

## Supplementary Methods

The clinical findings obtained in this study included demographic characteristics and vascular risk factors such as hypertension, diabetes, dyslipidemia, smoking, obesity, and previous history of stroke. Stroke severity was measured using the National Institutes of Health Stroke Scale by well-trained neurologists upon admission and discharge. The risk of paradoxical embolism score was calculated to evaluate the possibility of stroke associated with patent foramen ovale (PFO). The characteristics of magnetic resonance imaging performed within 7 days of the onset of symptoms were classified according to stroke lesion patterns, vascular territories, and angiographic findings. Stroke lesion patterns were categorized as single solitary, single territory multiple, multiple territory scattered, and large territory and/or additional. Vascular territories were categorized as anterior (anterior and middle cerebral arteries), posterior (vertebrobasilar artery), and both. Angiographic findings were categorized according to the presence of major arterial occlusions. Patients whose stroke lesions were not confirmed were classified separately as having a transient ischemic attack. D-dimer levels were assessed using a quantitative D-dimer latex agglutination assay. The size of the left atrium was measured using echocardiography, and measurements exceeding 40 mm were considered indicative of left atrial enlargement. Frequent atrial premature complexes were defined as cases where atrial premature complexes exceeded 1% of all heartbeats.

Additional tests performed within 2 weeks of the onset of symptoms to determine the etiology of cryptogenic stroke were as follows: transcranial Doppler (TCD) sonography bubble study, transesophageal echocardiography (TEE), and contrast-enhanced chest computed tomography (CT). The TCD bubble study was performed using a TCD monitoring device (PMD 150; Spencer Technologies, Redmond, WA, USA) and two 2-MHz probes fixed in a metal headframe (Marc 1500; Spencer Technologies). Microembolic signals (MESs) were detected by the monitoring device and probes heading toward the bilateral middle cerebral artery at depths of 40–60 mm through the temporal window. Agitated saline was prepared by mixing two syringes, one with 9 mL of 0.9% saline and the other with 1 mL of air connected by a three-way stopcock, which was injected three times through the antecubi-

tal vein. The procedure was performed during the Valsalva maneuver and at rest. MESs were recorded and counted using the computer software embedded in the device. All standardized protocols were performed by skilled sonographers. Right-to-left shunt (RLS) was classified as grade I (1–10 microbubbles), grade II (11–30 microbubbles), grade III (31–100 microbubbles), grade IV (101–300 microbubbles), and grade V (>300 microbubbles); including “shower” or “curtain” shunt patterns, defined as “massive RLS” in this study. In all patients with massive RLS in the TCD bubble study, the presence of PFO was confirmed with the TEE bubble study evaluated by cardiologists. Contrast-enhanced CT was considered when extracardiac RLS was suspected, as in pulmonary arteriovenous malformation (PAVM).

Chronic lung parenchymal disease was defined as a composite of bronchiectasis, chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, and lung cancer. The diagnosis of chronic lung parenchymal disease before stroke was examined by reviewing medical records and tests, including chest CT and pulmonary function tests, and was deemed reliable only with official readings written by professional radiologists or pulmonologists or with biopsy confirmation. Lung cancer was defined as primary or metastatic lung cancer currently diagnosed, treated, or with a recurrence of prior inactive cancer within 6 months of stroke occurrence.

## Statistical analysis

The study population was divided into RLS source(+) and RLS source(-) groups. The RLS source(+) group consisted of patients with a confirmed etiology of intracardiac RLS or PAVM. The remaining RLS source(-) group consisted of patients whose RLS source was not identified using the TEE bubble study and additional tests, including chest CT. To compare the data between the two groups, the Pearson  $\chi^2$  test or Fisher's exact test was used for nominal variables, and the two-sample t-test or Mann-Whitney U test was used for continuous variables. Subgroup analyses comparing the characteristics between the patients with chronic lung parenchymal disease in the RLS source(-) group and conventional PFO in the RLS source(+) group were also performed. These analyses were performed using R software for Windows version 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria).