Journal of Neurology (2026) 273:95
https://doi.org/10.1007/s00415-026-13622-6

REVIEW

Accuracy of the central vein sign in diagnosis of multiple sclerosis: a systematic review and meta-analysis

Ludovico Baiamonte[1] · Francesco Sera[2] · Giovanna Bellante[1] · Paolo Ragonese[3] · Giuseppe Salemi[1]

Received: 2 December 2025 / Revised: 5 January 2026 / Accepted: 7 January 2026

Abstract

Background The central vein sign (CVS) is a promising imaging marker of multiple sclerosis (MS). We performed a systematic review and meta-analysis to evaluate the diagnostic accuracy of CVS-based rules in the differential diagnosis of MS and to identify the best cutoffs for these rules.

Methods PubMed, Embase, and Scopus were systematically searched for available evidence. Data extracted were entered into Bayesian models recommended by the Cochrane network. Summary sensitivity and specificity of CVS-based diagnostic rules across different positivity thresholds were calculated. A meta-regression for the role of gadolinium-based MRI protocols was also performed.

Results 3434 patients from 28 studies were included. Three CVS-based diagnostic rules were found: the first one considers the percentage of CVS + lesions (relative threshold method), the second one the presence of CVS in a given number of lesions selected in T2 sequences (select-n method), and the third one the presence of a given number of CVS + lesions in gradient-echo sequences (select-n* method). For relative threshold method, the best cutoff was 37.5% (sensitivity 97.3%, 95%CI 90.9–99.6%; specificity 90.4%, 95%CI 83.2–95.9%; Youden index 0.877); for select-n* method, 4 was the best threshold (sensitivity 87.1%, 95%CI 66.9–96.6%; specificity 88.2%, 95%CI 65.1–98.1%; Youden index 0.753). Use of gadolinium-based MRI protocols was irrelevant (for relative threshold method RDOR = 6.62, 95%CI 0.68–71.27; for select-n* method RDOR = 2.52, 95%CI 0.35–15.36).

Conclusions Relative threshold and select-n* methods are good predictors of MS diagnosis. This synthesis should support the use of CVS in clinical practice and prompt further research.

Keywords Central vein sign · Multiple sclerosis · Magnetic resonance · Meta-analysis · Diagnsotica criteria

Introduction

Multiple sclerosis (MS) is the most common autoimmune disorder of the central nervous system (CNS) and the paradigm of demyelinating diseases. MS diagnosis is based on clinical history and findings, supported by laboratory and magnetic resonance imaging (MRI) data [1]. The canonical approach in MRI studies oriented to MS is based on the acquisition of T2-weighted and gadolinium-enhanced images; however, the need to improve sensitivity and specificity of diagnosis has led the search for new imaging markers. Among them, one of the most studied markers has been the central vein sign (CVS), i.e., the visualization of a vein at the center of an MS plaque on MRI, as it can be observed in gradient-echo (GRE) sequences like T2*, FLAIR*, or susceptibility-weighted imaging (SWI) [2, 3]. Interest in CVS has been growing in recent years, and in 2015 the North American Imaging in Multiple Sclerosis (NAIMS) Cooperative Group developed a set of criteria to define CVS [4]. The CVS, a radiological counterpart of the perivenular distribution of MS lesions in pathological specimens [5, 6], has been included in the 2024 McDonald diagnostic criteria for MS [7]. Considering these recent developments, we undertook a systematic review and meta-analysis to provide

✉ Paolo Ragonese
ragonesepaolo@gmail.com

1 Department of Biomedicine, Neuroscience and Advanced Diagnostics, University of Palermo, Palermo, Italy

2 Department of Statistics, Computer Science and Applications "G. Parenti", University of Florence, Florence, Italy

3 Neurology and Multiple Sclerosis Center, Foundation Institute "G. Giglio", Contrada Pietrapollastra – Pisciotto, 90015 Cefalù, Italy

Published online: 25 January 2026

a comprehensive summary of available knowledge on this topic.

Methods

We performed a systematic review and meta-analysis, adhering to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [8], and reported our findings according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [9, 10]. Our review was registered on the International Prospective Register of Systematic Reviews (PROSPERO; CRD42025640253).

Data sources and search strategy

We systematically searched PubMed, Embase (via Ovid), and Scopus databases from inception to January 13, 2025. Polyglot tool [11], part of the Systematic review accelerator (SRA) [12], was used to assist us in string conversion between different sources. The search strings are included in eMethods 1 in Supplemental material. The reference list of studies selected for inclusion and published systematic reviews on the same topic were also screened for studies that met our inclusion criteria. Following search, retrieved titles were collated in Zotero software, then duplicated items were removed using the Deduplicator tool [13] of the SRA. Screening by title and abstract was performed by two reviewers using Rayyan software [14]; disagreements were solved by discussion between the two authors and, if needed, through involvement of a third one. The same method was used for full-text screening.

Eligibility criteria

We included studies that: 1) explored the accuracy of the diagnostic rules based on the CVS in distinguishing MS from other diseases; 2) provided enough information to extract a 2×2 contingency table with the values for true positive (TP), false positive (FP), false negative (FN), and true negative (TN) results; 3) had been peer reviewed; 4) had a prospective or retrospective design; 5) included patients with MS diagnosed according to 2010 or 2017 McDonald criteria (or at least did not declare to have used other criteria); 6) were based on the acquisition of T2*, FLAIR*, or SWI on high-field (1.5 or 3 T) or ultra-high-field (7 T) MRI. We excluded studies: 1) written in languages other than English; 2) involving pediatric patients or animal models; 3) including less than five cases; 4) involving cases of radiologically isolated syndrome (RIS). Reviews, guidelines, commentaries, conference proceedings and abstracts, and surveys were excluded too.

Data extraction

Data extraction was performed using an electronic sheet. To extract data from plots, Plotdigitizer tool was used [15]. Extracted data for each study regarded authors, publication year, country, patient enrollment strategy (prospective or retrospective), age limits, technical details of MRI acquisition (field strength, contrast agent use, used sequences, use of automated search of CVS), number of participants with and without MS, MS type, clinical condition of non-MS subjects, demographics (age, sex) of each group of participants, type of threshold used, threshold value(s), and the number of TP, FN, TN, and FP at each threshold reported by the study.

Outcomes

The main outcome was the identification of summary estimates of sensitivity and specificity of CVS-based diagnostic rules for MS and the identification of the best cutoff for each of these rules. Secondary outcomes included evaluation of the risk of bias and reproducibility of included studies and exploration of possible sources of heterogeneity.

Risk of bias assessment

Assessment of risk of bias and reproducibility of included studies was performed using the QUADAS-2 tool, as recommended by the Cochrane collaboration [8, 16]. QUADAS-2 results were presented as graphical displays and within the text.

Data synthesis and analysis

For each diagnostic rule, to account for studies providing the estimates of positivity and specificity at different thresholds, the model proposed by Jones et al. [17] was used. Through this model, based on multinomial distributions, summary true positive rate (TPR, i.e., sensitivity) and false positive rate (FPR, i.e., 1-specificity) were calculated for different positivity thresholds. Based on these data, Youden's J statistic was calculated and the optimal cutoff was found for each diagnostic rule.

To calculate the diagnostic odds ratio (DOR) of each diagnostic rule, the hierarchical summary receiving operating characteristic (HSROC) model by Rutter and Gatsonis [18] was used, including for each study only the threshold with the best Youden's index.

All analyses were run in a Bayesian framework. We used RStudio version 2023.12.1 [19] with R version 4.3.2 [20] and JAGS version 4.3.1 [21]. Ninety-five percent credible intervals were calculated for each estimated value. See eMethods 2–3 in Supplemental material for code used in these analyses.

Investigation of heterogeneity

For each decision rule, we investigated heterogeneity by performing a meta-regression analysis on the role of gadolinium-based vs gadolinium-free MRI protocols. The impact of NAIMS criteria was not explored, although it was planned in the review protocol, due to the strong interaction between the adoption of these criteria and publication year. For meta-regression, the aforementioned HSROC model with proper adaptations for covariates was used; results were expressed as ratio of the pooled DORs (RDOR) of groups of studies differing for the covariate of interest. See eMethods 4 in Supplemental material for the code used in this analysis.

Results

Study selection

The systematic search of articles identified 1336 records. After removal of duplicated ones, 764 records were left. After title and abstract screening, 39 records were considered potentially relevant; we were unable to retrieve 1 record, which was therefore excluded. The other three records were found through manual search. Following full-text screening, 28 articles were finally included [22–49]. A detailed flowchart of the study selection process according to PRISMA guidelines can be found in eFigure 1 in Supplemental material. eTable 1 in Supplemental material lists articles excluded on full text screening along with the reason for exclusion.

Study characteristics

Overall, 3434 participants were included from the 28 selected studies. Among studies providing enough information, participants' mean age was 45 years and women's proportion was 68.4%. Patient enrollment was performed prospectively in 16 studies [22, 23, 25, 26, 28, 30, 32–35, 37, 40, 43, 45, 48, 49] and retrospectively in 9 [24, 27, 29, 31, 36, 42, 44, 46, 47]. The publication years ranged from 2013 to 2025. Thirteen studies were conducted in European countries [22, 24, 28, 31, 34, 37, 38, 40, 43, 44, 46, 48, 49], 7 in American countries [25, 27, 29, 35, 42, 45, 47], 2 both in America and Europe [23, 36], 3 in Asia [26, 32, 39], and 3 in Africa [30, 33, 41]. See Table 1 for more details.

Definition of MS and no-MS cases

Two thousand and thirty-seven patients with MS were included in selected studies. In 15 studies, the 2017 version of McDonald criteria was used for diagnosis [22–32, 35–37, 44], 7 used the 2010 revision [38, 39, 42, 43, 45–47], and 1 study used both versions [34]. Eight studies included all MS types [22–26, 37, 38, 48] and three of them included clinically isolated syndrome (CIS) too [22, 24, 38]; six included only relapsing–remitting (RRMS) cases [28, 31, 40, 42, 43, 46], one both primary progressive (PPMS) and RRMS [44], and the remaining ones [27, 29, 30, 32–36, 39, 41, 45, 47, 49] did not specify the type of MS included.

Our review included 1397 no-MS participants (people with an MS mimic to be distinguished from MS); their disease status varied, including asymptomatic white matter lesions (WML), cerebral small vessel disease (CSVD), neuromyelitis optica spectrum disorder (NMOSD), myelin oligodendrocyte glycoprotein antibody-associated disease (MOGAD), migraine, and Fabry disease. For more information, see Table 2.

CVS detection methods

Thirteen studies used T2* sequences to detect CVS [22, 23, 30, 34–37, 43–46, 48, 49], 9 SWI sequences [26, 28, 29, 32, 33, 39–42], 3 both types of sequences [24, 26, 38], 2 FLAIR* sequences [25, 47], and one used SPGR sequences [31]. The acquisition protocol of 11 studies included the administration of paramagnetic contrast agent [22, 23, 25, 35–37, 39, 43, 45–47]. All studies but five [27, 41, 47–49] used NAIMS criteria for MS definition (Table 2). One study proposed automated detection methods for CVS along with manual ones [36] (however, for comparability reasons, only the results regarding manual detection were considered).

CVS-based diagnostic rules

Three main diagnostic rules based on CVS were found: the first based on the percentage of CVS + lesions; the second based on the presence of CVS in a given number of lesions selected in T2 imaging; the third basede on a minimum number of CVS + lesions as seen in GRE imaging. No uniformity in naming these methods was found across included studies; for convenience purposes and looking at most used denominations, we called them "relative threshold method", "select-n method" and "select-n* method", respectively. eTables 2–4 in Supplemental material give further details about methods used in each study and their naming.

Table 1 Characteristics of included studies

Study	Country	Age, years	Participants, no	Women, no. (%)	Patient enrollment	MS cases, no	MS types	No-MS cases, no	No-MS cases condition
Silvestri et al. [22]	Italy	19–77	112	89 (79,5)	Prospective	70	All (and CIS)	42	Migraine
Borrelli et al. [23]	Europe, USA	> 18	285	182 (63,9)	Prospective	185	All	100	MS mimics
Cagol et al. [24]	Europe	> 18	934	NA	Retrospective	570	All (and CIS)	364	MS mimics
Daboul et al. [25]	USA	18–65	78	55 (70,5)	Prospective	37	All	41	MS mimics
Gao et al. [26]	China	14–68	160	117 (73,1)	Prospective	73	All	87	NMOSD
Okromelidze et al. [27]	USA	NA	100	84 (84)	Retrospective	61	NA	39	MS mimics
Cortese et al. [28]	UK	> 18	53	NA	Prospective	29	RRMS	24	NMOSD
Ciotti et al. [29]	USA	NA	33	NA	Retrospective	19	NA	14	MOGAD
Ghany et al. [30]	Egypt	18–59	42	30 (71,4)	Prospective	14	NA	28	MS mimics
Tranfa et al. [31]	Italy	NA	109	84 (77,1)	Retrospective	73	RRMS	36	Fabry disease
Yavas et al. [32]	Turkey	18–65	169	103 (60,9)	Prospective	89	NA	80	Leukoaraiosis
Anan et al. [33]	Egypt	17–70	50	28 (56)	Prospective	30	NA	20	CSVD
Clarke et al. [34]	UK	> 18	37	27 (73)	Prospective	23	NA	14	MS mimics
Kaisey et al. [35]	USA	NA	30	24 (80)	Prospective	15	NA	15	MS mimics
Maggi et al. [36]	Switzerland, USA	NA	75	56 (74,7)	Retrospective	42	NA	33	MS mimics
Maggi et al. [37]	Europe	NA	39	30 (76,9)	Prospective	27	All	12	MS mimics
Sinnecker et al. [38]	Europe	18–85	606	413 (68,2)	NA	353	All (and CIS)	253	MS mimics
Al-Zandi et al. [39]	Iraq	NA	60	46 (76,7)	NA	30	NA	30	Leukoaraiosis
Cortese et al. [40]	UK	NA	34	NA	Prospective	18	RRMS	16	NMOSD
Darwish et al. [41]	Egypt	26–60	18	10 (55,6)	NA	9	NA	9	MS mimics
Hosseini et al. [42]	Canada	NA	35	26 (74,3)	Retrospective	17	RRMS	18	Healthy with WMLs
Maggi et al. [43]	Italy, Belgium	NA	83	54 (65,1)	Prospective	52	RRMS	31	Inflammatory vasculopathies
Samaraweera et al. [44]	UK	> 18	71	35 (49,3)	Retrospective	55	PPMS and RRMS	16	CSVD
Solomon et al. [45]	USA	NA	40	37 (92,5)	Prospective	20	NA	20	MS mimics
Campion et al. [46]	UK	NA	35	17 (48,6)	Retrospective	25	RRMS	10	CSVD
George et al. [47]	USA	18–70	87	45 (51,7)	Retrospective	68	NA	19	Healthy or MS mimics
Mistry et al. [48]	UK	35–76	37	20 (54,1)	Prospective	20	All	17	MS mimics
Mistry et al. [49]	UK	NA	22	NA	Prospective	13	NA	9	MS mimics

NMOSD: Neuromyelitis optica spectrum disorder, MOGAD: Myelin oligodendrocyte glycoprotein antibody-associated disease, CSVD: Cerebral small vessel disease, WML: White matter lesion, CIS: Clinically isolated syndrome, RRMS: Relapsing–remitting multiple sclerosis, PPMS: Primary progressive multiple sclerosis

Table 2 Evaluation methods of CVS in included studies

Study	MRI field strength, T	MRI sequence	Use of contrast agent	Use of NAIMS criteria	Diagnostic rules
Silvestri et al. [22]	3	T2*	Yes	Yes	Relative threshold method Select-n* method
Borrelli et al. [23]	3	T2*	Yes	Yes	Relative threshold method Select-n* method
Cagol et al. [24]	3	SWI/T2*	No	Yes	Relative threshold method Select-n method Select-n* method
Daboul et al. [25]	3	FLAIR*	Yes	Yes	Relative threshold method Select-n* method
Gao et al. [26]	3	SWI	No	Yes	Select-n* method
Okromelidze et al. [27]	3 and 7	SWI/T2*	No	No	Relative threshold method
Cortese et al. [28]	3	SWI	No	Yes	Relative threshold method
Ciotti et al. [29]	3 or 1.5	SWI	No	Yes	Relative threshold method
Ghany et al. [30]	1,5	T2*	No	Yes	Relative threshold method
Tranfa et al. [31]	3	SPGR	No	Yes	Select-n* method
Yavas et al. [32]	1.5	SWI	No	Yes	Relative threshold method
Anan et al. [33]	1.5	SWI	No	Yes	Relative threshold method
Clarke et al. [34]	3	T2*	No	Yes	Relative threshold method Select-n* method
Kaisey et al. [35]	3	T2*	Yes	Yes	Relative threshold method
Maggi et al. [36]	3	T2*	Yes	Yes	Relative threshold method Select-n method
Maggi et al. [37]	3	T2*	Yes	Yes	Relative threshold method Select-n method
Sinnecker et al. [38]	3	SWI/T2*	No	Yes	Relative threshold method Select-n* method
Al-Zandi et al. [39]	3	SWI	Yes	Yes	Relative threshold method
Cortese et al. [40]	3	SWI	No	Yes	Relative threshold method Select-n method
Darwish et al. [41]	3	SWI	No	No	Relative threshold method
Hosseini et al. [42]	7	SWI	No	Yes	Relative threshold method
Maggi et al. [43]	3 or 1.5	T2*	Yes	Yes	Relative threshold method
Samaraweera et al. [44]	3	T2*	No	Yes	Relative threshold method
Solomon et al. [45]	3	T2*	Yes	Yes	Select-n method Select-n* method
Campion et al. [46]	3	T2*	Yes	Yes	Relative threshold method Select-n* method
George et al. [47]	3	FLAIR*	Yes	No	Relative threshold method
Mistry et al. [48]	3	T2*	No	No	Relative threshold method Select-n* method
Mistry et al. [49]	7	T2*	No	No	Relative threshold method

MRI: Magnetic resonance imaging; SWI: Susceptibility-weighted imaging; FLAIR: Fluid-attenuated inversion recovery; SPGR: Spoiled gradient-echo sequence; NAIMS: North American Imaging in Multiple Sclerosis

The relative threshold method was the most used, since it was applied in 25 studies. The select-n method was used in five studies. The select-n* method was used in 11 studies. Among studies applying select-n* method, some specified that when the total number of lesions was lower than the chosen value of n, the rule was considered to be fulfilled if the majority of lesions were CVS + (see eTable 4 for further information).

Risk of bias

Risk of bias and applicability concerns were assessed for all included studies (eFigures 2–3 in Supplemental material). Patient selection and definition of the index test were the domains contributing more to risk of bias; the first one contributed also to applicability concerns in four studies [30–32, 36].

Diagnostic accuracy of relative threshold method

Sensitivities and specificities provided by the included studies for relative threshold method are shown in Fig. 1; predictive values (PV) and likelihood ratios (LR) are listed in eTable 5 of Supplemental material. As can be seen, the 40% threshold was the most used, followed by 50%.

Figure 3a shows the values of summary sensitivity (true positive rate, TPR) and 1-specificity (false positive rate,

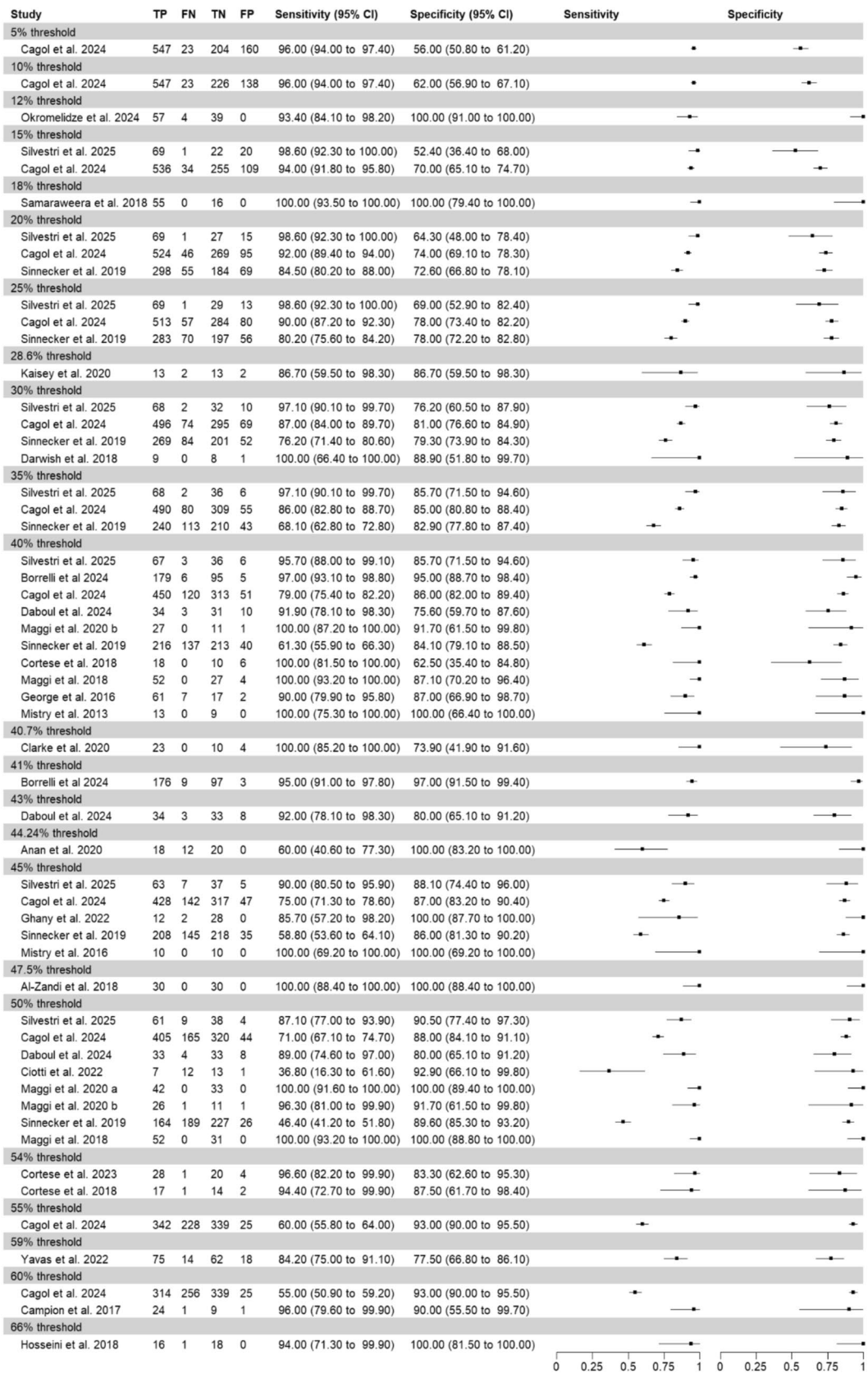

Fig. 1 Forest plot for sensitivity and specificity of relative threshold method in included studies

FPR) for different positivity thresholds as calculated with Jones' model. eTable 8 in Supplemental material lists values of sensitivity and specificity (CI) for some representative thresholds. As observed from Fig. 3b and eTable 8, our calculations showed that the threshold associated with the best Youden index is 37.5% (sensitivity 97.3%, 95%CI 90.9–99.6%; specificity 90.4%, 95%CI 83.2–95.9%; Youden index 0.877).

The HSROC model provided a DOR of 420.1 (95%CI 149.2–1586.7) for the relative threshold method (see eFigure 4 in Supplemental material for corresponding plot).

Diagnostic accuracy of select-n method

Due to high heterogeneity among selected studies, we decided it was not appropriate to perform a meta-analysis for select-n method. The parameters of diagnostic performance are reported in eTable 6 in Supplemental material.

Diagnostic accuracy of the select-n* method

Figure 2 and eTable 7 in Supplemental material show the sensitivity, specificity, PVs, and LRs of the select-n* method in the included studies. Summary of the TPR and FPR across different positivity thresholds are shown in Fig. 3c and eTable 9 in Supplemental material. As seen in Fig. 3d, select-4* was the rule with the best Youden index (sensitivity 87.1%, 95%CI 66.9–96.6%; specificity 88.2%, 95%CI 65.1–98.1%; Youden index 0.753).

Estimated DOR for the select-n* method was 57.8 (95%CI 23.4–172.1) (see eFigure 5 in Supplemental material for the HSROC plot).

Study	TP	FN	TN	FP	Sensitivity (95% CI)	Specificity (95% CI)
Select-1*						
Silvestri et al. 2025	70	0	10	32	100.00 (94.90 to 100.00)	24.20 (12.10 to 39.50)
Cagol et al. 2024	549	21	111	253	96.40 (94.40 to 97.70)	30.40 (25.80 to 35.50)
Daboul et al. 2024	33	4	14	27	89.00 (74.60 to 97.00)	34.00 (20.10 to 50.60)
Gao et al. 2024	70	3	42	45	95.90 (88.50 to 99.10)	48.30 (37.40 to 59.20)
Tranfa et al. 2022	57	16	36	0	78.10 (66.90 to 86.90)	100.00 (90.30 to 100.00)
Sinnecker et al. 2019	322	31	139	114	91.30 (87.80 to 94.00)	54.90 (48.60 to 61.20)
Select-2*						
Silvestri et al. 2025	68	2	24	18	97.10 (90.10 to 99.70)	57.60 (41.00 to 72.30)
Cagol et al. 2024	517	53	208	156	90.70 (88.00 to 93.00)	57.10 (51.90 to 62.30)
Daboul et al. 2024	32	5	21	20	86.00 (71.20 to 95.50)	51.00 (35.10 to 67.10)
Sinnecker et al. 2019	269	84	201	52	76.20 (71.40 to 80.60)	79.30 (73.90 to 84.30)
Solomon et al. 2018	18	2	13	7	90.00 (68.30 to 98.80)	65.00 (40.80 to 84.60)
Select-3*						
Silvestri et al. 2025	64	6	28	14	91.30 (82.30 to 96.80)	66.70 (50.50 to 80.40)
Borrelli et al 2024	179	6	93	7	97.00 (93.10 to 98.80)	93.00 (86.10 to 97.10)
Cagol et al. 2024	474	96	234	130	83.10 (79.80 to 86.10)	64.30 (59.10 to 69.20)
Daboul et al. 2024	30	7	28	13	81.00 (64.80 to 92.00)	68.00 (51.90 to 81.90)
Gao et al. 2024	67	6	77	10	91.80 (83.00 to 96.90)	88.50 (79.90 to 94.30)
Sinnecker et al. 2019	219	134	225	28	61.90 (56.80 to 67.10)	89.00 (84.40 to 92.50)
Solomon et al. 2018	17	3	16	4	83.00 (62.10 to 96.80)	81.00 (56.30 to 94.30)
Select-4*						
Silvestri et al. 2025	64	6	28	14	91.30 (82.30 to 96.80)	66.70 (50.50 to 80.40)
Cagol et al. 2024	434	136	260	104	76.20 (72.40 to 79.60)	71.40 (66.50 to 76.00)
Daboul et al. 2024	27	10	31	10	73.00 (55.90 to 86.20)	76.00 (59.70 to 87.60)
Sinnecker et al. 2019	177	176	239	14	50.20 (44.80 to 55.50)	94.50 (90.90 to 96.90)
Select-5*						
Silvestri et al. 2025	62	8	33	9	88.40 (78.70 to 94.90)	78.80 (63.20 to 89.70)
Cagol et al. 2024	396	174	312	52	69.40 (65.50 to 73.20)	85.70 (81.70 to 89.10)
Daboul et al. 2024	25	12	35	6	68.00 (50.20 to 82.00)	85.00 (70.80 to 94.40)
Sinnecker et al. 2019	144	209	248	5	40.90 (35.60 to 46.10)	98.20 (95.40 to 99.40)
Select-6*						
Borrelli et al 2024	165	20	95	5	89.00 (83.80 to 93.30)	95.00 (88.70 to 98.40)
Cagol et al. 2024	356	214	338	26	62.50 (58.30 to 66.40)	92.90 (89.70 to 95.30)
Daboul et al. 2024	24	13	40	1	65.00 (47.50 to 79.80)	98.00 (87.10 to 99.90)
Clarke et al. 2020	18	5	7	7	79.30 (56.30 to 92.50)	52.20 (23.00 to 77.00)
Campion et al. 2017	25	0	7	3	100.00 (86.30 to 100.00)	70.00 (34.80 to 93.30)
Mistry et al. 2016	13	0	7	0	100.00 (75.30 to 100.00)	100.00 (59.00 to 100.00)
Select-8*						
Silvestri et al. 2025	62	8	33	9	87.90 (78.70 to 94.90)	79.70 (63.20 to 89.70)
Select-10*						
Silvestri et al. 2025	52	18	39	3	73.90 (62.40 to 84.00)	93.90 (80.50 to 98.50)
Select-15*						
Silvestri et al. 2025	44	26	41	1	62.30 (50.50 to 74.10)	97.00 (87.40 to 99.90)

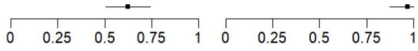

Fig. 2 Forest plot for sensitivity and specificity of select-n* method in included studies

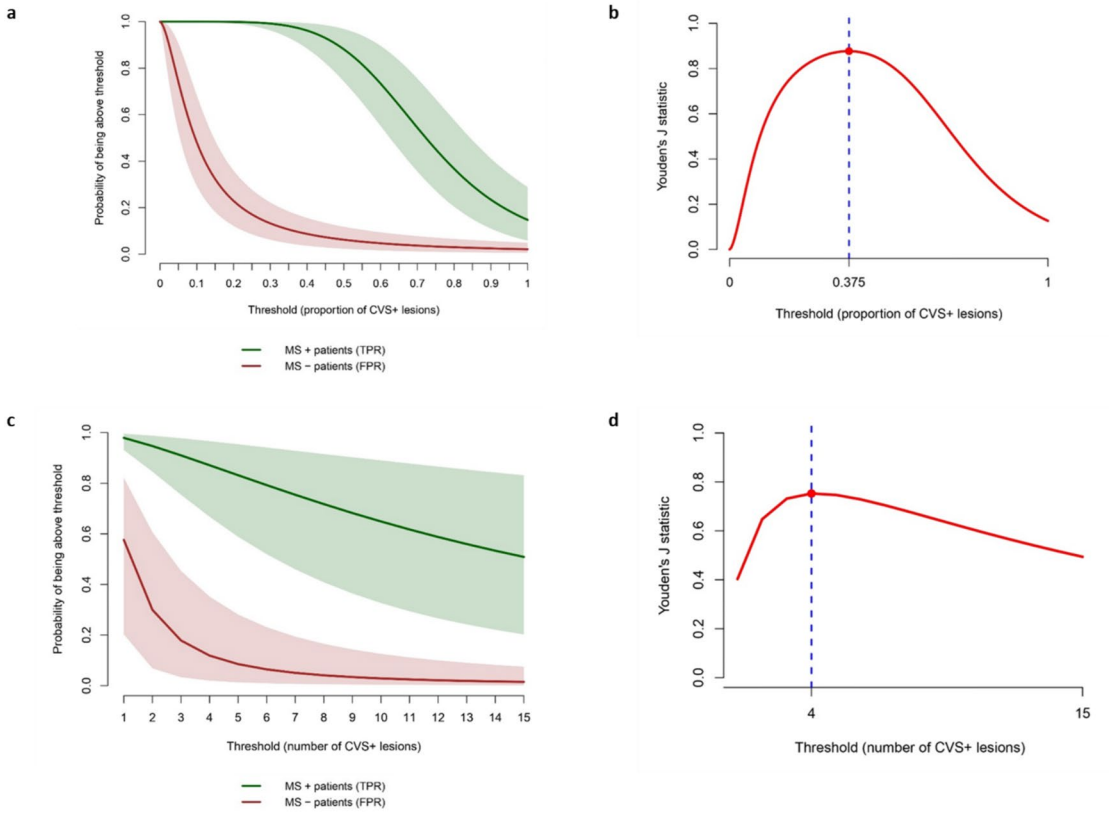

Fig. 3 Summary parameters of diagnostic accuracy for CVS-based methods. **a** and **c** are plots of true positive rate (TPR) and false positive rate (FPR) across different values of positivity threshold for relative threshold and select-n* method, respectively. The green line represents TPR, i.e., the probability of being above threshold for MS + subjects; the red line represents FPR, i.e., the probability of being above threshold for MS − subjects. **b** and **d** are plots of Youden index across different values of positivity threshold for relative threshold and select-n* method, respectively

Investigation of heterogeneity

For the relative threshold method, the adoption of gadolinium-based protocols of MRI acquisition did not show a relevant impact on DOR (RDOR = 6.62, 95%CI 0.68–71.27). In the same way, no significant difference was found for the select-n* method (RDOR = 2.52, 95%CI 0.35–15.36). See eFigures 6–7 in Supplemental material for the HSROC plots depicting these results.

Discussion

In this systematic review and Bayesian meta-analysis, which included 28 studies and 3434 participants, we collected the most robust and up-to-date evidence regarding the accuracy of CVS for MS diagnosis. Before our work, three systematic reviews with meta-analyses were published on this topic [50–52], the most recent one dating back to 2022. One more was published in 2024 [53], but it focused on the role of SWI sequences, including imaging markers other than CVS and excluding CVS detection with other GRE sequences. Other recent reviews provided no summary estimates of diagnostic accuracy of CVS [54, 55].

Inferences from our review point out some key findings. First, it can be observed that high heterogeneity exists in diagnostic rules based on CVS detection, not only in terms of their design, but also in their naming. This lack of shared rules and nomenclature represented a challenge to get pooled evidence and required adequate name standardization prior to further analyses.

Our results confirm the potential of CVS in MS diagnosis thanks to its high specificity and sensitivity, both with relative threshold and select-n* methods. With reference to the first method, we found an optimal cutoff at 37.5%; this value is quite similar to the 40% threshold, which is the most used threshold in included studies and the optimal one according to a past review on the topic [50]. Values of sensitivity and specificity estimated by this study for the 40% cutoff (90% and 89%, respectively) are comparable to those we obtained for the same threshold; in our results, superiority of the 37.5% value over the 40% one depends

on higher sensitivity. It should be noted that the other two reviews on CVS [51, 52] proposed a higher cutoff at 45%; however, they were based on a much lower number of included studies (6 and 7 vs 15 in the other cited review and 25 in ours). If our findings will be taken into account in future clinical practice, the small difference between 37.5 and 40% thresholds can be probably neglected and the latter can be regarded as a good compromise between diagnostic accuracy and ease of use.

It can be observed that for relative threshold method, sensitivity overcomes specificity at optimal cutoff; thus, adoption of 37.5% threshold could lead to a significant risk of false positives, as already highlighted by past reviews on the topic [52]. However, the use of CVS detection should not be seen as a stand-alone diagnostic method, but as a part of a system of criteria including also other imaging, laboratory, and clinical data; this approach should exclude at least some of the potential false positives due to their lack of suggestive characteristics from these other points of view.

Considering the select-n* method, the optimal threshold we found is 4, which is lower than the select-6* rule that has been proposed to be adopted in the revised McDonald criteria [7]. These results should be interpreted with caution, due to the limited number of included studies and heterogeneity in handling cases where the total number of lesions was lower than n (see above and eTable 4). However, our findings could strengthen the need for a reflection over this threshold, which has rarely been explored directly by existing studies and probably for this reason has been neglected until now.

Moreover, the fact that DOR is higher for relative threshold than for select-n* criterion, as also sensitivity and specificity at optimal cutoffs are, could raise some issues about the choice of the second method rather than the first one. Nevertheless, we are aware that practical reasons limit the use of the relative threshold rule (to get the percentage of CVS + , all lesions should be evaluated, whereas for select-n* method only n lesions have to be considered).

With regard to results of meta-regression, the absence of significant results for the impact of gadolinium use over diagnostic performance of CVS-based criteria is relevant, since the need for contrast agent administration could represent a barrier to their wide adoption. However, these results should be interpreted with caution, since our analysis, due to numerosity issues, did not stratify available data for the specific sequence used; since the administration of gadolinium, at least at a conceptual level, can improve the visualization of vessels, we cannot exclude that future studies performed on a single sequence, for example SWI, will have different results.

Some open questions are left by our review. The first one concerns real-world validation of the diagnostic accuracy of CVS, especially in the context of the new MS diagnostic criteria: large-scale adoption of CVS-based diagnostic rules will face new challenges, in terms of both reader expertise outside specialized centers and between-center variability of MRI protocols that could negatively affect their diagnostic performance. Another open issue is the role of automated assessment of CVS, which can become more and more relevant as methods based on artificial intelligence will be applied to MRI data interpretation and could revamp the role of relative threshold criterion, whose main limit has been its time-consuming nature.

Moreover, even if the accuracy of CVS in differential diagnosis between MS and other disorders is clear, we provided an overall estimate of diagnostic performance, since in most included studies the control group was composed of participants affected by diverse disorders. New studies will be needed in the future to explore the potential variability of CVS accuracy in distinguishing MS from specific diseases (for example, NMOSD, CSVD, or migraine).

The role of CVS + lesions in the spinal cord will also deserve future attention; even if the presence of perivenular plaques has been reported on pathological specimens, to our knowledge, only one study has addressed this theme until now, probably because the need of 3 T scans to visualize veins in this area [56].

Strengths and limitations

Unlike previous review on the same topic, we adopted statistical methods taking into account studies supplying accuracy estimates at multiple thresholds and thus we were also able to find the best cutoff for each rule in a reliable way. Moreover, to our knowledge, this review is the first to perform meta-analysis not only for relative threshold criterion, but also for select-n* rule, which has been recognized as the most feasible in clinical practice. In addition, our effort to categorize heterogeneous diagnostic rules in three broad groups can prompt standardization of future primary research and constitute a benchmark for new reviews on the same topic.

However, our review is not free from limits. Heterogeneity of included studies is one of them and involves both protocols of MRI acquisition and controls selection, with different MS mimics considered in each case. We tried to address some of these issues with use of meta-regression, but probably additional subgroup analyses could have been useful, with particular reference to stratification for specific MRI sequences used in included studies, since it cannot be excluded that different MRI protocols might influence CVS detection rate. For a wider discussion on these technical issues, which are partially outside the scope of this review, see Chaaban et al. [54]. Low number of studies presenting results for select-n* method is another issue, at least if compared with those regarding relative threshold method; this posed a challenge to the convergence of Bayesian models,

solved by increasing the number of iterations of Markovian chains, and limited the accuracy of our estimates, with broad credible intervals for TPR and FPR.

Another limitation of the study is the language restriction to English-written articles; however, the low number of studies excluded for this reason (one during full text screening and another during previous phases; see also Supplementary material) should limit the concerns of language bias. It should be also noticed that a more thorough assessment of inter-rater agreement (which was not quantitatively evaluated) during study selection, data extraction, and risk-of-bias evaluation would have improved the methodology of our review; however, disagreements were managed with adequate discussion among authors, thus limiting their impact.

Lack of information for automated assessments of CVS + lesions in another limit of our review. Finally, due to limited number of studies, we could not perform meta-regression regarding the role of MRI field strength and this could be a problem, since it is clear that the power of field strength changes the accuracy of CVS detection [54].

Conclusions

The results of this systematic review and meta-analysis show that CVS-based diagnostic rules are reliable in MS diagnosis, with an optimal cutoff of 37.5% for the relative threshold method and of four for the select-n* method. However, further research will be needed in the future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00415-026-13622-6.

Author contributions Dr Baiamonte and all of the authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: Baiamonte; acquisition, analysis, or interpretation of data: all authors; drafting of the manuscript: Baiamonte; critical review of the manuscript for important intellectual content: Sera, Ragonese, Salemi; statistical analysis: Sera, Baiamonte; supervision: Ragonese, Salemi.

Funding None.

Data availability Complete data of the analyses performed in the study are available upon reasonable request.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethical approval The manuscript does not contain newly published clinical studies or patient data; all included studies were approved by the local ethics committee.

References

1. Thompson AJ, Banwell BL, Barkhof F et al (2018) Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 17:162–173. https://doi.org/10.1016/S1474-4422(17)30470-2
2. Levraut M, Landes-Chateau C, Mondot L et al (2025) The Kappa free light chains index and central vein sign: two new biomarkers for multiple sclerosis diagnosis. Neurol Ther 14:711–731. https://doi.org/10.1007/s40120-025-00737-7
3. Rimkus CDM, Otsuka FS, Nunes DM et al (2024) Central vein sign and paramagnetic rim lesions: susceptibility changes in brain tissues and their implications for the study of multiple sclerosis pathology. Diagnostics 14:1362. https://doi.org/10.3390/diagnostics14131362
4. Sati P, Oh J, NAIMS Cooperative et al (2016) The central vein sign and its clinical evaluation for the diagnosis of multiple sclerosis: a consensus statement from the North American imaging in multiple sclerosis cooperative. Nat Rev Neurol 12:714–722. https://doi.org/10.1038/nrneurol.2016.166
5. Machado-Santos J, Saji E, Tröscher AR et al (2018) The compartmentalized inflammatory response in the multiple sclerosis brain is composed of tissue-resident CD8+ T lymphocytes and B cells. Brain 141:2066–2082. https://doi.org/10.1093/brain/awy151
6. Rae-Grant AD, Wong C, Bernatowicz R, Fox RJ (2014) Observations on the brain vasculature in multiple sclerosis: a historical perspective. Mult Scler Relat Disord 3:156–162. https://doi.org/10.1016/j.msard.2013.08.005
7. Montalban X, Lebrun-Frénay C, Oh J et al (2025) Diagnosis of multiple sclerosis: 2024 revisions of the McDonald criteria. Lancet Neurol 24:850–865. https://doi.org/10.1016/S1474-4422(25)00270-4
8. Deeks J, Bossuyt P, Leeflang M, Takwoingi Y (2023) Cochrane handbook for systematic reviews of diagnostic test accuracy, Version 2.0 (updated July 2023). Cochrane
9. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA (2022) PRISMA2020: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev 18:e1230. https://doi.org/10.1002/cl2.1230
10. McInnes MDF, Moher D, Thombs BD et al (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 319:388–396. https://doi.org/10.1001/jama.2017.19163
11. Clark JM, Sanders S, Carter M et al (2020) Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. J Med Libr Assoc. https://doi.org/10.5195/jmla.2020.834
12. Clark J, Glasziou P, Del Mar C et al (2020) A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol 121:81–90. https://doi.org/10.1016/j.jclinepi.2020.01.008
13. Forbes C, Greenwood H, Carter M, Clark J (2024) Automation of duplicate record detection for systematic reviews: deduplicator. Syst Rev 13:206. https://doi.org/10.1186/s13643-024-02619-9
14. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. Syst Rev 5:210. https://doi.org/10.1186/s13643-016-0384-4
15. (2025) PlotDigitizer
16. Whiting PF, Rutjes AWS, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
17. Jones HE, Gatsonsis CA, Trikalinos TA et al (2019) Quantifying how diagnostic test accuracy depends on threshold in a

meta-analysis. Stat Med 38:4789–4803. https://doi.org/10.1002/sim.8301
18. Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 20:2865–2884. https://doi.org/10.1002/sim.942
19. RStudio Team (2019) RStudio: Integrated Development for R
20. R Core Team R: A language and environment for statistical computing
21. Plummer M JAGS: Just another gibbs sampler
22. Silvestri G, Roccatagliata L, Tazza F et al (2025) The "central vein sign" to differentiate multiple sclerosis from migraine. Headache. https://doi.org/10.1111/head.14902
23. Borrelli S, Martire MS, Stölting A et al (2024) Central vein sign, cortical lesions, and paramagnetic rim lesions for the diagnostic and prognostic workup of multiple sclerosis. Neurol Neuroimmunol Neuroinflamm 11:e200253. https://doi.org/10.1212/NXI.0000000000200253
24. Cagol A, Cortese R, Barakovic M et al (2024) Diagnostic performance of cortical lesions and the central vein sign in multiple sclerosis. JAMA Neurol 81:143–153. https://doi.org/10.1001/jamaneurol.2023.4737
25. Daboul L, O'Donnell CM, Amin M et al (2024) A multicenter pilot study evaluating simplified central vein assessment for the diagnosis of multiple sclerosis. Mult Scler J 30:25–34. https://doi.org/10.1177/13524585231214360
26. Gao C, Su L, Li H et al (2024) Susceptibility-weighted image features in AQP4-negative-NMOSD versus MS. Mult Scler Relat Disord 82:105406. https://doi.org/10.1016/j.msard.2023.105406
27. Okromelidze L, Patel V, Singh RB et al (2023) Central vein sign in multiple sclerosis: a comparison study of the diagnostic performance of 3T versus 7T MRI. AJNR Am J Neuroradiol 45:76–81. https://doi.org/10.3174/ajnr.A8083
28. Cortese R, Prados Carrasco F, Tur C et al (2023) Differentiating multiple sclerosis From AQP4-neuromyelitis optica spectrum disorder and mog-antibody disease with imaging. Neurology. https://doi.org/10.1212/WNL.0000000000201465
29. Ciotti JR, Eby NS, Brier MR et al (2022) Central vein sign and other radiographic features distinguishing myelin oligodendrocyte glycoprotein antibody disease from multiple sclerosis and aquaporin-4 antibody-positive neuromyelitis optica. Mult Scler J 28:49–60. https://doi.org/10.1177/13524585211007086
30. Abdel Ghany H, Karam-Allah A, Edward R et al (2022) Sensitivity and specificity of central vein sign as a diagnostic biomarker in Egyptian patients with multiple sclerosis. Neuropsychiatr Dis Treat 18:1985–1992. https://doi.org/10.2147/NDT.S377877
31. Tranfa M, Tortora M, Pontillo G et al (2022) The central vein sign helps in differentiating multiple sclerosis from its mimickers: lessons from Fabry disease. Eur Radiol 32:3846–3854. https://doi.org/10.1007/s00330-021-08487-4
32. Yavaş HG, Sağtaş E (2022) Central vein sign: comparison of multiple sclerosis and leukoaraiosis. Turk J Med Sci 52:1933–1942. https://doi.org/10.55730/1300-0144.5541
33. Anan RA, El-Adalany MA, Belal TMI, Elmogy SAE-D (2020) Detection of central veins using susceptibility weighted imaging (SWI) for discrimination between lesions in multiple sclerosis (MS) and cerebral small vessel disease (CSVD). Egypt J Radiol Nucl Med 51:63. https://doi.org/10.1186/s43055-020-00185-3
34. Clarke MA, Samaraweera AP, Falah Y et al (2020) Single test to arrive at multiple sclerosis (STAR-MS) diagnosis: a prospective pilot study assessing the accuracy of the central vein sign in predicting multiple sclerosis in cases of diagnostic uncertainty. Mult Scler J 26:433–441. https://doi.org/10.1177/1352458519882282
35. Kaisey M, Solomon AJ, Guerrero BL et al (2021) Preventing multiple sclerosis misdiagnosis using the "central vein sign": a real-world study. Mult Scler Relat Disord 48:102671. https://doi.org/10.1016/j.msard.2020.102671
36. Maggi P, Fartaria MJ, Jorge J et al (2020) CVSnet: a machine learning approach for automated central vein sign assessment in multiple sclerosis. NMR Biomed 33:e4283. https://doi.org/10.1002/nbm.4283
37. Maggi P, Absinta M, Sati P et al (2020) The "central vein sign" in patients with diagnostic "red flags" for multiple sclerosis: a prospective multicenter 3T study. Mult Scler J 26:421–432. https://doi.org/10.1177/1352458519876031
38. Sinnecker T, Clarke MA, Meier D et al (2019) Evaluation of the central vein sign as a diagnostic imaging biomarker in multiple sclerosis. JAMA Neurol 76:1446. https://doi.org/10.1001/jamaneurol.2019.2478
39. Al-Zandi SH, Fayadh NAH, Al-Waely NKN (2018) Central vein sign detected by SWI at 3 T MRI as a discriminator between multiple sclerosis and leukoaraiosis. Egypt J Radiol Nucl Med 49:158–164. https://doi.org/10.1016/j.ejrnm.2017.09.003
40. Cortese R, Magnollay L, Tur C et al (2018) Value of the central vein sign at 3T to differentiate MS from seropositive NMOSD. Neurology. https://doi.org/10.1212/WNL.0000000000005256
41. Darwish EAF, Ibrahim YA, Zamzam DA (2018) Value of central vein sign in discriminating multiple sclerosis plaques from other white matter lesions. Egypt J Radiol Nucl Med 49:165–171. https://doi.org/10.1016/j.ejrnm.2017.12.008
42. Hosseini Z, Matusinec J, Rudko DA et al (2018) Morphology-specific discrimination between MS white matter lesions and benign white matter hyperintensities using ultra-high-field MRI. Am J Neuroradiol. https://doi.org/10.3174/ajnr.A5705
43. Maggi P, Absinta M, Grammatico M et al (2018) Central vein sign differentiates multiple sclerosis from central nervous system inflammatory vasculopathies. Ann Neurol 83:283–294. https://doi.org/10.1002/ana.25146
44. Samaraweera APR, Falah Y, Pitiot A et al (2018) The MRI central vein marker; differentiating PPMS from RRMS and ischemic SVD. Neurol Neuroimmunol Neuroinflamm 5:e496. https://doi.org/10.1212/NXI.0000000000000496
45. Solomon AJ, Watts R, Ontaneda D et al (2018) Diagnostic performance of central vein sign for multiple sclerosis with a simplified three-lesion algorithm. Mult Scler 24:750–757. https://doi.org/10.1177/1352458517726383
46. Campion T, Smith RJP, Altmann DR et al (2017) FLAIR* to visualize veins in white matter lesions: a new tool for the diagnosis of multiple sclerosis? Eur Radiol 27:4257–4263. https://doi.org/10.1007/s00330-017-4822-z
47. George IC, Sati P, Absinta M et al (2016) Clinical 3-tesla FLAIR* MRI improves diagnostic accuracy in multiple sclerosis. Mult Scler J 22:1578–1586. https://doi.org/10.1177/1352458515624975
48. Mistry N, Abdel-Fahim R, Samaraweera A et al (2016) Imaging central veins in brain lesions with 3-T T2*-weighted magnetic resonance imaging differentiates multiple sclerosis from microangiopathic brain lesions. Mult Scler J 22:1289–1296. https://doi.org/10.1177/1352458515616700
49. Mistry N, Dixon J, Tallantyre E et al (2013) Central veins in brain lesions visualized with high-field magnetic resonance imaging: a pathologically specific diagnostic biomarker for inflammatory demyelination in the brain. JAMA Neurol. https://doi.org/10.1001/jamaneurol.2013.1405
50. Castellaro M, Tamanti A, Pisani AI et al (2020) The use of the central vein sign in the diagnosis of multiple sclerosis: a systematic review and meta-analysis. Diagnostics 10:1025. https://doi.org/10.3390/diagnostics10121025
51. Bhandari A, Xiang H, Lechner-Scott J, Agzarian M (2020) Central vein sign for multiple sclerosis: a systematic review and meta-analysis. Clin Radiol 75:479.e9-479.e15. https://doi.org/10.1016/j.crad.2020.01.011

52. Suh CH, Kim SJ, Jung SC et al (2019) The "Central Vein Sign" on T2*-weighted images as a diagnostic tool in multiple sclerosis: a systematic review and meta-analysis using individual patient data. Sci Rep 9:18188. https://doi.org/10.1038/s41598-019-54583-3
53. Afkandeh R, Abedi I, Zamanian M (2024) Detection of multiple sclerosis lesions by susceptibility-weighted imaging-a systematic review and meta-analyses. Clin Radiol 79:e1522–e1529. https://doi.org/10.1016/j.crad.2024.09.009
54. Chaaban L, Safwan N, Moussa H et al (2022) Central vein sign: a putative diagnostic marker for multiple sclerosis. Acta Neurol Scand 145:279–287. https://doi.org/10.1111/ane.13553
55. Kerpel A, Tal S, Menascu S et al (2025) Acquisition of susceptibility-sensitive sequences for central vein sign detection: a systematic review of MRI protocols. Mult Scler Relat Disord. https://doi.org/10.1016/j.msard.2025.106608
56. Jensen-Kondering U, Larsen N, Huhndorf M et al (2022) Central vein sign in patients with inflammatory lesion of the upper cervical spinal cord on susceptibility weighted imaging at 3 tesla. Preliminary results. Magn Reson Imaging 93:11–14. https://doi.org/10.1016/j.mri.2022.07.013