



University
of Glasgow

Dittrich, R., Ritter, M.A., Kaps, M., Siebler, M., Lees, K.R., Larrue, V., Nabavi, D.G., Ringelstein, E.B., Markus, H.S. and Droste, D.W. (2006) *The use of embolic signal detection in multicenter trials to evaluate antiplatelet efficacy: signal analysis and quality control mechanisms in the CARESS (Clopidogrel and Aspirin for Reduction of Emboli in Symptomatic carotid Stenosis) trial.* Stroke, 37 (4). pp. 1065-1069. ISSN 0039-2499

<http://eprints.gla.ac.uk/21021/>

Deposited on: 23 January 2012



**The Use of Embolic Signal Detection in Multicenter Trials to Evaluate
Antiplatelet Efficacy : Signal Analysis and Quality Control Mechanisms in the
CARESS (Clopidogrel and Aspirin for Reduction of Emboli in Symptomatic
carotid Stenosis) Trial**

Ralf Dittrich, Martin A. Ritter, Manfred Kaps, Mario Siebler, Kennedy Lees, Vincent Larrue, Darius G. Nabavi, E. Bernd Ringelstein, Hugh S. Markus and Dirk W. Droste

Stroke 2006, 37:1065-1069: originally published online March 2, 2006

doi: 10.1161/01.STR.0000209331.39894.3d

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2006 American Heart Association. All rights reserved. Print ISSN: 0039-2499. Online
ISSN: 1524-4628

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://stroke.ahajournals.org/content/37/4/1065>

Subscriptions: Information about subscribing to *Stroke* is online at
<http://stroke.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters
Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax:
410-528-8550. E-mail:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/reprints>

The Use of Embolic Signal Detection in Multicenter Trials to Evaluate Antiplatelet Efficacy

Signal Analysis and Quality Control Mechanisms in the CARESS (Clopidogrel and Aspirin for Reduction of Emboli in Symptomatic carotid Stenosis) Trial

Ralf Dittrich, MD; Martin A. Ritter, MD; Manfred Kaps, MD; Mario Siebler, MD; Kennedy Lees, FRCP; Vincent Larrue, MD; Darius G. Nabavi, MD; E. Bernd Ringelstein, MD; Hugh S. Markus, FRCP; Dirk W. Droste, MD

Background and Purpose—The CARESS (Clopidogrel and Aspirin for Reduction of Emboli in Symptomatic carotid Stenosis) trial proved the effectiveness of the combination of clopidogrel and aspirin compared with aspirin alone in reducing presence and number of microembolic signals (MES) in patients with recently symptomatic carotid stenosis. The present study aimed at installing primary and secondary quality control measures in CARESS because MES evaluation relies on subjective judgment by human experts.

Methods—As primary quality control, centers participating in CARESS evaluated a reference digital audio tape (DAT) before the study containing both MES and artifacts. Interobserver agreement of classifying signals as MES was expressed as proportions of specific agreement of positive ratings ($ps \pm$ values). For all DATs included in CARESS ($n=300$), online number of MES and off-line number of MES read by the central reader were compared using correlation coefficients. As secondary control, a sample of 16 of 300 DATs was cross-validated by another independent reader (post-trial validator).

Results—For the reference tape, the cumulative $ps \pm$ value was 0.894 based on 12 of 14 observers. Two observers with very different results improved after a training procedure. Agreement between post-trial validator and central reader was $ps + = 0.805$, indicating very good agreement. Correlation between online evaluation and off-line evaluation of DATs was very good overall (cumulative $\rho = 0.84$; $P < 0.001$).

Conclusion—Multicenter studies using MES as outcome parameter are feasible. However, primary and secondary quality control procedures are important. (*Stroke*. 2006;37:1065-1069.)

Key Words: antiplatelet agents ■ carotid stenosis ■ stroke ■ ultrasonography, Doppler, transcranial

Clinically silent cerebral microembolic signals (MES) can frequently be detected by transcranial Doppler (TCD) sonography in patients with symptomatic carotid stenosis.¹⁻⁵ Several small studies have shown that MES predict recurrent stroke or transient ischemic attack (TIA) and stroke alone in patients with symptomatic carotid stenosis.⁵⁻⁸ MES-positive patients with symptomatic carotid stenosis therefore define a subset of high-risk patients for stroke recurrence. In this group, MES offer an attractive surrogate marker to evaluate antiplatelet efficacy. The CARESS (Clopidogrel and Aspirin for Reduction of Emboli in Symptomatic carotid Stenosis) trial demonstrated a superior efficacy of clopidogrel plus

aspirin compared with aspirin alone in reducing the presence and the number of MES after 1 week of treatment.⁹ Interobserver agreement in the identification of MES has been shown to be high within centers and between very experienced centers. However, disagreement can occur, particularly for MES with small intensity increases.^{10,11} The standard of evaluation of MES is still the off-line evaluation of a recorded investigation by a human expert.¹² Although there have been efforts to develop intelligent software capable of automatically classifying events as MES or artifacts, it is still felt that the technique is premature.¹²⁻¹⁴ For this reason, as for most multicenter trials involving human experts' judgment (eg, neu-

Received August 30, 2005; final revision received November 28, 2005; accepted December 14, 2005.

From the Department of Neurology (R.D., M.A.R., D.G.N., E.B.R., D.W.D.), University Hospital of Münster, Germany; Department of Neurology (M.K.), University Hospital of Giessen, Germany; Department of Neurology (M.S.), University Hospital of Düsseldorf, Germany; Division of Cardiovascular and Medical Sciences (K.L.), University of Glasgow, United Kingdom; Department of Neurology (V.L.), University of Toulouse, France; Centre for Clinical Neuroscience (H.S.M.), St. George's University of London, United Kingdom; and Department of Neurology (D.W.D.), Centre Hospitalier de Luxembourg.

The first 2 authors contributed equally to this study.

Correspondence to Martin A. Ritter, MD, Department of Neurology, University of Münster, Albert-Schweitzer-Straße 33, 48129 Münster, Germany. E-mail ritterm@uni-muenster.de

© 2006 American Heart Association, Inc.

Stroke is available at <http://www.strokeaha.org>

DOI: 10.1161/01.STR.0000209331.39894.3d

roimaging), signal analysis during CARESS was performed in a single reference center by an observer blinded to clinical and patient identity details. However, to the study required individual centers to screen for the presence, or absence, of MES.

The aim of the present study was to implement primary and secondary quality control procedures for the evaluation of MES. Primary quality control mechanisms before patient recruitment were designed to ensure consistency in criteria used to detect MES across centers. Secondary mechanisms were designed to ensure reproducibility of the off-line analysis performed by the central reader.

Methods

Primary Quality Control: Interobserver Agreement for a Reference Tape

Before randomization of the first patient, agreement in interpretation of MES between centers was determined using a reference tape. This tape contained MES of patients with carotid artery stenosis of varying intensities recorded in the ipsilateral middle cerebral artery and also artifacts derived from subjects coughing or talking, tapping the probe, and electronic interference. Recordings from patients with additional sources of MES were not used.

All recordings had been made previously for research purposes.^{3,13} The Doppler signal had been recorded with an EME TCD device (TC 4040; Nicolet/EME) using a 2-MHz transducer on digital audio tape (DAT). The machine used a 128-point fast Fourier transform (FFT) analysis and used a graded color scale to display the intensity of the received Doppler signal. FFT time frame overlap was 67%. A sample volume of 4 to 5 mm in length was used, and ultrasound emission power was set to 22% of maximum output. Copies of this tape were sent to each center with an interest to participate in the trial. The settings of the ultrasound devices in the centers were the settings that the centers used for MES evaluation in daily routine. The following types of ultrasound devices were used for playback in the respective centers: DWL Multidop X4 and Nicolet/EME TC4040. Signal processing was performed at the playback stage using FFTs, filter settings, and window overlaps as specified in Table 1.

Published international consensus criteria for MES detection were used by each observer.¹² Although a decibel threshold has been shown to improve agreement,^{10,11} no decibel threshold was used for

this study because different analysis packages on different TCD systems measure decibel levels in different ways, making comparisons of measured intensities unreliable. Each observer noted the exact time of MES occurrence in (hh:mm:ss) according to the counter on the DAT recorder. A time window of ± 1 s was allowed for allocating a MES correctly. The decibel value measured in the reference center was used in analysis to determine whether disagreement in signal evaluation was dependent on signal intensity.

In total, 14 different experienced observers from 8 different centers were involved:¹ Münster, Germany (6 investigators)²; Glasgow, United Kingdom (2 investigators)³⁻⁸; Bern, Switzerland; Düsseldorf, Germany; Gießen, Germany; London, United Kingdom; Lübeck, Germany; and Toulouse, France (1 investigator each).

Because the classification of MES depends on the judgment of a human observer, we classified an event as definite MES in case of agreement of ≥ 9 different investigators on the same signal. These events were taken as the "gold standard." Interobserver agreement was expressed as proportions of specific agreement for positive ratings ($ps+$):¹⁵ $ps+ = 2MES_b / 2MES_b + MES_1 + MES_2$, where MES_b indicates the number of events classified as MES by both observers; MES_1 indicates number of events classified as MES only by the first observer but not by the second observer; and MES_2 indicates number of events classified as MES only by the second observer but not by the first observer.

$ps\pm$ values provide probabilities that if one randomly selected observer makes a positive rating (in this case, declaring a signal to be a MES at a defined time), another randomly selected observer will also make a positive rating (ie, declare the same signal as MES). With multiple observers, the proportion declares the probability in a cumulative manner. $ps+$ values are comparable with widely used κ -statistics. However, κ -values are not meaningful in the case of MES because no specific negative rating is made during MES analysis.¹⁵

Secondary Quality Control: Cross-Validation of Off-Line Reading Results by Post-Trial Validator

Relevant Study Design

Patients were eligible for the CARESS trial if they had recently symptomatic (with TIA or stroke within the last 3 months) carotid stenosis $\geq 50\%$ based on ultrasound criteria. In subjects meeting these entry criteria, a screening TCD was performed, and subjects were eligible for randomization if ≥ 1 MES were detected during a 1-hour recording. The detailed study design has been reported previously.⁹

TABLE 1. Setup of Doppler Devices in Participating Centers

Center	1	2	3	4	5	6	7	8	9	10	11
	601	201	701	703	705	702	708	1301	1305	1303	1302
Doppler device	EME Pioneer	EME Pioneer	EME Pioneer	DWL Multidop	EME Pioneer	DWL Multidop	DWL Multidop	EME Pioneer	EME Pioneer	EME Pioneer	EME Pioneer
Transducer size	2 cm	2 cm	2 cm	1.5 cm	1 cm	1.5 cm	1.5 cm	2 cm	2 cm	2 cm	2 cm
Transmission frequency	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz	2 MHz
Sample volume	5 mm	5 mm	5 mm	5 mm	5 mm	7.5 mm	5 mm	5 mm	5 mm	5 mm	5 mm
Scale setting (above/below baseline)	120/120 cm/s	140/140 cm/s	130/30 cm/s	200/80 cm/s	150/10 cm/s	120/10 cm/s	150/90 cm/s	70/70 cm/s	70/70 cm/s	70/70 cm/s	70/70 cm/s
High-pass filter	258 Hz	254 Hz	106 Hz	104 Hz	254 Hz	104 Hz	104 Hz	254 Hz	106 Hz	254 Hz	254 Hz
FFT (points)	128	128	128	128	256	64	64	128	128	128	128
Software used	TCD 2.4	TCD 2.4	TCD 2.4	TCD 8	TCD 2.4	TCD 8	TCD 8	TCD 2.4	TCD 2.4	TCD 2.4	TCD 2.4
Window overlap	66%	72%	67%	50%	50%	50%	50%	66%	55%	67%	67%

TABLE 2. Stratification of DATs for Secondary Quality Control

Stratum	Online	Off-Line	n	No. in Stratum
1	Negative	Positive	3	29
2	Positive	Negative	3	35
3	Positive	1–7 MES	1	12
4	Positive	8–19 MES	1	49
5	Positive	≥20 MES	3	33

Column 1 gives the No. of the stratum; column 2 (online) gives the results of the online reading; column 3 (off-line), the results of the central reader.

The results of the off-line reader were further stratified according to the No. of MES found per tape (1–7 MES; 8–19 MES; ≥20 MES).

Positive indicates investigator found MES; negative, investigator did not find MES; n, No. of tapes randomly selected for cross-validation; No. in stratum, absolute No. of tapes within stratum.

During CARESS, 3 recordings within 1 week were performed on-site in the randomizing centers and were stored on DAT. DATs were analyzed off-line in the central reader. The central reader was blinded to patient identity and to which tape of a given patient corresponded to which visit.

Assessment of Off-Line Reader

For secondary quality control, 16 DATs of 300 recorded for the core study were transferred to London (the reference center for the United Kingdom) after the central reading procedure and were evaluated in the same fashion as in Münster by 1 experienced observer. The second reference reader was termed “post-trial validator.”

Five tapes of the 16 were randomly selected by the central data management center in Paris while the study was ongoing. Eleven tapes were selected after the trial based stratified by center and MES count. All tapes had been allocated to 1 of 5 strata and at least 1 tape was selected from each stratum as well as 1 from each center. The strata and the number of tapes within each stratum are given in Table 2.

Stratification ensured that no tape was selected without any MES (ie, 142 tapes negative on both online and off-line analysis). A selection of tapes was made with a range containing a few too many MES. The number of MES and the exact time of their occurrence were documented, thus enabling a signal-to-signal comparison with the results of the central reader, and $ps\pm$ values were calculated.

Correlation of Off-Line and Online Reading Results

After the end of the CARESS trial and unblinding of the reading results, the online findings documented by the study site investigator were compared with the results of the central reader. A signal-to-signal comparison was not possible because of the fact that online reading did not record the exact time points at which MES occurred. Intercenter agreement was therefore expressed by both numbers of MES detected and the decision whether a tape was MES positive or not. Correlation coefficients were calculated for each center independently, as well as for all centers cumulatively.

TABLE 3. Results of Primary Control: Results From First Master Tape (reference tape)

Investigator	1	2	3	4	5	6	7	8	9	10	11	12	13	14
No. of MES	115	92	99	99	99	91	99	96	82	90	98	117	135	95
No. of events classified as MES only by this individual observer	5	0	0	3	0	5	0	1	0	1	1	37	66	0
No. of MES/91 MES identified by ≥9 observers	91	88	90	85	89	81	90	89	79	83	88	25	27	89

In the first line, the No. of the respective investigator is given. In the second line, the absolute No. of events the individual investigator classified as MES is recorded. In the third line, the No. of events recorded as MES by the individual observer, which none of the other observers recorded as MES, is given. The fourth line indicates the No. of MES identified by the individual investigator out of the 91 MES defined as “gold standard.”

Results

Primary Quality Control: Interobserver Agreement for a Reference Tape

A total of 290 different signals on the reference tape were classified as MES by ≥1 observer. Ninety-one signals were classified as MES by ≥9 observers. Table 3 shows the results separately for each individual observer.

Two observers (12 and 13) had results very different from the remaining 12 observers. These observers detected only 25 and 27 of the 91 reference MES. A total of 145 signals were detected by only 1 or both of these 2 observers. The performances of these 2 observers were independent of signal intensity. The other 12 observers detected ≥79 of the 91 reference MES (86%). From the remaining 56 signals that were detected by ≥1 of the 12 observers who performed consistently, only 7 signals had intensities >5 dB as measured by the reference center, indicating that disagreement occurred mainly in signals with low intensity.

Excluding observers 12 and 13, the $ps\pm$ value was 0.894, signifying a 90% likelihood that if 1 of a pair of observers declared an event as MES, the other observer also declared the event as MES. After including investigators 12 and 13, the overall $ps\pm$ value (agreement between all observers) fell to 0.694. The results of the centers apart from 12 and 13 were considered excellent.

Observers 12 and 13 were therefore considered outliers requiring retraining. Retraining was performed by 1 of the reference centers. The observers were then re-evaluated using a second reference tape created in an analogous way as the first one. Two observers from the 12 observers with good agreement on the first tape agreed on 75 MES on the second reference tape. For this tape, $ps+$ was 0.95 for investigator 12 and 0.71 for investigator 13, showing marked improvement.

Secondary Quality Control: Cross-Validation of the Off-Line Reading Results With the Findings of the Post-Trial Validator

The $ps\pm$ value for the 5 randomly selected tapes during the study was 0.815. An additional 11 tapes were validated after the study. For the 3 stratum-1 tapes judged MES negative during the online analysis but found positive during off-line analysis, all were found MES negative by the post-trial validator. For the 3 stratum-2 tapes judged MES positive during the online analysis but MES negative during the off-line analysis, all 3 were also found MES negative by the post-trial validator's analysis. For the 1 tape from stratum-3 (ie, MES positive online and MES positive off-line, but with

only few MES) the post-trial validator agreed with all 6 signals of the central reader but also found 2 more signals. There was good agreement for the 1 tape from stratum-4 (ie, MES positive on-line, as well as MES positive off-line, with a considerable number of MES); the post-trial validator and central reader agreed on 7 signals, and each validator just found 1 extra. There was also good agreement on the 3 tapes in the high-count stratum. The overall κ value for these 11 tapes was 0.802. The corresponding value for all 16 tapes was $\kappa = 0.805$. In summary, there was very good agreement between central reader and post-trial validator.

Correlation of Off-Line and Online Reading Results

For 6 of 107 randomized patients, the central reading center did not report off-line MES during baseline recording. Therefore, these 6 patients were excluded from the per-protocol analysis to avoid bias because only online MES-positive patients at baseline met the entry criteria. The numbers of MES registered during central reading versus online analysis within each tape are given separately for each center in the Figure.

There was an excellent correlation between the online reader and the central reader in 6 of 11 centers ($\rho \geq 0.94$ for 6 out of 11 centers). For 3 centers, agreement was very good in 1 ($\rho = 0.88$), good in 1 ($\rho = 0.69$), and borderline in 1 ($\rho = 0.42$; $P = 0.057$) case. For 2 centers, there was no correlation

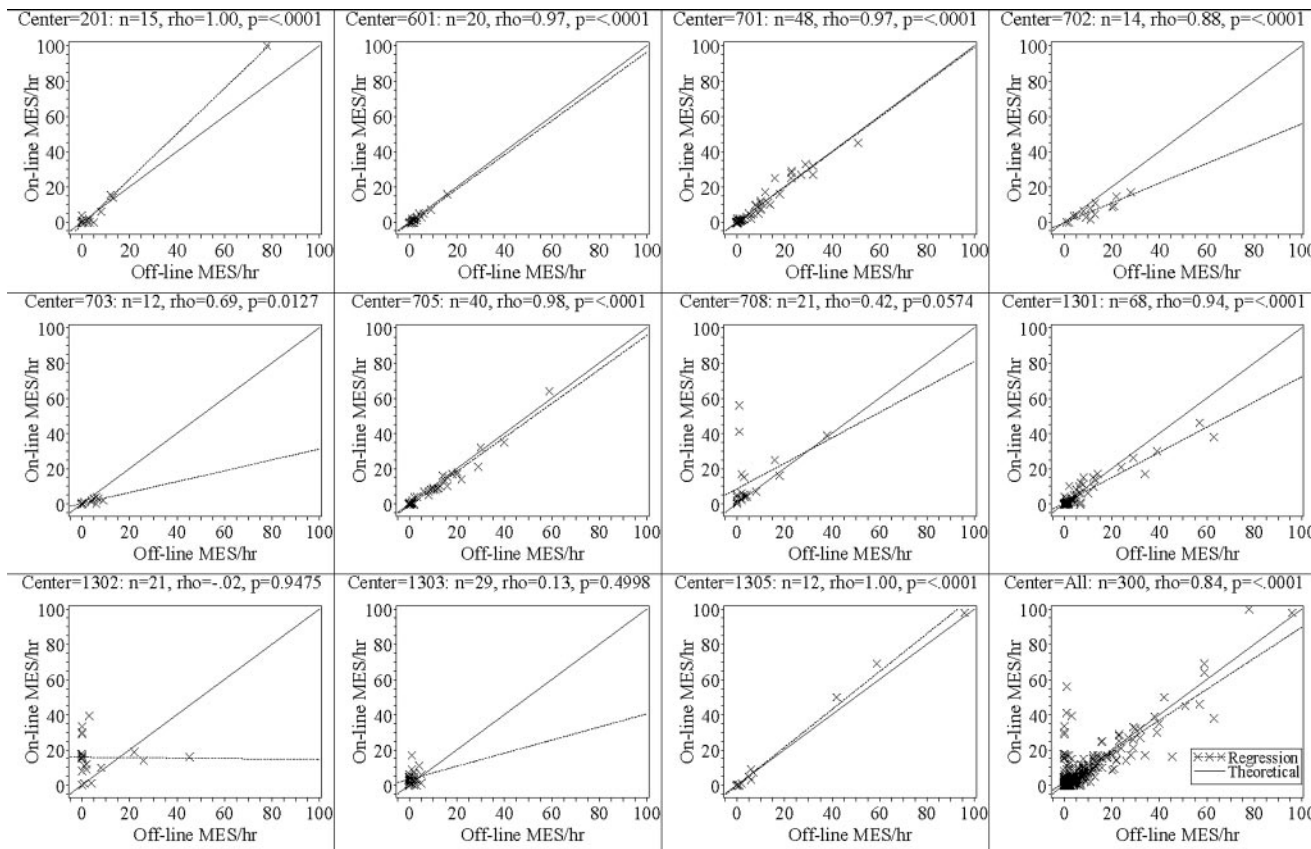
between off-line and online analysis ($\rho = -0.02$ and 0.13). In both of these centers, the ultrasound technician had changed during the trial. The correlation of all centers combined compared with the central reader was excellent ($\rho = 0.84$; $P < 0.001$).

The main source of disagreement was, as expected, low intensity of the signals (≤ 5 dB as measured by the EME machine used for central analysis). This was true as well for the comparison between the online reader and the central reader as well as for the comparison of the post-trial validator and the central reader.

Discussion

CARESS is the first multicenter study to use MES detection as a surrogate end point to evaluate antiplatelet efficacy. The trial design required individual centers to screen subjects for the presence of MES before randomization. This required good levels of agreement in detection of MES between centers. In this study, we demonstrated that with appropriate quality control measures set in place prospectively, such multicenter studies are feasible.

Similar issues apply to all studies using neuroimaging as a surrogate end point. For the interpretation of computed tomography (CT) or MRI scans as surrogate markers of disease, agreement rates of 0.6 to 0.7 are considered very acceptable if κ statistics are applied.¹⁶⁻¹⁸ κ values are comparable with



Correlation graphs for all centers individually and combined. The figure shows the scatter plots and resulting correlation graphs for all participating centers individually as well as in the lower right corner the cumulative results for all centers compared with the central reading center. In the first line of each field center number, number of tapes contributed to the study (n), resulting coefficient (ρ), and resulting P values are given. Each x in the graphs indicates coordinates for the pair of results of absolute MES counts in 1 tape achieved by the online reader (y axis) and the central reader (x axis).

κ statistics.¹⁵ Therefore, our study shows that equivalent or higher levels of agreement can be reached in MES reading than in CT or MRI reading.

We demonstrated that an important quality control feature is an assessment of performance before study commencement. Although most observers demonstrated high levels of agreement at this stage, 2 observers reported MES very different from the others. The high overall level of agreement for most observers is consistent with previous international reproducibility studies, which have shown good agreement among most centers except for a few outliers.^{10,11} In our study, the identification of outliers before enrollment of any patients allowed retraining of technicians in these centers, after which performance markedly improved. Subsequent comparison of online and off-line analysis demonstrated poor correlation in only 2 centers, both of which had a change in ultrasound technician during the study. This emphasizes the importance of re-evaluating any new technicians during the study, something that was not performed in CARESS.

Previous studies have shown high levels of interobserver reproducibility in the detection of MES.^{10,11} These were cross-sectional studies and involved a number of internationally recognized and expert centers. We demonstrated similar levels of agreement between most centers in this study. Our study extends these findings in a number of ways. First, we included all centers in the CARESS trial, not all of which had such extensive experience in MES detection. Second, we evaluated the effect of retraining less well-performing centers and demonstrated this resulted in a marked improvement in performance. Third, in this prospective study, we monitored the performance of centers during the trial by comparing their online analysis with the off-line analysis of the central reading center. This demonstrated the need for continued monitoring and particularly re-evaluation and, if necessary, training if technicians performing the recordings change during the study. This is a standard procedure in other clinical trials using neuroimaging.^{16–18} This was something that was not performed in the CARESS study.

There are a number of potential limitations in our study. First, 290 signals were declared as MES by ≥ 1 observer, but only 91 MES were considered as “reference MES” against which statistical tests were performed. This seemingly high level of disagreement in the number of MES is well explained by the fact that the 2 outliers were responsible for 145 of 199 of the signals not considered as MES for the purpose of the study and that only 7 of the remaining 56 signals had high signal intensities. This demonstrates that the disagreement occurred mainly in the low-intensity range, something that was expected before the trial.

Therefore, not using a decibel threshold for MES could be regarded as the second limitation. However, because of the different equipment used, it was not possible to use a fixed value. The lack of an intensity threshold also did allow us to evaluate agreement across the full range of MES intensity and demonstrated the relationship between disagreement and intensity. A third limitation could be that not all centers used the same ultrasound device for signal analysis. This could lead to a systematic error for centers that use a device different from the central reading center. However, our study does not

support this assumption because the results of the readers for the reference tape and during the trial using DWL or EME machines were not very different.

In summary, we have shown that it is possible to standardize MES reading and that MES are a robust tool suitable as a surrogate marker for clinical trials as long as independent quality control mechanisms are installed.

References

1. Siebler M, Sitzer M, Rose G, Bendfeldt D, Steinmetz H. Silent cerebral embolism caused by neurologically symptomatic high-grade carotid stenosis. Event rates before and after carotid endarterectomy. *Brain*. 1993; 116:1005–1015.
2. Blaser T, Hofmann K, Buerger T, Effenberger O, Wallech CW, Goertler M. Risk of stroke, transient ischemic attack, and vessel occlusion before endarterectomy in patients with symptomatic severe carotid stenosis. *Stroke*. 2002;33:1057–1062.
3. Droste DW, Dittrich R, Kemeny V, Schulte-Altendorneburg G, Ringelstein EB. Prevalence and frequency of microembolic signals in 105 patients with extracranial carotid artery occlusive disease. *J Neurol Neurosurg Psychiatry*. 1999;67:525–528.
4. Markus HS, Thomson ND, Brown MM. Asymptomatic cerebral embolic signals in symptomatic and asymptomatic carotid artery disease. *Brain*. 1995;118:1005–1011.
5. Markus HS, MacKinnon A. Asymptomatic embolization detected by Doppler ultrasound predicts stroke risk in symptomatic carotid artery stenosis. *Stroke*. 2005;36:971–975.
6. Valton L, Larrue V, le Traon AP, Massabau P, Geraud G. Microembolic signals and risk of early recurrence in patients with stroke or transient ischemic attack. *Stroke*. 1998;29:2125–2128.
7. Babikian VL, Wijman CA, Hyde C, Cantelmo NL, Winter MR, Baker E, Pochay V. Cerebral microembolism and early recurrent cerebral or retinal ischemic events. *Stroke*. 1997;28:1314–1318.
8. Siebler M, Nachtmann A, Sitzer M, Rose G, Kleinschmidt A, Rademacher J, Steinmetz H. Cerebral microembolism and the risk of ischemia in asymptomatic high-grade internal carotid artery stenosis. *Stroke*. 1995;26:2184–2186.
9. Markus HS, Droste D, Kaps M, Larrue V, Lees K, Siebler, Ringelstein EB. Dual antiplatelet therapy with clopidogrel and aspirin in symptomatic carotid stenosis evaluated using Doppler embolic signal detection; the multicentre CARESS study. *Circulation*. 2005;111:2233–2340.
10. Markus H, Bland JM, Rose G, Sitzer M, Siebler M. How good is intercentre agreement in the identification of embolic signals in carotid artery disease? *Stroke*. 1996;27:1249–1252.
11. Markus HS, Ackerstaff R, Babikian V, Bladin C, Droste D, Grosset D, Levi C, Russell D, Siebler M, Tegeler C. Intercentre agreement in reading Doppler embolic signals. A multicentre international study. *Stroke*. 1997; 28:1307–1310.
12. Ringelstein EB, Droste DW, Babikian VL, Evans DH, Grosset DG, Kaps M, Markus HS, Russell D, Siebler M. Consensus on microembolus detection by TCD. International Consensus Group on Microembolus Detection. *Stroke*. 1998;29:725–729.
13. Kemeny V, Droste DW, Hermes S, Nabavi DG, Schulte-Altendorneburg G, Siebler M, Ringelstein EB. Automatic embolus detection by a neural network. *Stroke*. 1999;30:807–810.
14. Cullinane M, Reid G, Dittrich R, Kaposzta Z, Ackerstaff R, Babikian V, Droste DW, Grosset D, Siebler M, Valton L, Markus HS. Evaluation of new online automated embolic signal detection algorithm, including comparison with panel of international experts. *Stroke*. 2000;31:1335–1341.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
16. Coutts SB, Demchuk AM, Barber PA, Hu WY, Simon JE, Buchan AM, Hill MD; VISION Study Group. Interobserver variation of ASPECTS in real time. *Stroke*. 2004;35:e103–e105.
17. Van Straaten EC, Scheltens P, Knol DL, van Buchem MA, van Dijk EJ, Hofman PA, Karas G, Kjartansson O, de Leeuw FE, Prins ND, Schmidt R, Visser MC, Weinstein HC, Barkhof F. Operational definitions for the NINDS-AIREN criteria for vascular dementia: an interobserver study. *Stroke*. 2003;34:1907–1912.
18. Motto C, Ariztu E, Boccardi E, De Grandi C, Piana A, Candelise L. Reliability of hemorrhagic transformation diagnosis in acute ischemic stroke. *Stroke*. 1997;28:302–306.