

BMJ Open A reliability study of colour-Doppler sonography for the diagnosis of chronic cerebrospinal venous insufficiency shows low inter-rater agreement

Maurizio A Leone,^{1,2} Olga Raymkulova,¹ Ausiliatrice Lucenti,¹ Alessandro Stecco,³ Laura Bolamperti,¹ Lorenzo Coppo,¹ William Liboni,⁴ Gianandrea Rivadossi,⁴ Giuseppe Zaccala,⁵ Maurizio Maggio,⁶ Fabio Melis,⁷ Claudia Giaccone,⁷ Alessandro Carriero,³ Piergiorgio Lochner⁸

To cite: Leone MA, Raymkulova O, Lucenti A, *et al*. A reliability study of colour-Doppler sonography for the diagnosis of chronic cerebrospinal venous insufficiency shows low inter-rater agreement. *BMJ Open* 2013;**3**:e003508. doi:10.1136/bmjopen-2013-003508

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-003508>).

Received 30 June 2013
Revised 17 September 2013
Accepted 4 October 2013



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Maurizio A Leone;
maurizio.leone@
maggioreosp.novara.it

ABSTRACT

Objective: Chronic cerebrospinal venous insufficiency (CCSVI) has been extremely variable, associated with multiple sclerosis in colour-Doppler sonographic studies. We aimed to evaluate inter-rater agreement in a colour-Doppler sonography venous examination.

Design: Inter-rater agreement study.

Setting: First-referral multiple sclerosis centre.

Participants: 38 patients with multiple sclerosis and 55 age-matched (± 5 years) controls.

Intervention: Sonography was carried out in accordance with Zamboni's five criteria by eight sonographers with different expertise, blinded to the status of cases and controls. Each participant was evaluated by two operators.

Primary and secondary outcome measures:

Inter-rater agreement was measured through the κ statistics and the intraclass correlation coefficient.

Results: The agreement was no higher than chance for criterion 2—reflux in the deep cerebral veins ($\kappa = -0.02$) and criterion 4—flow not Doppler detectable in one or both the internal jugular veins (IJVs) or vertebral veins (VVs; -0.09). It was substantially low for criterion 1—reflux in the IJVs and/or VVs (0.29), criterion 3—IJV stenosis or malformations (0.23) and criterion 5—absence of IJV diameter increase when passing from the sitting to the supine position (0.22). The κ value for CCSVI as a whole was 0.20 (95% confidence limit -0.01 to 0.42). Intraclass correlation coefficients for the measure of cross-sectional area ranged from 0.05 to 0.25. Inter-rater agreement was low for CCSVI experts ($\kappa = 0.24$; -0.11 to 0.59) and non-experts (0.20; -0.33 to 0.73); neurologists (0.21; -0.06 to 0.47) and non-neurologists (0.18; -0.20 to 0.56); cases (0.19; -0.14 to 0.52) and controls (0.21; -0.08 to 0.49). Zamboni-trained neurosonographers ascertained CCSVI more frequently than the non-trained neurosonographers.

Conclusions: Agreement was unsatisfactory for the diagnosis of CCSVI as a whole, for each of its five criteria and according to the different subgroups. Standardisation of the method is urgently needed prior

Strengths and limitations of this study

- Its blinded nature, the wide spectrum of examined participants, including patients with multiple sclerosis (MS) with different severity, healthy and other disease controls and different ages, and the control of potential sources of variability due to the equipment (we used only one colour-Doppler sonography (CDS) instrument), and the observed (we minimised any possible variability between the first and the second examination). Furthermore, the study evaluated several possible determinants of inter-rater agreement, including specialty and clinical experience of the operators, MS course and severity, and training practice.
- We were not able to randomise the operator sequence for practical reasons.
- Despite several precautions, blinding was incomplete since our CDS operators were able to identify controls in a percentage higher than chance.

to its further application in studies of patients with multiple sclerosis or other neurological diseases.

INTRODUCTION

The new entity of 'chronic cerebrospinal venous insufficiency (CCSVI)' has been described as a clinical syndrome comprising stenoses of the jugular and/or azygos veins, characterised by collateral venous outflows and reduced cerebral blood flow.¹ It has been associated with multiple sclerosis (MS) in a study using an extracranial and transcranial colour-Doppler sonographic examination,² whereas subsequent studies have produced conflicting results and even questioned its existence.^{3 4} Leaving aside the original



report, the prevalence of CCSVI detected with colour-Doppler sonography (CDS) in case-control studies ranged from 0% to 84% in patients with MS and from 0% to 36% in controls.^{3 5–14} Non-controlled studies found even higher prevalences; in one multicentre study,¹⁵ CCSVI prevalence ranged from 74% to 96% in six MS centres. This extreme heterogeneity is still devoid of a clear explanation. It may depend on any of the three determinants of the variability of a diagnostic procedure: the participant observed, the measuring instrument and procedure, and the observer. The possible sources of observed (patients) variability include differences in age, clinical forms and disability or several physiological factors, such as head position, hydration status, use of arm abduction to relax cervical musculature and respiration. Employment of different equipments and technical aspects of the procedure of CDS examination may also determine heterogeneity. However, since CDS is a highly observer-dependent examination, we postulated that interoperator variability is a major source of heterogeneity. If the reproducibility of CDS examination is not known, one cannot rely on the results of any study of CCSVI. For this reason, we set out to evaluate the inter-rater agreement of CDS venous examination. Our secondary aims were to evaluate such agreement in subgroups of cases and controls, different types of MS and in the function of some of an operator's characteristics.

METHODS

Informed consent was signed by all participants.

Patients

We prospectively enrolled all consecutive patients (age >18) presenting to our MS centre (a first referral MS centre) from March to September 2011 (N=243). They were asked to participate in the study, irrespective of the severity, duration or treatment of their disease. Those who accepted (N=185) were listed and subsequently summoned for the ultrasound examination. Exclusion criteria are listed elsewhere.¹² Thirty patients with MS were examined during the run-in period before the study onset, 9 had one of the exclusion criteria, 18 had already been examined at another centre, 12 refused and for 78 the repeat CDS examination was unpractical, leaving 38 patients available for the inter-rater study. Comparison of the 38 cases with the 147 non-included did not disclose any statistically significant difference for age, gender, place of birth, type of MS, treatment, age at onset and disease duration (data not shown). All patients were visited and diagnoses were confirmed according to the revised McDonald criteria.¹⁶ Patients with a relapse or those using steroids during the previous 30 days were excluded. Demographic and clinical information included age, gender, age at onset, Expanded Disability Status Scale (EDSS),¹⁷ disease duration, clinical course¹⁸ and other parameters not evaluated in this study.¹²

Controls

Forty-seven healthy controls, matched to cases by age (± 5 years), were selected from students, university personnel, relatives of patients admitted to the hospital for diseases other than MS, and their friends. They were examined to rule out MS or other neurological diseases. None had a relative suffering from MS. Representativeness was extended by enrolling a further eight controls with neurodegenerative diseases (5 amyotrophic lateral sclerosis and 3 spastic paraparesis). The exclusion criteria for the cases were applied to the controls.

Ultrasound examinations

The CDS study was performed with a GE Vivid 7 scanner with a 7.5–10 MHz high-resolution linear array transducer for extracranial measurements and a 2–3 MHz probe for transcranial evaluation of venous drainage (GE Healthcare, Milwaukee, Wisconsin, USA). CDS was conducted initially in the supine position and afterwards in the upright position, allowing for 2–4 min before any measurement. The head was kept in line with the neck and in slight hyperextension. Sonographers examined the flow characteristics of the internal jugular vein (IJV) and the vertebral vein (VV) on the right side before, and the direction of flow in the deep cerebral veins (DCVs). The system settings were adjusted for the analysis of low-velocity signals, and the pulse repetition frequency was thus reduced to facilitate venous vessel detection. The participant's head was held straight with appropriate head and arm supports to avoid venous compression. A large amount of ultrasonic gel was used and special care was taken to avoid compressing the neck. The CDS investigation was carried out in accordance with the five criteria suggested by Zamboni *et al.*²

1. Reflux in the IJVs and/or VVs in the supine and sitting positions. Reflux in any vein >0.88 s was considered 'pathological'. Flow was assessed during a period of apnoea following normal exhalation and not during Valsalva manoeuvre. The probe was located in a longitudinal and axial plane at the thyroid gland (J2 point), which was maintained when participants changed to the upright position.
2. Reflux in the DCVs. Flow characteristics in at least one DCV were measured; flow in a reverse direction >0.5 s was considered 'pathological'.
3. High-resolution B-mode evidence of IJV stenosis or malformations (septum, valve malformation, flap, membrane and annulus). Stenosis was defined as a cross-sectional area (CSA) <0.3 cm², measured at the thyroid gland (J2).
4. Flow not Doppler detectable in one or both the IJVs or VVs following deep inspiration in the supine and upright positions.
5. Absence of physiological diameter increase of the IJV when passing from the sitting to the supine position.

CCSVI status assessment

A participant was considered CCSVI positive if ≥ 2 criteria were met, according to the original study.²

Participants not assessed for criterion 2 due to the lack of expertise of some sonographers were assumed to have failed to fulfil this criterion. Participants assessed for only four criteria were thus placed in three groups¹⁹: no CCSVI (4 negative criteria), CCSVI (at least 2 positive of the other 4 criteria) and borderline CCSVI (1 positive of the other 4 criteria).

Procedure of the study

Eight CDS operators (6 neurologists, 1 internist and 1 radiologist) participated in the study. Four worked in our hospital (2 in the neurology department, 1 in the radiology department and 1 in the internal medicine departments), and four in other hospitals (all in the neurological departments). None of them work in a MS centre. We classified the clinical experience of the operators according to the number of ultrasound examinations performed in the last 15 years (≤ 1500 / >1500) and their specific CCSVI expertise (training at Zamboni's laboratory at the University of Ferrara, or other laboratories that use Zamboni's technique vs no specific training).

Much effort was directed to ensure blinding. Our MS centre is situated in a building apart from the CDS laboratory. Operators were blinded to the status of cases and controls. One 'outsider' (OR) was in charge of the whole procedure. She transferred participants to the laboratory, measured the blood pressure and heart rate, comfortably positioning them on a tilt chair and covering them with a blanket to conceal any hints such as injection marks potentially allowing for group assignment, and moved any aids out of the room. She alone was allowed to speak to the participants. Only at this point was the first operator allowed to enter the laboratory room. Participants were instructed not to talk to the operators, and operators were not allowed to talk to them.

At the end of the first examination, after the operator left the room, the participant was free to move and rest for 10 min. Thereafter, he/she was repositioned, and the second operator was allowed to enter the lab room. Before starting the second examination, the outsider remeasured the blood pressure and heart rate. The operators filled in the study forms immediately after each examination, first indicating their guess on the status of each participant (case or control) and later the CDS features. Every examination session included 3–6 patients examined by a single pair of CDS operators. The order in which the operators of each pair performed the examination was not randomised but was based on their availability. The number of examinations performed by each operator varied from 6 to 23.

Statistical analysis

Agreement was evaluated through the percentage agreement and κ statistics for categorical variables and the intraclass correlation coefficient (ICC)²⁰ for continuous variables. Comparisons between groups were assessed

using parametric (χ^2 tests and Student t test) and non-parametric methods (Mann-Whitney U tests) where appropriate (deviation from normal distribution according to the Shapiro-Wilk test). Bland and Altman plots²¹ were used to judge agreement across the range of a value. Blindness was evaluated with the Bang Index.²² Data were analysed with SAS²³ and R.²⁴

RESULTS

We surveyed 93 participants (60 women and 33 men). There were 38 patients with MS (23 women and 15 men) with a mean age of 44.9 years (SD=10.1, range 18–62) and 55 controls (37 women and 18 men) with a mean age of 40.9 (SD=12.7, range 24–74; p =ns). The mean age of onset was 31.8 years (10.3; 10–59) for the 27 patients with relapsing-remitting (RR) MS and 31.9 years (11.6; 14–52) for the 11 patients with primary and secondary progressive MS (p =ns). The mean disease duration was 11.4 years (7.2; 2–26) for patients with RR-MS and 17.3 years (10.6; 5–41) for patients with progressive MS (p =ns). Median EDSS was 2 (IQ: 1–2.5; range 1–6) for RR-MS and 6.5 (IQ: 5–7; range 2.5–8) for progressive MS ($p<0.0001$). The mean blood pressure was 127 mm Hg (SD=14.1)/78 (9.6) before starting the first examination and 125 (13.6)/78 (9.5) before starting the second examination. Heart rates were 78 (9.7) and 76 (9.4) bpm, respectively. None of the differences were statistically significant.

The agreement was no higher than chance for criteria 2 and 4, and was substantially low for the other three criteria (table 1). Table 2 shows the ICCs for the CSA as measured by the two raters. ICCs ranged from 0.05 to 0.25; they were statistically significant only in the sitting position at the left side. The δ CSA (criterion 5) was calculated by subtracting the CSA measured in the supine position from that in the sitting position; ICC was 0.19 (95% confidence limit -0.01 to 0.38 ; $p=0.04$) at the right side and 0.13 (-0.08 to 0.33 ; $p=0.11$) at the left side. The difference between the measurements of the CSAs by the two raters against the mean of the measurements (Bland and Altman plot) is shown in figure 1. The possibility of a proportional bias, that is, a different agreement between the two observers depending on the mean of the measure, was explored using a least square regression line fitted to the plot. With the exception of the upright left position (slope=0.01), all the other slopes were different from zero, indicating a possible proportional bias: supine left -0.27 , supine right -0.43 and upright right -0.58 .

The first operator found 30 participants positive for CCSVI, 59 negative and 4 borderline. The second operator found 26 participants positive for CCSVI, 62 negative and 5 borderline. After excluding the borderline participants, the percentage of agreement between the two operators was 66%, κ 0.20 (-0.01 to 0.42 ; table 3).

To further unravel the possible sources of disagreement, we studied the inter-rater agreement according to

**Table 1** Agreement for the five CCSVI criteria between the first and the second operator

Criterion (N)	Concordant+	Concordant–	Percentage of agreement	κ	95% CL
I (93)	12	55	72.0	0.29	0.08 to 0.50
II (64)	0	61	95.3	–0.02	–0.05 to 0.01
III (93)	29	28	61.3	0.23	0.03 to 0.43
IV (93)	0	75	80.7	–0.09	–0.15 to 0.3
V (93)	3	76	85.0	0.22	–0.06 to 0.50

CCSVI, chronic cerebrospinal venous insufficiency; CL, confidence limit.

several variables (N=84, borderline excluded). Operators were divided according to their expertise with CDS and CCSVI. In the case of specific CCSVI expertise, the κ value was 0.24 (–0.11 to 0.59) for experts (31 comparisons), 0.20 (–0.33 to 0.73) for non-experts (N=16) and 0.26 (–0.03 to 0.54) for comparisons between one expert and one non-expert (N=37).

As to the general CDS expertise, the κ value was 0.16 (–0.20 to 0.52) for high experts (29 comparisons), and 0.26 (–0.04 to 0.55) for comparisons between one high expert and one low expert (N=42); no calculation was possible for the non-expert (N=13) since one of the two operators did not find any positive patients. The κ value was similar between neurologists (0.21, –0.06 to 0.47; N=57) and non-neurologists (0.18, –0.20 to 0.56; N=27). The κ value was similar between cases (0.19, –0.14 to 0.52) and controls (0.21, –0.08 to 0.49). It was 0.09 (–0.30 to 0.49) for RR patients and 0.34 (–0.25 to 0.94) for the progressive patients. It improved with the disability score: it was –0.40 (–0.69 to –0.10) for patients with EDSS 0–1.5 (N=12), 0.23 (–0.40 to 0.86) for EDSS 2–3.5 (N=11) and 0.50 (–0.10 to 1.1) for EDSS \geq 4 (N=8). Lastly, in the search for a possible practice effect, we divided the patients into three groups, according to the time period of the study. The κ value was 0.19 (–0.19 to 0.56) for the first tertile (N=28), 0.05 (–0.33 to 0.43) for the second tertile (N=28) and 0.33 (–0.03 to 0.67) for the last tertile (N=28).

Zamboni-trained sonographers ascertained CCSVI more frequently than non-trained sonographers, whereas general CDS expertise and the tertile period of the study had no effect (table 4).

The efficacy of our blinding procedure was different between cases and controls (N=84, borderline excluded): the first operator correctly guessed 18/35

cases (51%) and 36/49 controls (74%; p=0.036). The second operator correctly guessed 18 cases (51%) and 39 controls (80%; p=0.006). The Bang Index was 0.16 (–0.13 to 0.45) for cases, and 0.36 (0.15 to 0.57) for controls for the first operator. It was 0.29 (–0.01 to 0.58) for cases, and 0.39 (0.19 to 0.60) for controls for the second operator. These figures indicate a lack of blinding for controls, and a better blinding for the first operator compared with the second.

DISCUSSION

Inter-rater agreement for CCSVI has never been systematically analysed so far, though scattered information is available from case–control studies. We found an unsatisfactory agreement for the diagnosis of CCSVI with an overall κ of 0.20. For Zamboni's five criteria, the agreement was no higher than chance for two criteria (2 and 4), little more than slight for two criteria (3 and 5) and fair for one (criterion 1), according to Landis and Koch's classification.²⁵ Agreement for CCSVI was 0.75 in a study that blindly evaluated 28 participants,¹⁹ and 0.79 in another case–control study.¹¹ We found the worst agreement for criteria 2 and 4, whereas it was better for criteria involving a measurement (as for criterion 5) or a direct visualisation of a venous anatomical abnormality (as criterion 3). This seems to indicate that criteria 2 and 4 are more prone to subjective interpretation. Criterion 2 was also the most critical in the study by Tsiygoulis *et al*,²⁶ together with criterion 5, though their κ values (0.14–0.48) were much higher than ours. Other studies^{27–28} reported only an unreliable per cent agreement. Measurement of CSA is crucial to ascertain the vein stenosis comprised in defining criterion 3; furthermore, the difference between CSA diameter in the

Table 2 Agreement for the jugular vein CSA measurement between the first and the second operator (N=93)

Internal jugular CSA, cm ²	First operator	Second operator	ICC	95% CL	p Value
	Mean, \pm SD, range	Mean, \pm SD, range			
Supine, right side	0.50, \pm 0.37, 0.05–1.96	0.62, \pm 0.46, 0.06–2.43	0.11	–0.09 to 0.31	0.14
Supine left side	0.49, \pm 0.33, 0.06–1.77	0.54, \pm 0.39, 0.03–2.06	0.11	–0.09 to 0.31	0.15
Upright right side	0.16, \pm 0.16, 0–0.73	0.18, \pm 0.21, 0–0.60	0.05	–0.16 to 0.25	0.32
Upright left side	0.18, \pm 0.20, 0–1.06	0.18, \pm 0.20, 0.02–1.28	0.25	0.05 to 0.43	0.008

CL, confidence limit; CSA, cross-sectional area; ICC, intraclass correlation coefficient with 95% CL and p value.

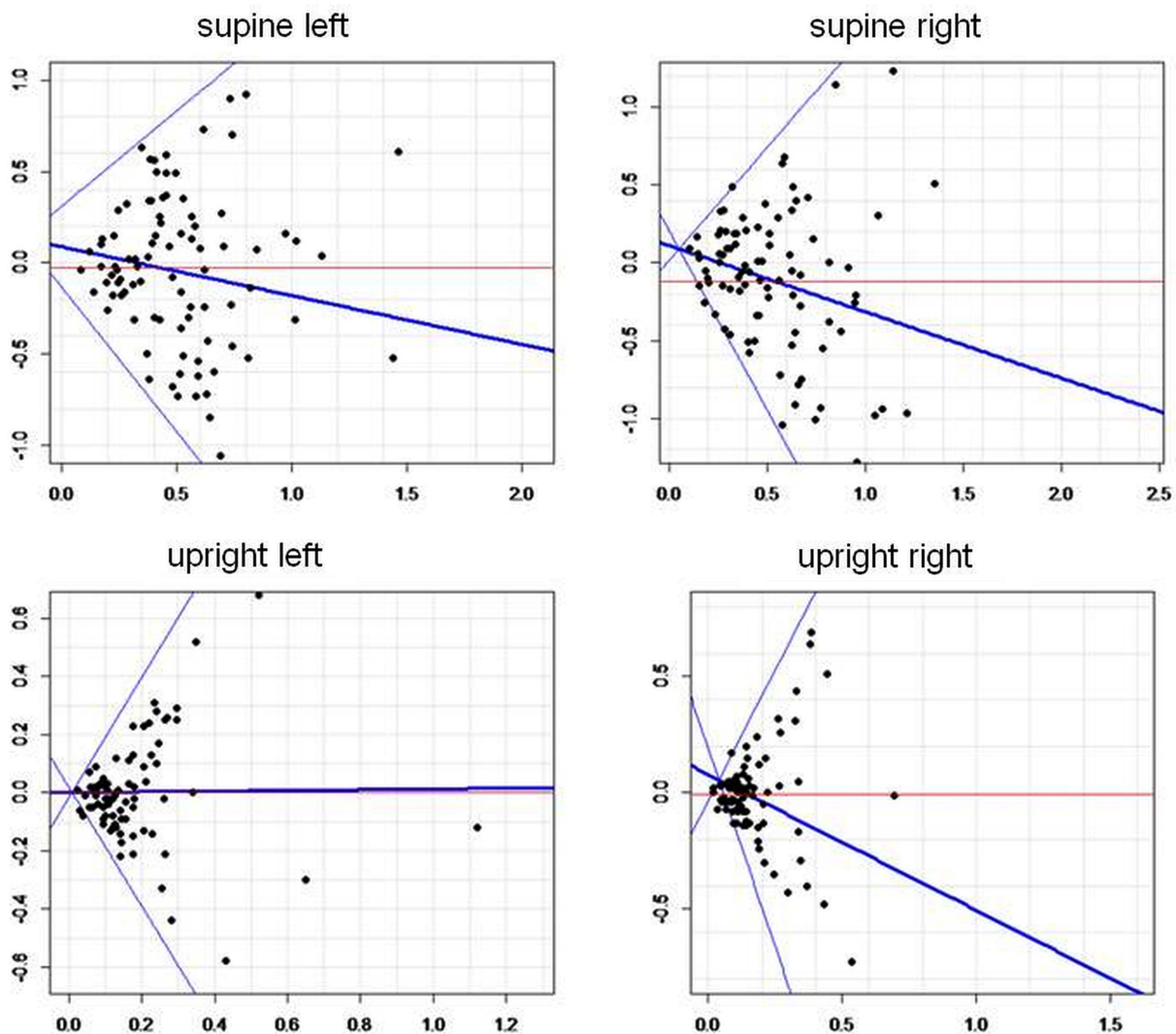


Figure 1 Difference between the measurements of jugular vein cross-sectional area (cm^2) by the first and second operators (y-axis) against the mean of the measurements (x-axis; Bland and Altman plots). Blue line: regression line with upper and lower 95% confidence limit. Red line: mean difference.

supine and sitting positions defines criterion 5. In our study, the only reproducible CSA measure (significant even after Bonferroni correction) was at the left side in the upright position, which was probably a chance result. All measurements, except at the left side in the upright position, were suggestive of a proportional bias,

that is, the agreement between raters is diminished as CSA rises (figure 1). In arterial CDS, the reproducibility of the degree of stenosis is generally much better than that found by us for veins. The κ values for agreement in measurement of the internal carotid artery peak systolic velocity as a categorical variable was 0.85–0.95 in one study.²⁹

We were able to evaluate several possible determinants of inter-rater agreement. Expertise is the most obvious one. We classified the clinical experience of operators based on the total number of ultrasound exams they had performed and their training with Zamboni's courses. Furthermore, we divided the study period in tertiles, on the assumption that an operator's ability would increase with the number of examinations performed. Although only expertise with Zamboni's technique prior to the study (not the number of CDS or the study period) influenced the ability of our operators to detect CCSVI, this did not influence their agreement, which

Table 3 Overall agreement for the diagnosis of CCSVI (chronic cerebrospinal venous insufficiency; N=84, borderline excluded)

	Second operator		Total
	CCSVI positive	CCSVI negative	
First operator			
CCSVI positive	12	17	29
CCSVI negative	12	43	55
Total	24	60	84



Table 4 Percentage of patients with CCSVI positive by expertise and training of the CDS operators (N=84, borderline excluded)

Expertise of the operator	CCSVI positive N (%)	Total N° CDS performed
≥1500 CDS—Zamboni training	23 (33.8)	68
<1500 CDS—Zamboni training	16 (51.6)	31
≥1500 CDS—no Zamboni training	8 (25.0)	32
<1500 CDS—no Zamboni training	6 (16.2)	37
Time period of the study		
First tertile	18 (32.1)	56*
Second tertile	14 (25.0)	56*
Third tertile	21 (37.5)	56*
Total	53 (31.6)	168

≥1500 CDSs vs <1500, p=0.98; Zamboni training versus no Zamboni training, p=0.01.

*p=0.36.

CCSVI, chronic cerebrospinal venous insufficiency; CDS, colour-Doppler sonography.

was similar among operators with different expertise and along the three study periods.

Our study has some limitations. First, we were not able to randomise the operator sequence for practical reasons; however, the prevalence of CCSVI was similar for the first (35%) and second operator (29%), which is an indirect marker of stability of clinical conditions. Second, despite several precautions (see Methods section), our CDS operators were able to identify controls in a percentage higher than chance. This incomplete blinding could also explain why agreement was better for progressive than RR patients, indicating that operators were more prone to agree on CCSVI diagnosis when examining more disabled patients. Third, in addition to the rater, disagreement is amenable to the observed (patient). Variation of the state of the cervical venous system during the daytime and on different days cannot be ruled out. For this reason, we tried to minimise any possible variability due to patients between the first and the second examination by performing all the examinations on the same day, and in the same room. Furthermore, the blood pressure and heart rate were similar before the first and second examinations. The equipment is not a possible source of variability in our study, since we used only one instrument.

In conclusion, we found a low agreement among CDS operators for the diagnosis of CCSVI as a whole or for any of its five criteria. This low reproducibility, associated with a possibly low accuracy of CDS compared to catheter venography, makes this technique of limited diagnostic value, unless accurate and reproducible sonographic criteria are established and verified.^{30 31} Our study evaluated operators from different centres, from different specialties and with different levels of

clinical experience. For this reason, our results imitate the circumstances of a clinical trial and may also be applied to the ‘real world’. Our work may have clinical and research implications. First, low agreement for the CDS CCSVI diagnosis could partly explain the huge difference in prevalence of the condition found in case-control and non-controlled studies. Second, this poor agreement becomes a crucial point when assessing patients for clinical trials; had our group of patients been screened by one operator only, 29 discordant patients out of 84 would have been enrolled in a trial without a certain diagnosis. Third, expertise with Zamboni’s procedure, but not general CDS expertise, improved the ability to detect CCSVI. This suggests that the expertise obtained by studying the arterial trunks is insufficient for a correct approach to venous vessels.

Author affiliations

¹MS Centre, SCDU Neurology, Head and Neck Department, AOU ‘Maggiore della Carità’, Novara, Italy

²IRCAD, Interdisciplinary Research Centre of Autoimmune Diseases, Novara, Italy

³Institute of Diagnostic and Interventional Radiology, AOU ‘Maggiore della Carità’, Novara, Italy

⁴Fondazione ‘Un passo insieme’, Valdellatorre, Italy

⁵Department of Medicine, AOU ‘Maggiore della Carità’, Novara, Italy

⁶SC Neurology, Civile Hospital, Ivrea, Italy

⁷SC Neurology, Maria Vittoria Hospital, Torino, Italy

⁸Department of Neurology, Tappeiner Hospital, Merano, Italy

Acknowledgements The authors are grateful to their patients with multiple sclerosis and controls who underwent the tedious double examination.

Contributors MAL, AS, AC and FM conceived and designed the study. LB, LC, WL, GR, GZ, MM and PL performed the examinations. OR clinically evaluated the patients. MAL, AL and CG analysed the data. MAL, WL, OR and FM wrote the manuscript. All other authors critically reviewed the manuscript. All authors approved the final version of the manuscript. MAL takes full responsibility for the data, the analyses and interpretation, and the conduct of the research. He had full access to all of the data.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests OR was supported by the CRT Foundation, Torino. AL was supported by the Fondo Maffeo-Fondazione della Comunità del Novarese, Novara. MAL serves as associate editor for the European Journal of Neurology and received funding for travel from Aventis, Merck-Serono and Esaote.

Patient consent Obtained.

Ethics approval The study was approved by the Ethical Committee of the ‘AOU Maggiore della Carità’, Novara (# 28/11).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Extra data can be accessed via the Dryad data repository at <http://datadryad.org/> with the doi:10.5061/dryad.7k048.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

1. Zamboni P, Galeotti R. The chronic cerebrospinal venous insufficiency syndrome. *Phlebology* 2010;25:269–79.

2. Zamboni P, Galeotti R, Menegatti E, *et al.* Chronic cerebrospinal venous insufficiency in patients with multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2009;80:392–9.
3. Laupacis A, Lillie E, Dueck A, *et al.* Association between chronic cerebrospinal venous insufficiency and multiple sclerosis: a meta-analysis. *CMAJ* 2011;183:1203–12.
4. Khan O, Filippi M, Freedman MS, *et al.* Chronic cerebrospinal venous insufficiency and multiple sclerosis. *Ann Neurol* 2010;67:286–90.
5. Marder E, Gupta P, Greenberg BM, *et al.* No cerebral or cervical venous insufficiency in US veterans with multiple sclerosis. *Arch Neurol* 2011;68:1521–5.
6. Auriel E, Karni A, Bornstein NM, *et al.* Extra-cranial venous flow in patients with multiple sclerosis. *J Neurol Sci* 2011;309:102–4.
7. Baracchini C, Perini P, Causin F, *et al.* Progressive multiple sclerosis is not associated with chronic cerebrospinal venous insufficiency. *Neurology* 2011;77:844–50.
8. Zaniewski M, Kostecki J, Kuczmik W, *et al.* Neck duplex Doppler ultrasound evaluation for assessing chronic cerebrospinal venous insufficiency in multiple sclerosis patients. *Phlebology* 2013;28:24–31.
9. Amato MP, Saia V, Hakiki B, *et al.* No association between chronic cerebrospinal venous insufficiency and pediatric-onset multiple sclerosis. *Mult Scler* 2012;18:1791–6.
10. Blinkenberg M, Akeson P, Sillesen H, *et al.* Chronic cerebrospinal venous insufficiency and venous stenoses in multiple sclerosis. *Acta Neurol Scand* 2012;126:421–7.
11. Patti F, Nicoletti A, Leone C, *et al.* Multiple sclerosis and CCSVI: a population-based case control study. *PLoS ONE* 2012;7:e41227.
12. Leone M, Raymkulova O, Naldi P, *et al.* Chronic cerebrospinal venous insufficiency is not associated with multiple sclerosis and its severity: a blind-verified study. *PLoS ONE* 2013;8:e56031.
13. Barreto AD, Brod SA, Bui TT, *et al.* Chronic cerebrospinal venous insufficiency: case-control neurosonography results. *Ann Neurol* 2013;73:721–8.
14. Imperiale D, Melis F, Giaccone C, *et al.* Chronic cerebrospinal venous insufficiency in multiple sclerosis: a sonographer-blinded case-control study. *Clin Neurol Neurosurg* 2013;115:1394–8.
15. Bastianello S, Romani A, Viselner G, *et al.* Chronic cerebrospinal venous insufficiency in multiple sclerosis: clinical correlates from a multicentre study. *BMC Neurol* 2011;11:132.
16. Polman CH, Reingold SC, Edan G, *et al.* Diagnostic criteria for multiple sclerosis: 2005 revisions to the 'McDonald Criteria'. *Ann Neurol* 2005;58:840–6.
17. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33:1444–52.
18. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* 1996;46:907–11.
19. Zivadinov R, Marr K, Cutter G, *et al.* Prevalence, sensitivity, and specificity of chronic cerebrospinal venous insufficiency in MS. *Neurology* 2011;77:138–44.
20. Fleiss JL. The measurement and control of misclassification error. In: Fleiss JL, ed. *Statistical methods for rates and proportions*. 2nd edn. New York, NY: John Wiley & Sons, 1981:140–54.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
22. Bang H, Nib L, Davis CE. Assessment of blinding in clinical trials. *Control Clin Trials* 2004;25:143–56.
23. Lu L, Shara N. Reliability analysis: calculate and compare intra-class correlation coefficients (ICC) in SAS. <http://www.nesug.org/proceedings/nesug07/sa/sa13.pdf>
24. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, <http://www.R-project.org>.
25. Landis JR, Koch GC. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
26. Tsivgoulis G, Mantatzis M, Bogiatzi C, *et al.* Extracranial venous hemodynamics in multiple sclerosis: a case-control study. *Neurology* 2011;77:1241–5.
27. Monti L, Menci E, Ulivelli M, *et al.* Quantitative colour Doppler sonography evaluation of cerebral venous outflow: a comparative study between patients with multiple sclerosis and controls. *PLoS ONE* 2011;6:e25012.
28. Ciccone MM, Galeandro AI, Scicchitano P, *et al.* Multigate quality Doppler profiles and morphological/hemodynamic alterations in multiple sclerosis patients. *Curr Neurovasc Res* 2012;9:120–7.
29. Corriveau MM, Johnston KW. Interobserver variability of carotid Doppler peak velocity measurements among technologists in an ICAVL-accredited vascular laboratory. *J Vasc Surg* 2004;39:735–41.
30. Simka M, Ludyga T, Latacz P, *et al.* Diagnostic accuracy of current sonographic criteria for the detection of outflow abnormalities in the internal jugular veins. *Phlebology* 2013;28:285–92.
31. Zamboni P, Morovic S, Menegatti E, *et al.* Screening for chronic cerebrospinal venous insufficiency (CCSVI) using ultrasound—recommendations for a protocol. *Int Angiol* 2011;30:571–97.