Phone and Video-Based Modalities of Central Blinded Adjudication of Modified Rankin Scores in an Endovascular Stroke Trial

Elena López-Cancio, MD, PhD; Mercè Salvat, RN; Neus Cerdà, MS; Marta Jiménez, MD; Javier Codas, MD; Laura Llull, MD; Sandra Boned, MD; Luis M. Cano, MD; Blanca Lara, MD; Carlos Molina, MD, PhD; Erik Cobo, MD, PhD; Antoni Dávalos, MD, PhD; Tudor G. Jovin, MD; Joaquín Serena, MD, PhD; on behalf of REVASCAT investigators

Background and Purpose—The standard outcome measure in stroke research is modified Rankin scale (mRS) evaluated by local blinded investigators. We aimed to assess feasibility and reliability of 2 central adjudication methods of mRS in the setting of a randomized endovascular stroke trial.

Methods—This is a secondary analysis derived from the Randomized Trial of Revascularization With Solitaire FR Device Versus Best Medical Therapy in the Treatment of Acute Stroke Due to Anterior Circulation Large Vessel Occlusion Presenting Within Eight Hours of Symptom Onset (REVASCAT) trial cohort. Primary outcome was distribution of mRS at 90 days. Local evaluation was done by certified investigators masked to treatment assignment using structured face-to-face interviews. In addition, central assessment was performed by 2 independent raters via structured phone interview (n=120) and via video recordings of the face-to-face interviews with local investigators (n=106). Interrater agreement was evaluated using kappa and discordance statistics. Sensitivity analyses for the primary end point using different adjudication approaches were performed. Correlation between mRS obtained with each modality and 24-hour follow-up infarct volumes was studied.

Results—Using local evaluation as the reference, higher agreement rates were noted with central video than with central phone evaluations (k = 0.92 [0.88–0.96] versus 0.77 [0.72–0.83]). Discrepancies in mRS scoring between local and central raters (phone- and video-based) were similar in both treatment allocation arms. Sensitivity analyses showed benefit of endovascular treatment irrespective of adjudication method, but higher odds ratios were observed with local evaluations. Final infarct volume was similarly correlated with mRS across all 3 evaluation modalities.

Conclusions—Central adjudication of mRS is feasible, reducing interrater variability and avoiding potential problems related to lack of blinding. Our findings may have implications in the planning of future randomized acute stroke trials, especially in those including nonpharmacological interventions.

Clinical Trial Registration—URL: http://www.clinicaltrials.gov. Unique identifier: NCT01692379.

(Stroke. 2015;46:3405-3410. DOI: 10.1161/STROKEAHA.115.010909.)

Key Words: central adjudication ◼ endovascular ◼ odds ratio ◼ Rankin ◼ stroke trial
not been assessed properly in a clinical trial setting. Recently, McArthur et al published the feasibility and reliability of a video-based modality for central remote mRS adjudication in a virtual multicenter stroke trial using a group adjudication approach with 4 central assessors. Agreement between the centrally adjudicated and local evaluations was good. They also demonstrated the feasibility of using translated interviews simulating an international multicenter study. However, the use of external video-based adjudication methodology in a real-world trial has not been published yet.

To address the above mentioned concerns related to in-person evaluation, REVASCAT (Randomized Trial of Revascularization With Solitaire FR Device Versus Best Medical Therapy in the Treatment of Acute Stroke Due to Anterior Circulation Large Vessel Occlusion Presenting Within Eight Hours of Symptom Onset) was designed to include both local and central evaluation methods of the primary outcome. This secondary analysis from REVASCAT study aimed to determine feasibility of central mRS adjudication and compare different central adjudication methods (phone- and video-based). In addition, because the true disability status cannot be determined, the 3 methods of outcome adjudication were also compared with a more objective measure of cerebral impairment, core laboratory–evaluated 24-hour infarct volume on computed tomographic scan or magnetic resonance imaging.

Methods

Study Design
REVASCAT enrolled 206 patients randomized to thrombectomy with Solitaire device versus medical management alone. Eligible patients had contraindications to intravenous alteplase or had received intravenous alteplase therapy within 4.5 hours without revascularization within 30 minutes of alteplase infusion. Primary end point was distribution of mRS scores at 90 days (±14 days). Details on study protocol and main results of the study have been already published.

Modified Rankin Scale Evaluations
The primary outcome variable was evaluated twice in each patient by both local and central certified assessors. Locally, each site designated one or more mRS-certified neurologists, not involved in patient management, to evaluate the mRS score in a face-to-face visit. Local investigators were asked to follow a specific structured interview based on the Rankin Focused Assessment.

Methodology for central adjudication of mRS scores varied along REVASCAT study enrollment: during the first period of the study, an external mRS-certified nurse (M. Salvat) evaluated mRS scores by telephone call using the same structured interview as local investigators and recorded the interviews by audiotape. After 50 patients centrally adjudicated in this manner, based on a predetermined convergence to Normal distribution. First, discrepancies were evaluated globally between local and central raters, including phone- and video-based raters (in the 55 patients with both central assessments, the average score [phone+video]/2 was used). Thereafter, because of the different methodology between phone- and video-based evaluations, discrepancies of local investigators with each of central assessors were evaluated also separately. All analyses were performed excluding deaths (mRS=6). In patients evaluated by the central video-based rater, percentage of total agreement with local evaluations was also studied taking into account quality of interviews (poor, acceptable, reliable). Primary analysis in REVASCAT was based on central evaluator through video recording with local investigator’s assessment as the default modality in case the video evaluation was not available as per steering committee decision before data unblinding. Sensitivity analyses for the main outcome were also performed based on scores given by both central evaluators (sensitivity analysis I) and local investigators (sensitivity analysis II). All analyses were in the intention-to-treat population. Effect size measure was a cumulative logistic regression odds ratio (shift analysis) and 95% confidence interval, adjusted for minimization factors and intravenous alteplase use.

Correlation (Spearman) of final infarct volume with mRS scores obtained with each of the 3 mRS assessments was performed.

Results
From November 2012 to December 2014, 206 patients were enrolled in REVASCAT trial in 4 centers in Catalonia, Spain. All patients had available outcomes at 90 days performed via at least 2 methods. Excluding deaths, 171 patients had an evaluation of mRS at 90 days performed by masked local investigators at each site (in 5 patients, the local evaluation was made by phone because of patient’s impossibility to attend the hospital). During the first period of the study, 120 patients also underwent a central phone-based assessment of functional status at 90 days. In the second part of the study, 106 patients located outside the enrolling hospitals, were not aware of mRS scores given by local investigators, and introduced their assessments in an independent part of electronic case report form.

Central video rater also evaluated, in a blinded manner, quality of face-to-face interviews performed by local investigators and scored them as (1) poor (local investigator did not follow structured interview systematically, central rater had to display video clip >2× to decide on a final assessment); (2) acceptable (structured interview is used but central rater has some uncertainties in evaluation between 2 adjacent items, having to view part of the video clip at least twice); (3) reliable (structured interview is completely followed, and central rater has no hesitation in ascribing a final adjudication after reviewing video clip only once).

Final Infarct Volume Measurement
Infarct volumes at 24 hours on computed tomography or magnetic resonance imaging were adjudicated blinded to clinical data by image core laboratory. In case both computed tomography and magnetic resonance imaging were available, for uniformity sake, computed tomographic data were used. Volumes were calculated through software-generated volumetric calculations (Quantomo11) with manual adjustments when deemed necessary. In case of coexistence of hemorrhage and infarction, combined (infarction and hemorrhage) volumes were used.

Statistical Analysis
Global interrater agreement in mRS scoring between local and central raters was evaluated with weighted kappa statistics. Magnitude and direction of discrepancies (difference between central rater and local rater) were also studied. To detect possible bias regarding lack of blinding, those discrepancies were compared between both treatment allocation arms with t test because sample size was big enough to use convergence to Normal distribution. First, discrepancies were evaluated globally between local and central raters, including phone- and video-based raters (in the 55 patients with both central assessments, the average score [phone+video]/2 was used). Thereafter, because of the different methodology between phone- and video-based evaluations, discrepancies of local investigators with each of central assessors were evaluated also separately. All analyses were performed excluding deaths (mRS=6). In patients evaluated by the central video-based rater, percentage of total agreement with local evaluations was also studied taking into account quality of interviews (poor, acceptable, reliable).
were video-recorded during the face-to-face interviews at local sites, and a single central assessor evaluated mRS in video recordings. There were 55 patients who received all 3 evaluations (local, phone-based, and video-based).

Feasibility of Central Adjudication of mRS During the Trial
All patients or proxies consented to mRS evaluation by any method, anonymity was maintained in all patients, and the files containing the functional assessment (audio and video clips) were stored as back-up copies. For central phone adjudication, a notification was sent via fax to the central evaluator, including the name of the patient and a direct relative, contact phones, and the date to be called after randomization. Phone calls were made from a central office of the Catalan Stroke Program. Remote video-based assessment required the implementation of specific technology in the 4 participating centers consisting on a light laptop equipped with webcam and Internet connectivity to directly store the video clips and afterward upload them to the transmission system. The portability of the laptop facilitated that the interviews could be made in any outpatient office. Video protocol was brief, and no technical issues were encountered with respect to recording, storing, or uploading the video files. The file transfer protocol allowed secure transfer of video files to the central assessor. Following the protocol, there were no need to edit the video files, and the entire process, including set-up and transmission, took no longer than a few minutes per case.

Agreement Between Local Investigators and Central Assessors
The cross-tabulation of pair ratings by local investigators and central assessors is represented in Table 1. The percentage of total agreement (diagonal cells) was higher using video-based assessments than using phone-based assessments for all mRS scores. Globally, total agreement between local and central assessor was obtained in 62.5% of cases (95% confidence interval [CI], 53.2–71.2), \( k_w = 0.77 \) (95% CI, 0.72–0.83) using phone calls and in 86.8% of cases (95% CI, 78.8–92.6), \( k_w = 0.92 \) (95% CI, 0.88–0.96) using video recordings. In the group of 55 patients that received both central assessments, agreement between both central raters was 67.3% (95% CI, 53.3–79.3), \( k_w = 0.78 \) (95% CI, 0.69–0.88). Compared with local investigator, phone assessor gave a lower score in 22.5% and a higher score in 15% of cases; video assessor gave a lower or higher score in the same percentage of cases (6.6%) compared with local investigator.

Magnitude and direction of discrepancies in each treatment arm is represented in Table 2. Globally, mean difference in mRS scoring between local and central raters (phone- and video-based) was comparable, regardless of treatment allocation (\( P = 0.3466 \)), but differed in those patients evaluated by video recordings (\( P = 0.0075 \)). Percentage of direction of discrepancies is represented in Table 3.

Quality of Local mRS Interviews as Assessed by the Central Video-Based Assessor
Quality of local face-to-face interviews as assessed by central video evaluator were poor in 11/106 cases (10.4%), acceptable in 19/106 cases (17.9%), and reliable in 76/106 (71.7%). The higher the quality of face-to-face interviews, the better the agreement between video-based central assessor and local investigator (36.4%, 84.2%, and 92.1% for poor, acceptable, and reliable clips, respectively). Percentage of poor quality interviews was higher in the first period after implementation of video recordings compared with the last period (13.3% versus 7.2%).

Sensitivity Analyses for the Primary Outcome of REVASCAT Trial
Sensitivity analyses for primary outcome of REVASCAT trial using different adjudication methods of final mRS scores are represented in Figure. All analyses showed benefit of

Table 1. Cross-Tabulation of Pair Ratings of mRS by Local Investigators and Central Assessors

<table>
<thead>
<tr>
<th>Central Evaluation (Face-to-Face Interview), n (%)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central phone (n=120)</td>
<td>4 (80.0%)</td>
<td>7 (28.0%)</td>
<td>1 (4.0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 (20.0%)</td>
<td>11 (44.0%)</td>
<td>8 (32.0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7 (28.0%)</td>
<td>12 (48.0%)</td>
<td>5 (20.0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4 (16.0%)</td>
<td>19 (76.0%)</td>
<td>2 (13.3%)</td>
<td>2 (8.0%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 (4.0%)</td>
<td>8 (53.3%)</td>
<td>2 (8.0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5 (33.3%)</td>
<td>21 (84.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central video-recording (n=106)</td>
<td>7 (100.0%)</td>
<td>4 (20.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12 (60.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 (15.0%)</td>
<td>17 (85.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 (5.0%)</td>
<td>3 (15.0%)</td>
<td>24 (100.0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>11 (100.0%)</td>
<td>3 (12.5%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>21 (87.5%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

mRS indicates modified Rankin scale.
outcomes are desired to be reanalyzed centrally for the purposes of pooled analyses or for the purposes of conducting analyses using different end points.

Reliability of central adjudication was evaluated assessing interrater agreement of central assessors with face-to-face evaluations at local sites, the latter being generally considered the standard methodology. Agreement was good for both modalities, but it was higher with video-based adjudications than phone-based adjudications. Agreement between local and central video assessor was higher compared with CARS study, where they reported a $k_0$ of 0.80 (0.75–0.84) for adjudicated panel score versus local score at 90 days. In CARS study, there was initial disagreement from ≥1 of the 4 panel members in around 50% of evaluations, reaching unanimous agreement after discussion in majority of cases. Furthermore, in our relatively small sample size, agreement between local and video-based assessors was 100% for mRS scores 0, 3, and 4. Phone assessor gave lower scores (corresponding to lower disability states) than local investigators more frequently than video-based assessor. This may be explained by the fact that central video rater is reviewing the assessment that was performed at the same point in time and in the same environment as the local investigator, receiving exactly the same information from the patient and proxy (if present); this is not the case with phone adjudication, which was performed at different time points. In addition, for those patients who are too disabled to provide the information themselves, questions may have been answered by a different proxy. Importantly, the central video assessor is able to observe and evaluate patient global status and some abilities (eg, walking) directly while phone assessor is not. Furthermore, behavioral aspects of stroke recovery, such as neglect or anosognosia, with impact on patient’s account of their ability to function can be better assessed by direct visual assessment.

Another advantage of central video-based adjudication is that it provides quality control of face-to-face interviews performed at local sites. Although there were no video clips unable to be scored, agreement was better when local investigators followed correctly the structured interview, as mandated by the study protocol. In case an investigator was not performing mRS assessment properly as judged by the central assessor, feedback could be given to local investigators to improve training of local investigators for future interviews. Indeed, we observed a reduction in poor quality interviews across the study. Ideally, the expert feedback should be given in real time, allowing the investigators to reperform the interview following the central assessor advice. Therefore, the possibility of scheduling the outpatient visit with availability of central assessor to perform the evaluation in real time would be of great interest. This could be studied for implementation in future studies thanks to the advent of new technologies as high-speed Internet, 2-way video conferencing, and the ubiquitous use of smart phones.

### Table 3. Direction of Discrepancies Between Local and Central Raters in Each Treatment Arm

<table>
<thead>
<tr>
<th></th>
<th>Endovascular Treatment</th>
<th>Medical Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>84</td>
<td>87</td>
</tr>
<tr>
<td>Local scores lower than central, n (%)</td>
<td>12 (14.3%)</td>
<td>12 (13.8%)</td>
</tr>
<tr>
<td>Same rating, n (%)</td>
<td>60 (71.4%)</td>
<td>56 (64.4%)</td>
</tr>
<tr>
<td>Local scores higher than central, n (%)</td>
<td>12 (14.3%)</td>
<td>19 (21.8%)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>0.4346</td>
<td></td>
</tr>
</tbody>
</table>

Central scores included phone- and video-based. In those patients with both scores, the average central score [(phone+video)/2] was used.
The few observed disagreements between video assessor and local investigators are to be expected given the inherent interrater variability of this adjudication method. However, we must be alert for potential bias because of lack of blinding within endovascular stroke trials in which treatment allocation is open. Therefore, we compared disagreement between local and central raters in both treatment arms. Although not significant differences were noted in the whole cohort, in the group of patients evaluated by video, mean differences in mRS scoring varied among treatment arms and were in the direction of benefit of endovascular treatment. Although the reason for these observed discrepancies is unclear, there is a possibility that these differences may in part be explained by lack of blinding by evaluators working at recruiting centers.

The statistical analysis plan in REVASCAT study did not a priori establish which of the 3 outcome adjudication methods will be the main method used for the primary outcome. Based on the unequivocally blinded nature of the central video adjudication in conjunction with high interrater reliability rates between local adjudications and central video adjudications, before the first interim analysis, the blinded steering committee of REVASCAT decided to consider the central video analysis as primary method used for end point adjudication with local evaluation used as default method in case the former was missing. Sensitivity analyses were preplanned using different adjudication methods for mRS. These analyses reflected the effect of selecting different end point adjudication methods on trial results. Had local mRS assessments been chosen as primary outcome adjudication method, the treatment effect of endovascular treatment in REVASCAT would have been higher than reported.

Because all 3 measurement methods are prone to errors and the true disability status of each patient cannot be ascertained with certainty, we sought to validate these 3 distinct ways of obtaining the mRS against another more objective variable that is known to strongly correlate with neurological recovery, which is 24 hour infarct volume. We found similar and good degree of correlation across all evaluation modalities that did not favor any of the evaluation methods.

The main limitation of the present study is that central adjudication modalities varied along the study (phone in the first period and video in the last), and only 55 patients received both central evaluations. This fact prevented direct comparisons between the 2 central assessors in the whole cohort and hampered sensitivity analyses of primary outcome of the trial using only 1 central adjudication method. Another limitation is that being only 1 central assessor in each modality, we cannot be certain whether the findings apply to a rating pattern of particular central assessors; however, it represents also a strength point because of the elimination of interobserver variability among different central assessors.

In summary, central adjudication of mRS using video recordings is feasible and easy to implement in a stroke trial setting, improving quality of assessment of primary outcome and avoiding potential bias. In addition, it confers the advantage of permanent data recording. These results should be considered in the planning of future randomized acute stroke trials, especially in those with open treatment allocation.

Table 4. Correlation Coefficients (Spearman) Between Final Infarct Volume (FIV) and mRS Scores

<table>
<thead>
<tr>
<th></th>
<th>All Patients (n=55)*</th>
<th>Endovascular Arm (n=28)</th>
<th>Medical Arm (n=27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>0.458</td>
<td>0.331</td>
<td>0.453</td>
</tr>
<tr>
<td>Central phone</td>
<td>0.471</td>
<td>0.399</td>
<td>0.426</td>
</tr>
<tr>
<td>Central video</td>
<td>0.472</td>
<td>0.401</td>
<td>0.437</td>
</tr>
</tbody>
</table>

*For this analysis, only patients with the 3 evaluations (local, phone, and video) were selected to enable comparison of coefficients (same infarct volumes).
Sources of Funding
REVASCAT was funded by a local independent Catalan institution (Fundació Ictus Malaltia Vascular, www.fundacioictus.com/es) by means of an unrestricted grant from the manufacturer of the device (Covidien). This project has been partially supported by a grant from the Spanish Ministry of Health cofinanced by FEDER (Instituto de Salud Carlos III, RETICS-INVICTUS, RD 12/0014/008) as well as grant from the Generalitat de Catalunya (SGR 464/2014) to the GRBIO group.

Disclosures
Dr Cobo received nonfinancial research grant from Generalitat de Catalunya (research group GRBIO); modest honoraria from Fundació Ictus Malaltia Vascular; Institutional conflict of interest: Barcelona-Tech received a grant for statistical design of REVASCAT trial. Dr Dávalos received significant research grant from Covidien and modest honoraria from Silk Road (consultant); consultant/advisory board from Medtronic and Stryker Neurovascular (nonfinancial); and consultant from J&J and Neuravi (modest). The other authors report no conflicts.

References
Phone and Video-Based Modalities of Central Blinded Adjudication of Modified Rankin Scores in an Endovascular Stroke Trial

Elena López-Cancio, Mercè Salvat, Neus Cerdà, Marta Jiménez, Javier Codas, Laura Llull, Sandra Boned, Luis M. Cano, Blanca Lara, Carlos Molina, Erik Cobo, Antoni Dávalos, Tudor G. Jovin and Joaquín Serena

on behalf of REVASCAT investigators

Stroke. 2015;46:3405-3410; originally published online November 5, 2015;
doi: 10.1161/STROKEAHA.115.010909

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. 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/46/12/3405

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

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

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