) is the most common and critical tool in the diagnosis, determining prognosis, and monitoring treatment response in MS (Wattjes et al. 2015). However, the standard MRI usually does not provide a complete understanding of the underlying histopathological changes and may not have a strong relationship with the patient’s clinical symptoms (Hemond and Bakshi 2018; Zivadinov et al. 2008). Therefore, there is a substantial need for additional markers assisting in diagnosis, determining disease course, and prognosis of MS.

Compared to biomarkers in the serum, cerebrospinal fluid (CSF) biomarkers can better represent the pathological process in the central nervous system (CNS) as a result of proximity to the brain and the spinal cord (Ziemssen et al. 2019). Since neuronal and axonal damage are the principal leading factor of disability in MS, neuronal and glial biomarkers can play a major role in the clinical management of MS. However, applying CSF biomarkers in clinical practice requires thorough validation. Even though the role of many CSF biomarkers has been investigated in MS, inconsistent findings hinder their clinical application. CSF oligoclonal bands were reintroduced as CSF markers to the diagnostic criteria of MS of 2017 (Thompson et al. 2018).

Markers of axonal damage, including neurofilament light chain (NFL) and total tau protein (t-tau), as well as markers of glial activation, including chitinase-3-like protein 1 (CHI3L1 or YKL-40), glial fibrillary acidic protein (GFAP), and S100B are among the commonly investigated neuronal and glial biomarkers in MS (Housley et al. 2015). Table 1 summarizes specifications of these molecules. Herein, we aimed to apply a quantitative meta-analysis to compare the CSF levels of these markers in various groups of MS, including MS versus controls, clinically isolated syndrome (CIS) versus controls, CIS versus MS, relapsing-remitting MS (RRMS) versus progressive MS (PMS), and MS in relapse versus MS in remission. We also discussed the association of these markers with sex, age, disease severity, and disease duration (Brenner 2014; Correale and Fiol 2011; Didonna 2020; Housley et al. 2015; Huizinga et al. 2012; Ising et al. 2019; Kapaki et al. 2000; Lee et al. 2011; Liedtke et al. 1998; Maphis et al. 2015; Michetti et al. 2019; Starossom et al. 2019; Yuan et al. 2012).

We also searched for unpublished “grey” literature via OpenGrey. The search terms included “neurofilament” OR “tau Proteins” OR “glial fibrillary acidic protein” OR “S100 Calcium binding protein beta subunit” OR “CHI3L1 protein” AND “multiple sclerosis” AND “cerebrospinal fluid” and the equivalent terms (Figure S1). No language or date limit was applied. Medical Subject Headings (MeSH) and Emtree were used to retrieve results from PubMed and Embase, respectively. We also traced the reference list of the relevant articles to find additional eligible studies.

**Selection criteria**

Studies were included if (1) they were peer-reviewed articles, (2) biomarkers were measured quantitatively using enzyme-linked immunoassays (ELISA) or other immunoassays, such as Single molecule array (SiMoA), (3) they included data from a control group and/or compared different subtypes of MS, and (4) the exact values of the markers were either given within the manuscript or provided by the authors of the original study for performing meta-analyses. Exclusion criteria included (1) coefficient of variation of larger than 25% (2) pediatric MS, and (3) case reports, case series, letters, commentaries, abstracts, review articles, and animal and in vitro studies. Data selection was in concordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Two authors independently performed the screening and eligibility assessment. In case of disagreement, the two authors discussed and resolved the conflict.

**Data extraction**

Two authors extracted the following data independently. We extracted (1) bibliographic information (study title, year of publication, first author, study type, and country), (2) demographic and clinical features of the sample (number of patients and controls, age, sex, disease duration, mean expanded disability status scale [EDSS] score, and medication profile), (3) methodological details (diagnostic criteria, characteristics of the ELISA assay), and (4) levels of the biomarkers. We contacted the studies’ corresponding authors for further information if the absolute values of the levels of biomarkers that were not given in the manuscript.

**Quality assessment**

The quality of the included studies was assessed according to the Newcastle–Ottawa scale (NOS) (Wells et al. 2014). Based on this scale, studies can receive 0–9 stars based on their performance in sample selection, comparability of cases and controls, and assessment of outcome. Comparability of cases and controls was determined based on whether cases and controls are age- and sex-matched.

**Statistical methods**

As the included studies were conducted in a 25-year period and were susceptible to having different ELISA assays, we estimated a standardized mean difference (SMD) (Hedge’s g) and 95% confidence interval (CI) for each between-group comparison. The SMD of ≤0.2, 0.2–0.8, and ≥0.8 represented small, moderate, and large
Table 1: Specifications of the investigated biomarkers.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Specifications of the molecule</th>
<th>Sources within the CNS</th>
<th>Function</th>
<th>Mechanism of increasing</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFL</td>
<td>NFL is a member of neurofilaments family, which are intermediate filament neuronal cytoskeletal proteins. Neurofilaments of the CNS consist of four subunits: NFL, NFM, NFH, and α-internexin. While NFL and NFH have been investigated in MS, other subunits of neurofilaments have not been widely explored in MS.</td>
<td>Neurons (exclusively)</td>
<td>They have an essential role in regulating axonal diameter and transmission of neuronal electrical signals through axons. In MS, NFL is engulfed by macrophages and microglia, which may increase autoimmunity against neuronal antigens.</td>
<td>Released subsequent to the neuroaxonal injury</td>
<td>(Huizinga et al. 2012; Yuan et al. 2012)</td>
</tr>
<tr>
<td>GFAP</td>
<td>GFAP is the major intermediate cytoskeletal protein in astrocytes reflecting astrogliosis and astroglial damage.</td>
<td>Astrocytes (exclusively)</td>
<td>It plays a role in the formation of astrocytic processes and affect production of neurons and neurites. Clinical symptoms were more severe in GFAP null EAE mice (an animal model of MS), which is probably due to malformation of scars.</td>
<td>Released from reactive astrocytes</td>
<td>(Brenner 2014; Housley et al. 2015; Liedtke et al. 1998)</td>
</tr>
<tr>
<td>Total tau</td>
<td>Tau proteins are microtubule-associated proteins, which are mainly intracellular and soluble proteins.</td>
<td>Neurons (the main source)</td>
<td>Physiologically, they are responsible for the construction and stabilization of microtubules and axonal lengthening.</td>
<td>Neuronal and axonal injury, particularly neuroinflammation, is one of the main precipitating factors in increased release of tau and tau aggregation.</td>
<td>(Didonna 2020; Ising et al. 2019; Kapaki et al. 2000; Maphis et al. 2015)</td>
</tr>
<tr>
<td>CHI3L1</td>
<td>CLPs are glycoproteins, which are members of a family of hydrolases cleaving chitin. CHI3L1 is a CLP without chitinase enzymatic activity as a result of a mutation at its catalytic site.</td>
<td>Microglia</td>
<td>Physiologically, CHI3L1 plays a prominent role in inflammatory processes, tissue remodeling, and microglial polarization. In MS, it inducing oligodendrogenesis and CHI3L1 null EAE mice showed more severe clinical symptoms. Additionally, stimulation of isolated monocytes with CHI3L1 led to increased expression of inflammatory mediators, including IL-8, MCP-1, and CCL5 and resulted in increased migratory abilities of peripheral mononuclear cells in an in vitro BBB model.</td>
<td>Released by activated microglia and astrocytes and from hippocampal neuron subsequent to neuronal injury</td>
<td>(Correale and Fiol 2011; Lee et al. 2011; Starossom et al. 2019)</td>
</tr>
<tr>
<td>S100B</td>
<td>S100B, an inflammatory molecule and a biomarker of neuronal damage, is a member of the S100 protein family, which are Ca2+/binding proteins.</td>
<td>Astrocytes (the main source)</td>
<td>Extracellularly, based on its concentration, S100B can have both neurotrophic (at physiological concentrations) and neurotrophic effects (at higher concentrations) through activation of the</td>
<td>Its excretion can be triggered by demyelinating insults</td>
<td>(Barateiro et al. 2016; Michetti et al. 2012, 2019)</td>
</tr>
<tr>
<td></td>
<td>Other glial cells (to a lesser extent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Some neuron subpopulations (to a lesser extent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1: (continued)

<table>
<thead>
<tr>
<th>Marker specifications of the molecule</th>
<th>Sources within the CNS</th>
<th>Function</th>
<th>Mechanism of increasing</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>receptor for advanced glycation end products. At higher concentrations, S100B is a damage-associated molecular pattern molecule and promotes inflammation and oxidative stress by triggering microglial and astroglial activation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


effect sizes, respectively. Meta-analyses were performed for comparisons for which results from at least three individual datasets were available.

If the values reported in the manuscript were given as median and interquartile range (IQR) or median and range and we were not able to retrieve the mean ± standard deviation (SD) from the authors, we used statistical methods suggested by Luo et al. (2018) and Wan et al. (2014) to convert these values.

To assess heterogeneity between studies in the between-group meta-analyses, we used Cochran’s Q-test and the I²-index. The I²-indices of ≤25%, 26–75%, and ≥75% represented low, moderate, and high degrees of heterogeneity, respectively (Huedo-Medina et al. 2006). We utilized fixed effects models if the results were homogeneous (I² < 40% and p > 0.05) and random effect models according to the DerSimonian and Laird method (DerSimonian and Laird 1986) if otherwise (Borenstein et al. 2010).

To reduce the heterogeneity among individual studies, in the comparison of patients with MS or CIS and controls, we also performed subgroup analyses based on the type of control groups in different studies. The control groups were classified according to the “Consensus definitions and application guidelines for control groups in CSF biomarker studies in multiple sclerosis” to three subgroups: healthy controls (HCs), non-inflammatory neurological disease controls (NINDCs), and symptomatic controls (SCs) (Teunissen et al. 2013).

To further assess the causes of heterogeneity, for meta-analyses with significant heterogeneity and including 10 or more studies, we conducted sensitivity analysis to identify influential cases. Each time we omitted one study and recalculated the effect size (Leave-One-Out Analyses). We also used the diagnostic procedure proposed by Viechtbauer and Cheung (2010) to identify influential cases.

We conducted separate regression of mean age, sex (%females), mean disease duration (years), mean EDSS scores, sample size, and score of the quality assessment on SMD whenever the required data were available for 10 or more studies in the meta-analyses comparing patients with MS and controls.

Publication bias was initially assessed by visual observation of degree of funnel plot asymmetry. Then, we used Egger’s bias test (Egger et al. 1997) and Begg-Mazumdar Kendall’s tau (Begg and Mazumdar 1994) to confirm the visual perception from the funnel plot objectively. A p-value < 0.1 was considered as evidence of publication bias. Funnel plots (Figure S2) and Egger’s plots (Figure S3) are available in the supplementary material. When there was evidence of publication bias, we adjusted the effect sizes using the trim-and-fill method (Duval and Tweedie 2000).

Except for sensitivity analysis, all other statistical analyses were performed using STATA 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC). Forest plots, funnel plots, Egger’s plots, and bubble plots were designed using STATA16. Sensitivity analysis was carried out using “dmetar” package, R (R Core Team [2020], R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). Forest plots of the sensitivity analyses were designed using R. A p-value of <0.05 was considered statistically significant.

Results

Search results and features of the included studies

The initial search resulted in 1304 findings, and four records were identified through other sources. Three hundred and thirty two of the records were duplicates. During the title/abstract screening, 866 studies did not meet the eligibility criteria and were excluded. Full-texts of the remaining 109 articles were retrieved and reviewed for potential eligibility. Forty-two studies were excluded during the full-text screening process. Sixty-seven studies were included in the qualitative analysis. In the meta-analysis, we had to exclude two studies as they had sample overlap (Håkansson et al. 2018; Malmestrom et al. 2014; Martínez et al. 2015). Lastly, 64 studies were included in the quantitative analyses. Except for one study, which was in Czech (Fialová et al. 2018), the other studies were in English. Figure 1 illustrates the process of study selection according to the PRISMA guideline.
Study and patient characteristics

Supplementary Table S1 illustrates the features of the included studies in the meta-analysis. Levels of NFL were measured in 31 studies, GFAP in 17 studies, t-tau in 20 studies, CHI3L1 in 10 studies, and S100B in eight studies. Three studies did not have a control group. Results of quality assessment of the studies are depicted in Supplementary Table S2.

NFL

Patients with MS compared to controls

In 28 of the included studies, levels of NFL were compared between patients with MS (RRMS and PMS) \((N = 1830)\) and controls \((N = 798)\). The levels of NFL were significantly higher in the CSF of patients with MS compared to controls with a large effect size \((\text{SMD} [95\%\text{CI}] = 0.96 [0.72–1.20], p-value < 0.001)\) (Figure 2, Table 2). The heterogeneity of studies was significant. Subgroup analysis showed that studies with NINDCs had the lowest heterogeneity.

Patient samples of three out of the 28 included studies were a mix of MS and CIS patients (Fialová et al. 2018; Novakova et al. 2018; Rajda et al. 2020). We excluded these studies and reran the analysis using the remaining 25 studies. The effect size remained significant and the heterogeneity reduced by near 15% \((\text{SMD} [95\%\text{CI}] = 0.78 [0.59, 0.97], p-value < 0.001, I^2 = 71.66\%, p-value < 0.001)\).

Sensitivity analysis (leave-one-out analysis) showed that after omitting each one of the studies, the effect size remained significant, and the heterogeneity did not reduce significantly (Figure S4). Studies of Rajda et al.
(2020) and Hassanpour et al. (2020) were detected as influential cases. After omission of both of them, the $I^2$ index reduced by more than 40% and the effect size remained large and significant (SMD [95%CI] = 0.719 [0.585, 0.853], $p$-value < 0.001, $I^2$ = 49.36, $Q$ = 49.37, $p$-value = 0.003). Of note, after removing the influential cases, subgroup analysis did not reveal any significant difference between subgroups with various control types ($p$-value = 0.21).

Meta-regression showed that SMD of NFL negatively correlated with sex (%females) ($p$-value = 0.027), disease duration ($p$-value = 0.018), and sample size ($p$-value = 0.007).

Figure 2: Forest plot of meta-analysis of NFL in patients with MS versus controls.
There was a trend of correlation between SMD of NFL and mean age (p-value = 0.058) (Figure 3, Table 3). The heterogeneity between studies that were included in the comparison of patients with MS and controls could be partially explained by disease duration ($R^2 = 33.1\%$), sample size ($R^2 = 10.16\%$), and sex ($R^2 = 1.58\%$). No correlation was found between the effect size and EDSS scores or NOS scores.

The Egger’s test (p-value < 0.001) and the Begg’s test (p-value < 0.001) suggested publication bias (Table S3).

After the implementation of the trim-and-fill method, no studies were imputed, and the effect size did not change.

### CIS patients compared to controls and MS

Patients with CIS ($N = 278$) were compared to controls ($N = 301$) in eight of the included studies (Figure S5). Notably, CIS patients had higher levels of NFL in CSF compared to controls (SMD [95%CI] = 0.67 [0.38, 0.96],...
p-value < 0.001). The heterogeneity between studies was almost significant ($I^2 = 48.74\%$, p-value = 0.07). Interestingly, no significant difference was observed between CIS and MS patients (Figure S6).

Importantly, the 2017 revision of the McDonald diagnostic criteria facilitated discrimination between RRMS and CIS at the first visit and reduced the proportion of patients diagnosed with CIS (Schwenkenbecher et al. 2019). To remove the potential bias caused by this revision, we identified studies utilizing the 2017 revision of the McDonald criteria that investigated CIS patients. One study (Kušínerová et al. 2020) was identified. After the removal of this study, the results remained almost the same with higher NFL levels in CIS patients compared to controls (SMD [95%CI] = 0.66 [0.35, 0.97], test of heterogeneity: $I^2 = 65.74$, p-value = 0.008) and no significant difference between CIS and MS patients (SMD [95%CI] = 0.10 [-0.04, 0.24], p-value = 0.173, test of heterogeneity: $I^2 = 37.65$, p-value = 0.142).

RRMS compared to PMS

Comparison of CSF NFL levels between patients with RRMS and PMS (secondary and primary PMS [SPMS and PPMS]) was made in 14 of the included studies (Figure S7a). We did not find any significant difference between CSF levels of NFL in RRMS ($N = 752$) compared to PMS ($N = 462$). The heterogeneity among studies was significant ($I^2 = 58.34\%$). In the sensitivity analysis, the study of Norgren et al. (2004) was detected as an inflectional case. When this case was removed, the heterogeneity among studies reduced significantly ($I^2 = 14\%$). However, the difference between the two groups remained not significant (Figure S8).

Since in the initial analysis RRMS patients were included whether they were in relapse, remission, or their disease activity was not stated, to remove the potential bias caused by including RRMS patients in relapse, we reperformed the analysis using only studies reporting NFL levels in RRMS patients in remission (Figure S7b). No significant difference was observed between RRMS and PMS patients in the second analysis as well.

Patients in relapse compared to those in remission

Twelve of the included studies compared MS patients in relapse ($N = 459$) and remission ($N = 697$). Patients with MS in relapse had higher CSF NFL levels than those in remission (SMD [95%CI] = 0.69 [0.24, 1.15], p-value = 0.003) (Figure S9a). The heterogeneity between these studies was significant ($I^2 = 90.83\%$). In the sensitivity analysis, the study of Norgren et al. (2004) was detected as an inflectional case (Figure S10). When this case was removed, the heterogeneity among studies reduced significantly ($I^2 = 42\%$) and the effect size remained significant but reduced to SMD = 0.38 (95% CI = [0.17, 0.59], p-value < 0.001).

To eliminate the potential bias and heterogeneity caused by including different MS subtypes, we reperformed the analysis using only studies with definite RRMS patients (Figure S9b). In the second analysis, the RRMS patients in relapse had higher CSF NFL levels compared to those in remission as well (SMD [95%CI] = 0.39 [0.18, 0.60]).

GFAP

Patients with MS compared to controls

Sixteen of the included studies compared patients with MS with a control group. The levels of GFAP were significantly higher in the CSF of patients with MS ($N = 1016$) compared to controls ($N = 467$) (SMD [95%CI] = 0.55 [0.44, 0.67]) (Table 2). The effect size was significant in all subgroups (Figure 4). The heterogeneity was not statistically significant ($I^2 = 38.94\%$, p-value = 0.06). Since the $I^2$ index was near the cut-off value, the random-effects model was also performed, which did not change the effect size significantly. Subgroup analysis did not show a significant difference among various control groups (p-value = 0.89). Publication bias was not detected by either Egger’s or Begg’s tests.

The patient sample of one of the 16 included studies was a mix of MS and CIS patients (Novakova et al. 2018), while the other studies contained only MS patients. We excluded this study and reran the analysis using the remaining 14 studies. The effect size remained significant (SMD [95%CI] = 0.559 [0.394, 0.724], p-value < 0.001, $I^2 = 42.78\%$, p-value = 0.04).

Meta-regression showed no association with mean age, sex, EDSS scores, sample size, and NOS scores.

RRMS compared to PMS

Comparing CSF GFAP levels between patients with RRMS ($N = 705$) and PMS ($N = 241$) was made in nine of the included studies. CSF levels of GFAP were higher in PMS compared to RRMS (SMD [95%CI] = 0.72 [0.37, 1.06]). The heterogeneity between studies was significant (Figure S11a).

Like NFL, as in the initial analysis, RRMS patients were included whether they were in relapse, remission, or their disease activity was not stated, to remove the potential bias, we re-performed the analysis using only studies
Figure 3: Results of meta-regression.
Effects of age, gender (%females), mean disease duration (years), and mean expanded disability status scale (EDSS) score on the effect size of the comparisons of patients with MS and controls was assessed wherever for 10 or more original studies data was available.
reporting GFAP levels in RRMS patients in remission (Figure S11b). In the second analysis, PMS patients still had higher GFAP levels compared to RRMS patients in remission; however, the effect size was smaller (SMD [95% CI] = 0.39 [0.02, 0.76]), and the heterogeneity among studies reduced significantly.

**Patients in relapse compared to those in remission**

Data of four studies comparing CSF levels of GFAP in relapse and remission was available. We detected no significant difference in CSF levels of GFAP between patients in relapse and remission (Figure S12a).

Since including studies with different MS subtypes may result in potential bias and heterogeneity, we reperformed the analysis using only studies with definite RRMS patients (Figure S12b). Comparably, in the second analysis, no significant difference was noted between RRMS patients in relapse and remission; however, the heterogeneity among studies reduced significantly.

**Total tau**

**Patients with MS compared to controls**

We entered data from 17 included studies comparing patients with MS (N = 990) with controls (N = 615) into the meta-analysis. Overall, CSF t-tau levels were higher in patients with MS with a moderate effect size (SMD [95% CI] = 0.35 [0.04, 0.67], p-value = 0.03). Subgroup analysis showed that only in studies with HCs, t-tau was higher in the MS group. However, the subgroup analysis detected no significant difference between different control groups (p-value = 0.12). Notably, the heterogeneity of studies was significant (Figure 5).

Patient samples of two out of the 20 included studies were a mix of MS and CIS patients (Guimaraes et al. 2006; Novakova et al. 2018). We excluded these studies and re-ran the analysis using the remaining 18 records. The overall effect size remained significant (SMD [95%CI] = 0.41 [0.05, 0.77], p-value = 0.025, I² = 89.09%, p-value < 0.001). Publication bias was not detected.

Meta-regression showed that the SMD of t-tau was correlated with disease duration (p-value = 0.009) and mean EDSS scores (p-value = 0.004). The heterogeneity between studies that were included in the comparison of patients with MS and controls could be partially explained by mean EDSS scores (R² = 34.07%), disease duration (R² = 14.92%), and sex (R² = 9.53%). No correlation was found between the effect size and mean age, the score in the risk of bias assessment, and sample size (Table 3).

Sensitivity analysis (leave-one-out analysis) showed that omission of each one of four studies (Bartosik-Psujek et al. 2011; Bartosik-Psujek and Archelos 2004; Bartosik-Psujek and Stelmasiak 2006; Terzi et al. 2007) resulted in nonsignificant effect sizes (Figure S13). The study of Bartosik-Psujek et al. (2011) was detected as an influential case, after omitting

---

**Table 3: Results of meta-regression of meta-analyses comparing patients with multiple sclerosis and controls.**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Covariate</th>
<th>Number of studies</th>
<th>Cases</th>
<th>Controls</th>
<th>Slope</th>
<th>95% confidence interval (CI)</th>
<th>p-Value</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFL</td>
<td>Mean age</td>
<td>27</td>
<td>1840</td>
<td>798</td>
<td>−0.046</td>
<td>−0.094, 0.001</td>
<td>0.056</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Percentage of females</td>
<td>26</td>
<td>1835</td>
<td>787</td>
<td>−1.940</td>
<td>−3.664, 0.217</td>
<td>0.027</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>Disease duration</td>
<td>19</td>
<td>1217</td>
<td>582</td>
<td>−0.055</td>
<td>−0.100, 0.009</td>
<td>0.018</td>
<td>33.10</td>
</tr>
<tr>
<td></td>
<td>Mean EDSS score</td>
<td>20</td>
<td>1529</td>
<td>557</td>
<td>−0.051</td>
<td>−0.199, 0.096</td>
<td>0.495</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>28</td>
<td>1849</td>
<td>798</td>
<td>−0.005</td>
<td>−0.008, −0.001</td>
<td>0.007</td>
<td>10.16</td>
</tr>
<tr>
<td></td>
<td>NOS score</td>
<td>28</td>
<td>1849</td>
<td>798</td>
<td>−0.130</td>
<td>−0.347, 0.087</td>
<td>0.239</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GFAP</td>
<td>Mean age</td>
<td>16</td>
<td>1035</td>
<td>467</td>
<td>0.002</td>
<td>−0.022, 0.026</td>
<td>0.868</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Percentage of females</td>
<td>15</td>
<td>1022</td>
<td>462</td>
<td>−1.119</td>
<td>−2.416, 0.177</td>
<td>0.091</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Disease duration</td>
<td>10</td>
<td>497</td>
<td>203</td>
<td>−0.012</td>
<td>−0.0558, 0.028</td>
<td>0.558</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Mean EDSS score</td>
<td>13</td>
<td>906</td>
<td>375</td>
<td>0.061</td>
<td>−0.079, 0.202</td>
<td>0.392</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>16</td>
<td>1035</td>
<td>467</td>
<td>&lt;0.001</td>
<td>−0.002, 0.002</td>
<td>0.813</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>NOS score</td>
<td>16</td>
<td>1035</td>
<td>467</td>
<td>−0.099</td>
<td>−0.208, 0.012</td>
<td>0.082</td>
<td>NA</td>
</tr>
<tr>
<td>T-tau</td>
<td>Mean age</td>
<td>17</td>
<td>990</td>
<td>615</td>
<td>0.002</td>
<td>−0.111, 0.115</td>
<td>0.974</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Percentage of females</td>
<td>17</td>
<td>990</td>
<td>615</td>
<td>−2.40</td>
<td>−5.494, 0.680</td>
<td>0.126</td>
<td>9.53</td>
</tr>
<tr>
<td></td>
<td>Disease duration</td>
<td>13</td>
<td>723</td>
<td>477</td>
<td>0.221</td>
<td>0.055, 0.388</td>
<td>0.009</td>
<td>14.92</td>
</tr>
<tr>
<td></td>
<td>Mean EDSS score</td>
<td>10</td>
<td>691</td>
<td>330</td>
<td>0.764</td>
<td>0.238, 1.292</td>
<td>0.004</td>
<td>34.07</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>17</td>
<td>990</td>
<td>615</td>
<td>0.0005</td>
<td>−0.008, 0.006</td>
<td>0.882</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>NOS score</td>
<td>17</td>
<td>990</td>
<td>615</td>
<td>−0.039</td>
<td>0.381, 0.303</td>
<td>0.823</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

which the effect size did not remain significant, and the I² index was reduced by more than 10% (Figure S13).

CIS patients compared to controls and MS

Patients with CIS were compared to controls in six of the included studies. Notably, patients with CIS had higher levels of t-tau in CSF compared to controls (SMD [95% CI] = 0.42 [0.04, 0.81], p-value = 0.03) (Figure S14). However, no significant difference was observed between CIS and MS patients (Figure S15).

RRMS compared to PMS

Data for the comparison of CSF t-tau levels between patients with RRMS and PMS was available in 10 of the included studies. No significant difference was observed between RRMS and PMS (Figure S16). The heterogeneity between these studies was not significant.

Similar to NFL and GFAP, we reperformed the analysis using only studies reporting t-tau levels in RRMS patients in remission to remove the potential bias caused by including RRMS patients in relapse (Figure S16b). Comparably, no significant difference was observed between RRMS and PMS patients in the second analysis either.

Patients in relapse compared to those in remission

We entered data from six studies comparing patients in relapse and remission. The difference in CSF t-tau levels between patients in relapse and remission was not significant (Figure S17a). The heterogeneity between these studies was significant.
Comparably, the reperformed the analysis using only studies with definite RRMS patients did not show any significant difference was noted between RRMS patients in relapse and remission; however, the heterogeneity among studies reduced significantly (Figure S17b).

**CHI3L1**

**Patients with MS compared to controls**

Data from nine of the included studies were entered into the meta-analysis. The levels of CHI3L1 were significantly higher in the CSF of patients with MS (N = 486) compared to controls (N = 228) with a large effect size (SMD [95% CI] = 0.96 [0.80, 1.13]). In the subgroup analysis, the effect sizes remained significant in all subgroups (Figure 6a). The heterogeneity among different subgroups of various control groups was not significant (p-value = 0.35). The heterogeneity between studies was not significant. We did not find any evidence of publication bias.

**CIS patients compared to controls and MS**

CIS patients had higher CHI3L1 levels compared to controls (SMD [95%CI] = 0.48 [0.17, 0.80]) (Figure S18). CHI3L1 was the only marker that significantly differed between CIS and MS patients with higher levels in MS with

---

**Figure 5:** Forest plot of meta-analysis of total tau in patients with MS versus controls.
a moderate effect size (SMD [95%CI] = 0.51 [0.14, 0.89]) (Figure S19). The heterogeneity between these studies was not significant.

RRMS compared to PMS

Between the subgroups of MS, the comparison of CSF CHI3L1 levels between patients with RRMS and PMS was made in six of the included studies. However, no significant difference was detected between these two subgroups (Figure S20).

Patients in relapse compared to those in remission

We entered data from four studies comparing patients in relapse and remission. The difference in CSF CHI3L1 levels between patients in relapse and remission was not significant (Figure S21).

S100B

Patients with MS compared to controls

Data comparing CSF S100B levels between MS and controls were available in six of the included studies. The levels of S100B were significantly higher in the CSF of patients with MS compared to controls with a large effect size (SMD [95% CI] = 1.11 [0.27, 1.94]) (Figure 6b, Table 2). The heterogeneity between studies was significant. However, subgroup analysis did not reveal a significant difference among studies with different control groups (p-value = 0.06). We did not find any evidence of publication bias.

Discussion

To the best of our knowledge, this is the first meta-analysis of CSF levels of GFAP, t-tau, CHI3L1, and S100B in MS patients. Figure 7 summarizes the main findings of the meta-analyses.

Neuronal and glial damage biomarkers in MS patients compared to controls

We found that CSF levels of NFL, GFAP, t-tau, CHI3L1, and S100B were higher in MS patients than controls (with a large effect size for NFL, CHI3L1, and S100B and a moderate effect size for GFAP and t-tau). However, some studies reported no significant pathological alterations in levels of GFAP (Gunnarsson et al. 2011; Hakansson et al. 2017), t-tau (Colucci et al. 2004; Guimaraes et al. 2006; Jimenez-Jimenez et al. 2002; Mori et al. 2011; Pietroboni et al. 2017), CHI3L1 (Correale and Fiol 2011), and S100B (Malmstrom et al. 2003).

CSF NFL levels have been widely explored in MS. In line with our findings, previous meta-analyses confirmed higher levels of NFL in MS patients (Bridel et al. 2019; Cai and Huang 2018; Martin et al. 2019). CSF levels of NFL are suggested as a reliable diagnostic marker for MS, particularly after correction for age with an area under the curve (AUC) of 0.923 (Rajda et al. 2020). However, since every pathological process that damages axons can lead to an increase in NFL, it may be an unspecific biomarker.

Of note, we should interpret the results regarding changes of CSF levels of t-tau in MS cautiously as the sensitivity analysis showed evidence of a small-study effect, and the effect size did not remain significant after omitting the influential case. Individual studies assessing CSF levels of t-tau in MS are inconsistent, as outlined above. Whether tau accumulations have a role in the pathogenesis of the disease is not fully elucidated. Studies in experimental autoimmune encephalomyelitis (EAE) animal models of MS showed abnormal tau phosphorylation and association of tau aggregation with neuroaxonal loss (Anderson et al. 2008; Didonna 2020). Some studies suggested higher levels of phosphorylated (p)-tau in addition to the increased t-tau in MS (Bartosik-Psujek and Stelmasiak 2006), while some other studies reported no significant alteration in levels of p-tau (Szalardy et al. 2013; Valis et al. 2008).

Postmortem studies also confirmed pathological alterations in the release of neuronal and glial markers in MS. They suggested higher expression of GFAP in the cortex of MS patients (Petzold et al. 2002), aggregation of insoluble tau in the brain tissues of SPMS patients (Anderson et al. 2008), and higher expression of S100B in both white and gray matter of MS patients compared to controls. Expression of S100B in acute lesions was found to be nearly two times higher than subacute lesions (Petzold et al. 2002).

Neuronal and glial damage biomarkers in CIS and early MS

We found that CIS patients showed higher CSF levels of NFL, t-tau, and CHI3L1 compared to controls with a moderate effect size. However, only CHI3L1 was higher in MS compared to CIS. Higher CSF levels of CHI3L1 are also reported in RRMS compared to radiologically isolated syndrome (RIS) patients (Thouvenot et al. 2019). As production
### a) CHI3L1

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients with CDMS</th>
<th>Controls</th>
<th>SMD with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>HC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonneh-Barkay et al., 2010</td>
<td>10</td>
<td>340.646</td>
<td>176.189</td>
<td>12</td>
</tr>
<tr>
<td>Correa et al., 2010</td>
<td>48</td>
<td>96.5</td>
<td>24.4915</td>
<td>24</td>
</tr>
<tr>
<td>Burman et al., 2016</td>
<td>117</td>
<td>148.794</td>
<td>108.115</td>
<td>30</td>
</tr>
<tr>
<td>Novakova et al., 2016</td>
<td>59</td>
<td>136.237</td>
<td>51.9201</td>
<td>39</td>
</tr>
<tr>
<td>Sellebjerg et al., 2017</td>
<td>52</td>
<td>179.938</td>
<td>82.494</td>
<td>24</td>
</tr>
<tr>
<td>Hakansson et al., 2017</td>
<td>22</td>
<td>131.637</td>
<td>86.9758</td>
<td>22</td>
</tr>
<tr>
<td>Kulićević et al., 2020</td>
<td>42</td>
<td>139.35</td>
<td>68.8</td>
<td>15</td>
</tr>
<tr>
<td>Heterogeneity: $I^2 = 0.00%$, $H^2 = 0.73$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of $\theta \neq 0$: Q(6) = 4.38, p = 0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SC**

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients with CDMS</th>
<th>Controls</th>
<th>SMD with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Thouvenot et al., 2018</td>
<td>50</td>
<td>277.258</td>
<td>194.795</td>
<td>42</td>
</tr>
<tr>
<td>Huss et al., 2020</td>
<td>86</td>
<td>157.217</td>
<td>76.215</td>
<td>20</td>
</tr>
<tr>
<td>Heterogeneity: $I^2 = 0.00%$, $H^2 = 0.20$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of $\theta \neq 0$: Q(6) = 0.20, p = 0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Overall**

| Heterogeneity: $I^2 = 0.00\%$, $H^2 = 0.68$ |     |         |      |     |        |      |                  |            |
| Test of $\theta \neq 0$: Q(8) = 5.44, p = 0.71 |     |         |      |     |        |      |                  |            |
| Test of group differences: $Q_{(1)} = 0.86$, p = 0.35 |     |         |      |     |        |      |                  |            |

Fixed-effects inverse-variance model

### b) S100B

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients with CDMS</th>
<th>Controls</th>
<th>SMD with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>HC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rejdak et al., 2007</td>
<td>20</td>
<td>459.4</td>
<td>214.1</td>
<td>10</td>
</tr>
<tr>
<td>Rejdak et al., 2008</td>
<td>34</td>
<td>577.8676</td>
<td>851.3297</td>
<td>12</td>
</tr>
<tr>
<td>Bartosik-Pszüge et al., 2011</td>
<td>54</td>
<td>234</td>
<td>11.5</td>
<td>30</td>
</tr>
<tr>
<td>Heterogeneity: $I^2 = 93.03%$, $H^2 = 14.35$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of $\theta \neq 0$: Q(2) = 28.70, p = 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NINDC**

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients with CDMS</th>
<th>Controls</th>
<th>SMD with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Pezold et al., 2002</td>
<td>51</td>
<td>430</td>
<td>422</td>
<td>51</td>
</tr>
<tr>
<td>Baralirai et al., 2016</td>
<td>11</td>
<td>1122</td>
<td>112.2</td>
<td>13</td>
</tr>
<tr>
<td>Heterogeneity: $I^2 = 34.3%$, $H^2 = 3.72$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of $\theta \neq 0$: Q(1) = 3.72, p = 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SC**

<table>
<thead>
<tr>
<th>Study</th>
<th>Patients with CDMS</th>
<th>Controls</th>
<th>SMD with 95% CI</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Green et al., 1997</td>
<td>40</td>
<td>180</td>
<td>90</td>
<td>19</td>
</tr>
<tr>
<td>Misu et al., 2009</td>
<td>10</td>
<td>137.7</td>
<td>51.04</td>
<td>5</td>
</tr>
<tr>
<td>Heterogeneity: $I^2 = 37.11%$, $H^2 = 6.61$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of $\theta \neq 0$: Q(1) = 1.61, p = 0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Overall**

| Heterogeneity: $I^2 = 1.12\%$, $H^2 = 9.16\%$ |     |         |      |     |        |      |                  |            |
| Test of $\theta \neq 0$: Q(6) = 68.11, p = 0.00 |     |         |      |     |        |      |                  |            |
| Test of group differences: $Q_{(2)} = 5.63$, p = 0.06 |     |         |      |     |        |      |                  |            |

Random-effects DerSimonian-Laird model

**Figure 6:** Forest plots of meta-analyses of CHI3L1 and S100B in patients with MS versus controls.  
(a) Forest plot of meta-analysis of CHI3L1 in patients with MS versus controls, (b) Forest plot of meta-analysis of S100B in patients with MS versus controls.
of CHI3L1 is mainly triggered by inflammatory processes, increased oxidative stress (Mossakowski et al. 2015) and inflammation (Rossi et al. 2015) promoted by activated microglia and astrocytes in RRMS and PMS could be the underlying reason for differences in the levels of this marker between CIS and MS patients. The number of studies investigating CSF levels of GFAP and S100B in CIS patients was not adequate for performing a statistical analysis. However, no difference was reported between CSF levels of S100B in patients with CIS and controls (Hein Nee Maier et al. 2008) or MS (Martínez et al. 2015). Additionally, Martinez et al. did not find any significant difference in levels of NFL, t-tau, CHI3L1, GFAP, and S100B between patients with CIS and RRMS (Martínez et al. 2015).

In line with our findings, a recent meta-analysis found no significant difference between CSF NFL levels in CIS and MS (Bridel et al. 2019). Moreover, no significant difference was reported between CSF NFL levels in patients with RIS and HC as well (Pawlitzki et al. 2018). However, the rise of NFL levels seems to be among the early signs of MS. A recent study showed that MS patients had modest increased serum levels of NFL years before clinical onset (Bjornevik et al. 2020).

Our analysis also found that CIS patients had higher levels of t-tau than controls. Notably, the effect size of the comparison of CIS and controls was larger than the effect size of the comparison of MS and controls. While we did not find any significant difference in CSF t-tau levels between CIS and MS patients, some of the previous
Neuronal and glial damage biomarkers in differentiating subtypes of MS

For differentiating different subtypes of MS, only GFAP appeared a useful biomarker. Patients with PMS had higher GFAP levels compared to RRMS. On the contrary, CSF levels of NFL, CHI3L1, and t-tau did not significantly differ between RRMS and PMS patients. The number of studies investigating CSF levels of S100B in RRMS and PMS patients separately was not adequate for performing a statistical analysis. However, no significant difference was found in CSF levels of S100B between patients with RRMS and SPMS (Bartosik-Psujek et al. 2011; Mane-Martínez et al. 2016).

While in RRMS, the movement of adaptive immune cells from the periphery into the CNS is the principal pathological mechanism, in PMS, players of innate immunity, including astrocytes and microglia, have a more prominent role. During the pathogenesis of PMS, through producing neurotoxic molecules inducing immune-independent mechanisms, astrocytes are one of the main drivers of the disease activity (Baecher-Allan et al. 2018). Moreover, molecular biomarkers of reactive astrogliosis have shown promising results in differentiating RRMS and PMS (Barbour et al. 2017). This can be one of the explaining reasons for the higher GFAP levels, which reflect astrogliosis, in PMS patients compared to RRMS. Serum levels of GFAP were also higher in SPMS patients compared to RRMS (Hogel et al. 2020). Considering the potential positive correlation between age and GFAP levels, since subjects in most of the included studies were not age-matched, and PMS patients are generally older, higher levels of GFAP in PMS patients should be interpreted cautiously. The inconsistencies in the literature might also be due to confounding factors, such as age and sex.

Interpretation regarding the difference of NFL between subgroups of MS should be made cautiously. Our results are in line with a recent meta-analysis, including some unpublished data (Bridel et al. 2019). However, sensitivity analysis showed evidence of a small-study effect. After omitting the influential case, CSF NFL levels were higher in RRMS compared to PMS. This is in line with the meta-analysis of Martin S-J et al. which was performed on five original studies (Martin et al. 2019). Notably, the total number of PMS cases analyzed in our study was higher than both of the previous meta-analyses. Moreover, inconsistent findings were found in individual studies. The inconsistency can be explained by confounding factors, including age, sex, percentage of patients receiving disease-modifying treatments (DMTs), recent relapse, and EDSS scores (Williams et al. 2020). For instance, Gunnarson et al. showed that before treatment with natalizumab, there was no significant difference between CSF NFL levels in SPMS and RRMS; however, after the treatment, SPMS patients had higher CSF levels of NFL compared to RRMS (Gunnarsson et al. 2011). However, their relatively small sample size may hinder drawing a definite conclusion. Notably, higher CSF levels of NFL were a predictor for the earlier conversion of RRMS to SPMS (Gil-Perotin et al. 2019).

Our analysis did not find any significant difference between CSF levels of t-tau in RRMS compared to PMS. However, the findings of individual studies investigating this comparison are inconsistent. While lower levels of CSF t-tau are reported in the CIS/RRMS group compared to PMS (Novakova et al. 2018), some studies reported higher CSF t-tau in RRMS compared to SPMS (Jaworski et al. 2012; Kosehasanogullari et al. 2015), and a larger number of studies found no difference between these subtypes (Bartosik-Psujek et al. 2011; Bretschneider et al. 2005; Guimaraes et al. 2006; Terzi et al. 2007).
We also did not find any significant difference between CSF levels of CHI3L1 in RRMS compared to PMS. Similar to t-tau, probably as a result of the confounding factors, individual studies reported inconsistent findings of higher CSF CHI3L1 levels in the PMS (vs. RRMS) (Gil-Perotin et al. 2019; Huss et al. 2020) or in the RRMS (vs. PMS) groups (Correale and Fiol 2011).

Association of neuronal and glial damage biomarkers with incidence of relapse, disease activity, and severity

For discriminating between active and nonactive patients, only NFL was higher in patients in relapse compared to patients in remission with a moderate effect size, which is in line with the meta-analysis of Martin et al. (2019). We found no significant difference in CSF levels of GFAP, CHI3L1, and t-tau between patients in remission and relapse. This is in line with studies finding no correlation between CSF levels of GFAP and frequency of relapses (Norgren et al. 2004; Rosengren et al. 1995). However, Martinez et al. reported higher CSF levels of GFAP in patients with active disease (Hakansson et al. 2017). Individual studies were inconsistent regarding CHI3L1 levels (suggesting higher levels in patients in relapse (Novakova et al. 2016) or no significant difference (Malmestrom et al. 2014; Martinez et al. 2015). While due to the few original datasets available for meta-analysis of CSF levels of S100B in patients in relapse and remission, we were not able to perform statistical analysis; no difference in CSF levels of S100B was reported between patients in relapse and remission (Martinez et al. 2015).

NFL can be useful in determining disease activity as its release is increased during acute neuroaxonal injury and demyelination, which are postulated to be the underlying mechanism of acute relapses (Ross et al. 2013). Individual studies comparing NFL levels between patients in relapse and in remission were controversial. While some found higher levels of NFL in patients with active disease (Aeinehband et al. 2015; Hakansson et al. 2017; Malmestrom et al. 2003; Martinez et al. 2015; Sellebjerg et al. 2017), some found no significant difference (Hakansson et al. 2017). Notably, the increase in CSF levels of NFL persisted for several weeks after the onset of relapse (Burman et al. 2014). NFL was also a predictor of disease activity and was associated with the number of relapses (Gil-Perotin et al. 2019; Hakansson et al. 2017; Novakova et al. 2016).

It can be presumed that markers that are higher in MS patients and the levels of which are not affected by presence of relapse may be continuously secreted. As a result, they are independent of the time of lumbar puncture and can be a more reliable marker.

Disease severity (measured by EDSS or Multiple Sclerosis Severity Score [MSSS]) is reported to be positively correlated with CSF levels of NFL (Bergman et al. 2016; Norgren et al. 2004), GFAP (Axelsson et al. 2011; Burman et al. 2014; Malmestrom et al. 2003; Norgren et al. 2004; Petzold et al. 2002; Rosengren et al. 1995), t-tau (in CIS or RRMS) (Brettschneider et al. 2006), and CHI3L1 (Gil-Perotin et al. 2019; Novakova et al. 2016; Perez-Miralles et al. 2020). Early disability progression also correlated with NFL (Hakansson et al. 2017; Trentini et al. 2014), GFAP (Axelsson et al. 2011; Martinez et al. 2015; Norgren et al. 2004), CHI3L1 (Gil-Perotin et al. 2019; Martinez et al. 2015). Our meta-regression also showed a positive correlation between mean EDSS scores and the effect size of the meta-analysis of t-tau in MS compared to controls. However, the literature on the association of these markers with disease severity is not consistent. For example, a negative correlation between CSF tau levels and EDSS in RRMS or SPMS was also reported (Jaworski et al. 2012). However, several studies did not find a correlation between CSF NFL (Hakansson et al. 2017), t-tau (Colucci et al. 2004; Guimaraes et al. 2006; Szalardy et al. 2013; Terzi et al. 2007), and CHI3L1 (Sellebjerg et al. 2017) levels and disease severity.

Neuronal and glial damage biomarkers may also reflect disease duration. A positive correlation between CSF GFAP (Abdelhak et al. 2019; Novakova et al. 2018) and t-tau (Terzi et al. 2007) and disease duration and a negative correlation with NFL (Gil-Perotin et al. 2019) has been reported. Our meta-regression also showed a weak positive correlation between mean disease duration and the effect size of the meta-analysis of t-tau in MS compared to controls. However, some studies found no correlation between levels of NFL (Norgren et al. 2004), GFAP (Norgren et al. 2004), and t-tau (Guimaraes et al. 2006; Jimenez-Jimenez et al. 2002; Kapaki et al. 2000) and disease duration. No correlation was either found between t-tau CSF levels and age at onset (30).

These markers may correlate with markers of neuroinflammation. The CSF levels of NFL positively correlated with the number of inflammatory cells (Norgren et al. 2004), mononuclear cells (Hakansson et al. 2017), and quinolinic acid, which is a metabolite of the kynurenine pathway and a marker of neuroinflammation (Rajda et al. 2020) in CSF. CSF tau levels also correlated with immunoglobulin (IgG index (Bartsis-Psujek and Archelos 2004), while this finding was not supported by two other studies (Szalardy et al. 2013; Terzi et al. 2007). Neuroinflammation can aggravate neurodegeneration via inflammatory mediators in MS (Ising and Heneka 2018; Kempuraj et al. 2016).
Association of neuronal and glial damage biomarkers with age and sex

We found a trend for a negative correlation between patients’ mean age and the effect size of the comparison of CSF NFL levels between MS patients and controls. Age may be a key determinant of CSF levels of neuronal and glial damage biomarkers. Several studies reported a positive correlation between age and CSF NFL (Sellebjerg et al. 2017), GFAP (Axelsson et al. 2011; Martínez et al. 2015), CHI3L1 (Gil-Perotín et al. 2019; Sellebjerg et al. 2017), and S100B (Martínez et al. 2015) levels. However, a negative correlation between CSF NFL and age (Khademi et al. 2013; Martínez et al. 2015) or no correlation between age and NFL (Hakansson et al. 2017) or CHI3L1 (Huss et al. 2020) have also been reported. A recent meta-analysis showed that CSF NFL levels positively correlate with age in HC, but they had no or negative correlation with age in MS (Bridel et al. 2019). The pathological changes during MS affecting CSF NFL levels seem to be the underlying cause of the difference between HC and MS patients.

Our meta-regression showed a negative correlation between the percentage of females and the effect size of the comparison of CSF NFL levels between MS patients. Sex can be another determinant of CSF levels neuronal and glial damage biomarkers. Higher CSF levels of CHI3L1 and t-tau are reported in men with MS (Martínez et al. 2015). A recent meta-analysis found higher CSF NFL levels in men in HC and MS groups (Bridel et al. 2019). However, in PMS patients, CSF NFL levels were moderately higher in women (Sellebjerg et al. 2017).

Lastly, in addition to the CSF levels of neuronal and glial damage biomarkers, their blood level may also play a practical biomarker in MS. The CSF and blood levels of these biomarkers can be affected by DMT, and they may be potentially used in monitoring treatment response. These topics are further discussed in the supplementary material (Supplementary document S3).

Future directions

Future studies are needed to further investigate clinical applications of neuronal and glial biomarkers in MS. More studies are required to illuminate the role of these markers in distinguishing different subtypes of MS and the disease course. Determining cut-off values for different biomarkers in the diagnosis of MS and determining prognosis can be useful. While several studies have sought to answer this question, more studies with large sample sizes are required. Moreover, defining novel diagnostic and prognostic models combining levels of different biomarkers with each other or with MRI markers seems to be a promising tool in the diagnosis and prognosis of MS (Brettschneider et al. 2006; Huss et al. 2020). Additionally, the potential role of these markers in monitoring treatment response should be more explored (Gaetani et al. 2019).

Given the inflammatory role of molecules such as S100B and CHI3L1, modulation of these molecules’ activity may also provide novel therapeutic targets. For instance, suppressing the activity of CHI3L1 in vitro resulted in reduced release of inflammatory mediators from isolated monocytes (Correale and Fiol 2011). Moreover, inhibition of S100B in an ex-vivo demyelinating model also resulted in lower inflammatory responses (Barateiro et al. 2016).

Limitations

This study has some limitations. First, in several of the included studies, the MS and control groups were not age- and sex-matched. Given the possible impact of age, sex, the presence of a relapse, and use of DMTs, it is critical that groups of comparison be matched for the potential confounding factors in future studies. Second, the sample sizes of the included studies were relatively small. This highlights the substantial need for large studies assessing levels of these biomarkers in different subgroups of MS. Third, due to small sample sizes, in the comparisons made between patients in relapse and remission, studies with different MS subtypes, and in the comparisons made between RRMS and PMS patients, RRMS patients with active and non-active disease stage were combined to reach a larger total population. We addressed the potential bias caused by the combining in the analyses performed for NFL, GFAP, and t-tau, by re-calculating the comparisons using only studies with definite RRMS in the comparisons made between patients in relapse and remission and studies with RRMS in remission in the comparisons made between RRMS and PMS patients. However, for CHI3L1, the number of remaining eligible studies was not enough to compare these groups while removing the potential bias caused by the combination. Fourth, a large proportion of the studies were conducted with patients of European ethnicity. Therefore, more studies are required to assess the levels of these biomarkers in a variety of ethnic groups. Fifth, other neuronal and glial damage markers, including amyloid-beta, NFH, and p-tau, have also been investigated in MS (Kuhle et al. 2011; Szalardy et al. 2013). However, due to the insufficient number of studies investigating their CSF levels and variability of their measurement methods, we
could not conduct a meta-analysis on their levels in MS. Moreover, we did not investigate myelin basic protein (MBP), a myelin damage biomarker in MS. Sixth, we used statistical methods suggested by Luo et al. (23) and Wan et al. (24) to convert values given in median (IQR) to mean (SD). However, many meta-analyses have used this method (Bekiari et al. 2018; Kim et al. 2017; Ovadia et al. 2019). Lastly, despite multiple emails and reminders, some authors failed to respond to our request for their data. As a result, we had to exclude some studies due to not having the required data for conducting a meta-analysis.

**Conclusion**

To summarize, NFL, GFAP, t-tau, CHI3L1, and S100B were higher in MS patients compared to controls. Moreover, NFL, t-tau, and CHI3L1 may also be used in differentiating CIS from controls while CHI3L1 was the only marker with higher levels in MS than CIS. We found that GFAP can be helpful in differentiating RRMS and PMS. Additionally, NFL and CHI3L1 were higher in patients in relapse compared to remission. Our meta-regression suggested that age, sex, disease duration, EDSS scores, and quality of the individual studies could affect the effect sizes, which needs to be taken into account during the interpretation of individual findings.

Though not included in the meta-analysis, levels of neuronal and glial damage markers have been reported to correlate with disease severity, activity, duration, and imaging biomarkers. Some of them may also be used in monitoring treatment response in MS. However, considering the controversy between different studies, more studies with age- and sex-matched participants are required to elucidate the diagnostic and prognostic roles of these markers to assess a potential clinical utility in the future.

**Acknowledgements:** The authors thank Dr. Iris Kleerekooper, Dr. Maja Jagodic, and Dr. Sara Gil-Perotin for providing additional data for the meta-analysis.

**Author contributions:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

**Research funding:** This systematic review and meta-analysis was supported by a grant from Tehran University of Medical Sciences (No. 50866). LS was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund. AB was supported by PROGRES Q35, Charles University.

**Conflicts of interest statement:** Not applicable.

**Availability of data, material and codes:** The data used in this manuscript are available from the corresponding author upon reasonable request.

**References**


central inflammation is risk for RIS and CIS conversion to MS. Mult. Scler. 21: 1463–52.


Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/revneuro-2020-0145).