É. Tollard
T.E. Darsaut
F. Bing
F. Guilbert
G. Gevry
J. Raymond

# Outcomes of Endovascular Treatments of Aneurysms: Observer Variability and Implications for Interpreting Case Series and Planning Randomized Trials

**BACKGROUND AND PURPOSE:** Angiographic results are commonly used as a surrogate marker of success of coiling of intracranial aneurysms. Inter- and intraobserver agreement in judging angiographic results remain poorly characterized. Our goal was to offer such an evaluation of a grading scale commonly used to evaluate results of endovascular treatment of aneurysms.

**MATERIALS AND METHODS:** A portfolio of 90 angiographic images from 45 patients selected from the core lab data base of a randomized trial was sent to 12 observers on 2 occasions more than 3 months apart. The variability of a 3-value grading scale used to score angiographic results and of a final judgment regarding the presence of a recurrence was studied using $\kappa$ statistics.

**RESULTS:** Ten participants responded once and 6 responded twice. Agreement was poor to moderate ($\kappa = 0.28-0.5$) for senior and junior observers judging angiographic results immediately or 12–18 months after treatment. Agreement reached a reassuring "substantial" ($\kappa = 0.62$) level, with a dichotomous presence-absence of a major recurrence, and intraobserver agreement was better in experienced core lab assessors.

**CONCLUSIONS:** There is an important variability in the assessment of angiographic outcomes of endovascular treatments, rendering comparisons between publications risky, if not invalid. A simple dichotomous judgment can be used as a surrogate outcome in randomized trials designed to assess the value of new endovascular devices.

**ABBREVIATIONS:** CI = confidence interval; HELPS = Hydrocoil Endovascular Aneurysm Occlusion and Packing Study; RA = ruptured aneurysm; RCT = randomized controlled trial; UA = unruptured aneurysm

The goal of endovascular therapy of intracranial aneurysms is the prevention of rupture (in UAs) or of rebleeding (in RAs). This statement may seem trivial, but angiographic occlusion of aneurysms has been used as a surrogate marker of a satisfactory treatment for so long that many interventionists have come to believe that the goal of treatment is "to eliminate blood flow to as much of the aneurysm as can safely be achieved."[1]

If success is defined as a fully occluded aneurysm at long-term follow-up, then success is not always achieved.[2-3] A plethora of new devices, carrying unknown risks and benefits, have been introduced in the last 10 years. They purport to improve on anatomic results. None have been reliably shown to provide the benefits claimed by manufacturers, and the FDA has rarely required more than a case series of 100 patients or less to approve these new endovascular devices. Ideally, such devices should be shown to lead to better clinical outcomes than platinum coils or surgical clipping, using clin-

ical end points. Unfortunately, such trials would be so long and so large that little progress in this field would be possible. Hence, surrogate anatomic end points are commonly used in trials to judge the value of new devices in the treatment of aneurysms.[4-6]

Once we accept that we treat images, we have to define success of treatment in terms of imaging results and costs in terms of increased morbidity at the time of treatment.[7-8] This type of comparison is at least awkward, if not invalid, but we will confine the concerns of the present work to the difficulties encountered when we want a verdict regarding success of therapy in terms of imaging results. To use imaging results as end points of clinical trials of new devices, it is paramount that we determine how accurate, repeatable, or precise our measurement tools are.

Cloft et al[1] showed that agreement between observers evaluating postcoiling angiographic results was better when grading scales had fewer responses, and that substantial to almost-perfect agreement was possible when 2 observers trained and practiced together, using a worse/not worse dichotomous classification.

A recent editorial in *Radiology* called for more stringent evaluation of variability in test interpretations using larger numbers of observers with a broad range of experience.[9] Implications for the interpretation of case series in our literature, and for the design of clinical trials, will be discussed. We will also review other grading scales that have been proposed in the endovascular literature.

## Materials and Methods

### Cases

The primary aim of this work was to evaluate the intra- and interobserver variability in adjudicating outcomes of treatment according to an ordinal scale commonly used to assess angiographic results of coiling.[2] To minimize variability due to different angiographic equipment, number and type of projections, selection of series, and selection of final images from various series, as well as to encourage the participation of multiple observers, we assembled a portfolio of selected images from selected cases that could be sent electronically to various interpreters. All images were retrieved from the data base of the core lab of HELPS.[5] The types of coils used to treat aneurysms were unknown, but in this RCT, we can surmise that approximately half the cases had been treated with platinum coils; the other half was treated with a mixture of hydrogel-coated and platinum coils. Only cases that included at least 2 comparable images from comparable angiographic series, 1 immediately following treatment and another 12–18 months later, were eligible. One author (É.T.) selected and assembled the images ($n = 90$) and the cases ($n = 45$) for the portfolio, attempting to include 10 "recurrences," 10 "stable results," and 25 cases that were "difficult to judge or unclear" in similar proportions. These proportions were chosen to 1) cover a wide spectrum of difficulties in interpretation of results, and 2) to minimize well-known paradoxes of $\kappa$ statistics.[10-11] On each page of the electronic version sent to reviewers, 1 postembolization image and 1 follow-up image were displayed side-by-side. No clinical information was provided. Observers were given the task of grading each image according to a 3-value scale (complete occlusion, residual neck, residual aneurysm), graphically displayed on each page (see Raymond et al[12]). They were also asked to make a final judgment regarding the presence of a recurrence, again according to a 3-value scale (no recurrence, minor recurrence, major recurrence) by comparing the 2 images. The definition of a major recurrence was "a saccular recurrence of a size sufficient to allow retreatment." Any other increase in the residuum was to be labeled a minor recurrence.[2]

### Observers

The portfolio was sent twice electronically, at least 3 months apart, to 12 potential participants, all interventional neuroradiologists who work in 6 different centers from 4 different countries (United States, United Kingdom, France, and Canada). There were 6 senior observers (more than 10 years of experience in interventional neuroradiology, including 3 who had previous experience as core lab assessors [2 from the same lab] and 4 junior observers (less than 5 years of experience).

There was no training of observers for this task. Apart from the graphic display of the scale, they were not provided with precise definitions of categories.

### Statistics

The interrater agreement regarding the angiographic result of the intervention in 3 categories at 2 points in time, and the final judgment regarding the evolution of 2 angiographic results (no, minor, or major recurrence) in the same patient, were measured by the generalized $\kappa$ statistic.[13] All categories, such as "fair," "moderate," or "substantial," were qualified according to Landis and Koch.[13] The same approach was used to measure the agreement of senior and junior specialists separately. The variability of the interjudge agreement was studied by calculating $\kappa$ for each possible evaluator. The intrajudge agreement of the 6 experts who responded twice was also measured by the $\kappa$ statis-

tic. Because the primary end point of the trial included only major recurrences, agreement was also analyzed in 2 categories (major recurrence, yes or no). All tests were performed with SPSS version 19 (SPSS, Chicago, Illinois). CIs were calculated using 95% CI = $\kappa \pm 1.96$ standard error of $\kappa$.

## Results

Ten participants responded once, and 6 responded twice, allowing the study of interobserver agreement in 10 participants and intraobserver agreement in 6 participants.

Results are presented in On-line Tables 1–4. There was a wide variability in absolute results, with 0%–48% of cases being judged to have a residual aneurysm on the first posttreatment evaluation. On follow-up angiography, 21%–60% of cases were judged to present a residual aneurysm. Observers judged that a major recurrence occurred in 19%–48% of cases. The generalized $\kappa$ statistic increased from "fair" (0.276) for the first evaluation to "substantial" (0.619; 0.544−0.696) for the final judgment on the presence of a major recurrence, according to the Landis and Koch categories.[12]

Senior observers were not more concordant than junior observers. The intraobserver $\kappa$ statistic for major recurrence was better for core lab experts (1.0, 0.85, 0.69) than for other evaluators (0.71, 0.65, 0.61). Clinical examples are shown in the Figure.

## Discussion

It is commonly agreed that if a lesser morbidity can be achieved with endovascular management, this is accomplished at the price of an increased frequency of incomplete eradication of the lesion at the time of treatment or of recurrences on follow-up imaging studies.[2,14-16] Incomplete occlusions and recurrences probably mean an increased risk of future hemorrhagic events,[2,14,17] but the magnitude of these risks, compared with those of retreatments, are currently impossible to estimate precisely.[17] Imaging results of treatments of aneurysms remain clinically important because many physicians will justify the use of surgical clipping, or of new endovascular devices, on the alleged risks of residual or recurrent lesion with simple platinum coiling.

Difficult questions arise when one tackles the problem of agreement regarding a treatment outcome. Some kind of skepticism or relativism is difficult to escape. Surely there must be some reality regarding the presence or absence of a residual or recurrent aneurysm. In the absence of a "gold standard," however, a truth regarding the verdict of the test seems impossible to capture. In addition, all talks about "prevalence" or "positive cases" must be oblique because these concepts are now relative to a certain observer at a certain time.

Two types of judgments have been used thus far in assessing results of endovascular coiling: 1) an absolute judgment, a degree of occlusion, reflecting the status of aneurysm occlusion at a particular time, and 2) a comparative judgment, reflecting the "stability of occlusion" in a particular patient. Our aim in the present work was to measure the variability in the interpretation of angiographic images postcoiling of aneurysms using a specific, ordinal grading scale, as well as the variability in the adjudication of a considered important (still nonclinical) outcome, the presence or absence of a recurrence.
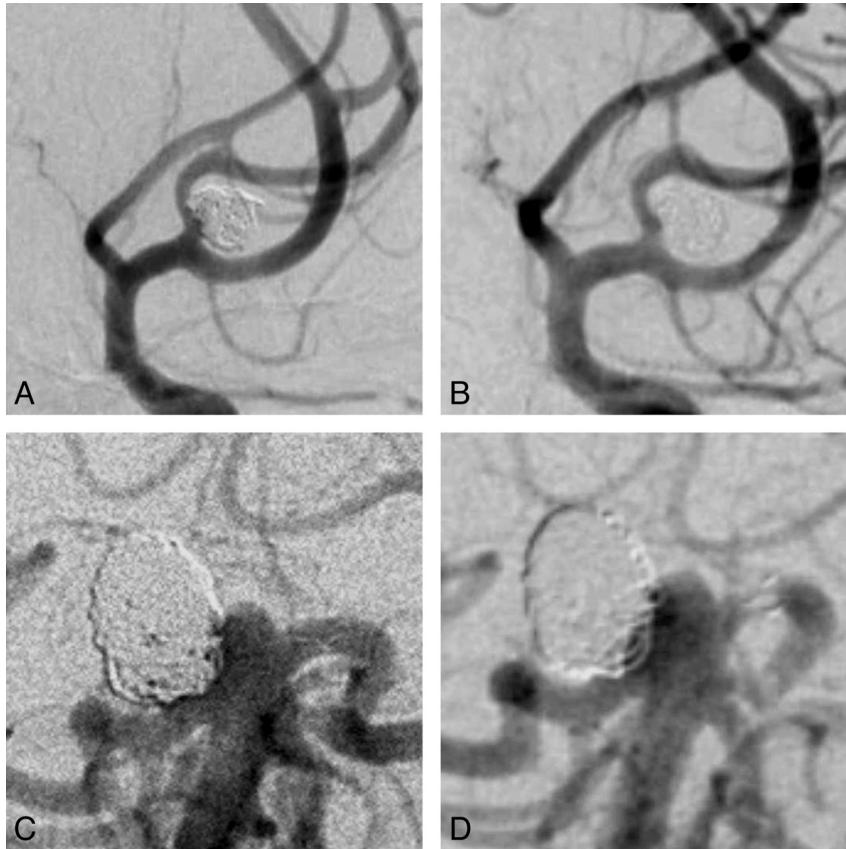
**Fig. 1** Cases: Selected images showing 1 case with perfect agreement between all observers (complete occlusion; no recurrence; *A, B*). The case illustrated in *C, D* showed maximal disagreement for both degree of occlusion and presence of a recurrence.

The present work has shown that agreement regarding an absolute judgment on the "completeness" of angiographic occlusion after coiling is, at best, only fair. Agreement improves to "moderate" with follow-up angiography, but the finding that the prevalence of residual aneurysms can still vary from 20%–60%, depending on the observer, is sufficient to invalidate any attempt at judging the value of different devices by comparing publications of case series performed in different institutions.

This disconcerting finding is basically a reminder that RCTs are an absolute necessity if valid comparisons between results of treatments are desired. The present work supports the notion that if angiographic results are to be used as a surrogate end point for clinical success of treatments, a core lab, with experienced observers, is necessary to minimize intraobserver variability. Kappa statistics also show that if anchoring of the judgment scale is difficult with a single time point, variability decreases when a second angiogram is available. Decreasing the number of categories, and judging the evolution between the 2 examinations, further decreases the variability in order to reach an acceptable level of "substantial agreement,"[13] as others have previously shown.[1] Yet if substantial intra- and interobserver agreement become possible when a core lab judges the value of 2 randomized treatment options within a trial, we must remember that the verdict of the trial is a relative judgment between the 2 options, and absolute results cannot be exported outside the trial itself for comparisons with published series.

### The Choice of Scale

There are a number of criteria that can be used to assess the value of outcome scales.[4] Numerous scales, summarized in the Table, have been proposed in the literature.[12,19-34] Some were designed for conventional angiography, others for MRA. If most use a different terminology, or use a numeric or pseudonumeric form, and if some aim to be more "objective" and precise—by multiplying the number of categories—or more intuitive, at least according to some authors, none have been reliably validated.[4] Reassuringly, most scales can be translated into the 3-value scale we have previously proposed[19] and perhaps, most important, many can be dichotomized to the more reproducible "major recurrence yes-no" judgment.

### Kappa Statistics

The $\kappa$ statistic was designed to provide a measure of agreement that takes into account the role of chance in the occurrence of concordant verdicts. The immediate consequence is the appearance of a number of "paradoxes," previously analyzed in detail by Feinstein and others.[10-11] When there is imbalance in the distribution of positive cases within the sample studied, agreement can be expected from chance alone and $\kappa$ values can be relatively low despite frequent concordance. One way to minimize these paradoxes is to assemble series of cases that include an approximately 50% "prevalence" of the judgment to be adjudicated.[10-11] This problem is enlightening, for it reveals the

**Classifications proposed by various authors to assess outcomes of coiling**

| Articles | Classification | Classes | Dichotomy Possible? |
|---|---|---|---|
| **Proposed for DSA** | | | |
| Raymond (1997)[19] | Complete | 4 | Y |
| | Dog ear | | |
| | Residual neck | | |
| | Residual aneurysm | | |
| Cognard (1999)[20] | 100% | 3 | Y |
| | 95–99% | | |
| | < 95% | | |
| Roy (2001)[12] | Complete | 3 | Y |
| | Residual neck | | |
| | Residual aneurysm | | |
| Johnston (2008)[21] | 100% = complete | 5 | No |
| | 91–99% = small residual neck | | (yes?) |
| | 70–90% = residual neck | | |
| | 1–69% = partial occlusion | | |
| | 0% = not occluded | | |
| Meyers (2010)[22] | ≥90% | 5 | Y |
| | 70–89% | | |
| | 50–69% | | |
| | 25–49% | | |
| | <25% | | |
| CLARITY (2010)[23] | A = complete occlusion | 5 | N |
| | B = subtotal occlusion with | | |
| | B1 complete neck coverage but contrast in coil mesh | | |
| | B2 incomplete neck coverage but no contrast in coil mesh | | |
| | B3 incomplete neck coverage and contrast in coil mesh | | |
| | C = incomplete occlusion = aneurysm remnant | | |
| **Proposed for MRA** | | | |
| Gönner (1998)[24] | Complete | 2 | Y |
| Anzalone (2000)[25] | Residual flow | | |
| Yamada (2004)[26] | | | |
| Boulin (2001)[27] | 100% | 3 | Y |
| | 95–99% | | |
| | <95% | | |
| Kahara (1999)[28] | No remnant | 2 | Y |
| | Remnant if cavity >1 mm (neck or body) | | |
| Leclerc (2002)[29] | Complete | 4 | N |
| | Small residual neck (1–3 mm) | | |
| | Large residual neck(>3 mm) | | |
| | Not assessable | | |
| Costalat (2006)[30] | Complete | 3 | Y |
| Gauvrit (2006)[31] | Residual neck | | |
| | Residual aneurysm | | |
| Cottier (2003)[32] | Complete | 4 | N |
| | Remnant size <3 mm | | |
| | Remnant size 4–5 mm | | |
| | Remnant size >5 mm | | |
| Majoie (2005)[33] | Remnant | 3 | N |
| | Recurrence | | |
| | Occlusion | | |
| Sprengers (2008)[34] | Adequate occlusion | 2 | Y |
| | Reopening | | |

artificial nature of our assessment of concordance. In addition to the number or proportion of "positive cases" that compose the assembly, an even more difficult question arises: What types of cases are to be included? Obviously, some selection, natural or artificial, is going to dramatically influence results. Cases with inadequate follow-up studies, or cases assessed by angiography immediately after treatment but by MRA at follow-up, could not be included in the present study, for example. As for any assessment of diagnostic tests, one is struck by how easy it would be to provide numbers that would only reflect manipulation of cases.

Here, as for assessment of accuracy of a diagnostic test, one may propose that the sample must also include a wide variety of "positive cases," perhaps organized along a certain spectrum of severity, and a variety of "negative cases," selected among cases that may have pertinence in the clinical context of interest.[10,11] In the present work, the inclusion of "unclear" cases has probably increased the difficulty in achieving agreement between observers.

A previous publication along similar lines has shown better agreement, with fewer (*n* = 2) observers, both trained and working closely together at the same center.[1] They used a

larger sample of cases ($n = 125$; 83 with follow-ups) from an observational study[35] and had access to all the angiographic series of each patient.[1]

We have only provided confidence intervals for $\kappa$ in On-line Table 4; the variance can become unreliable as sample size decreases or as $\kappa$ approaches unity, and a bootstrap resampling methodology for each comparison appeared unreasonable.[36] A quick glance at the Tables suffices to capture the extent of the variability in concordance between observers.

### Other Sources of Variability in Assessing Angiographic Outcomes of Trials

Precision in outcome assessment in clinical trials is affected by 3 main sources of "random error": observer, subject, and instrument variability. Unless the study is restricted to aneurysms of a certain size or location, or to centers using similar methods and equipment, little can be done to limit subject or instrument variability. Upon reflection, this type of variability is intrinsic to the phenomenon under study, and while it will, by necessity, impose methodologic adjustments, such as increasing the number of patients to be recruited in the study to show a difference between 2 groups, any restriction designed to artificially increase precision will impact on the validity of the results. Ten of 12 potential observers responded once and only 6 responded twice. While we can only speculate what the results would have been had missing responses been available for inclusion, this problem also raises the issue of how artificial this type of assessment is. How seriously observers worked to come to verdicts can always be questioned, and the context of assessment is certainly different from a typical clinical or core lab context.

The variability we observed in judging the extent of angiographic occlusion of treated aneurysms, and the presence of a recurrence at follow-up, was probably underestimated by the method we chose because there are many other sources of discrepancies in a core lab context: There are more images, from various series, using various projections and diverse equipment from various centers throughout the world.

Legitimate strategies to enhance precision include standardization of angiographic projections and techniques, using an operations manual, refining criteria defining the score classes, and training (and sometimes certifying) the observers.

Repeating the measurements by a number of observers, with resolution of discrepancies by consensus, can succeed in achieving a precision that is totally artificial, but that can be tailored to fit the immediate purpose of the study. It is unclear, however, if such verdicts are more valid.[9]

A final note on a compounding factor that will affect future trials and that can multiply the difficulties in minimizing variability is the increasing use of various noninvasive imaging strategies, often imposing a comparison between different imaging modalities at different time points in the same patient, a problem that was evaded in the present work.[37] Needless to say, the results we present only apply to patients followed by conventional angiography. It is unclear what to do if patients are to be followed by MR angiography. Perhaps this experience, along with that of others,[1] suggests that if we are to start all over with grading results according to noninvasive imaging, a simple classification scheme (presence or absence of a

large residuum; presence of absence of a major recurrence) would be the most reproducible.

## Conclusions

Agreement between observers adjudicating angiographic results of coiling is moderate at best. "Substantial" agreement, using a dichotomous verdict regarding the presence or absence of a major recurrence, through using 2 different angiographic studies, can be reached by an experienced core lab within the context of a randomized trial.

## References

1. Cloft HJ, Kaufmann T, Kallmes DF. **Observer agreement in the assessment of endovascular aneurysm therapy and aneurysm recurrence.** *AJNR Am J Neuroradiol* 2007;28:497–500
2. Raymond J, Guilbert F, Weill A, et al. **Long-term angiographic recurrences after selective endovascular treatment of aneurysms with detachable coils.** *Stroke* 2003;34:1398–403
3. Piotin M, Spelle L, Mounayer C, et al. **Intracranial aneurysms: treatment with bare platinum coils/aneurysm packing, complex coils, and angiographic recurrence.** *Radiology* 2007;243:500–08
4. Raymond J, White PM, Molyneux AJ. **Scales, agreement, outcome measures, and progress in aneurysm therapy.** *AJNR Am J Neuroradiol* 2007;28:501–02
5. White PM, Lewis SC, Gholkar A, et al. **Hydrogel-coated coils versus bare platinum coils for the endovascular treatment of intracranial aneurysms (HELPS): a randomised controlled trial.** *Lancet* 2011;377:1655–62
6. Molyneux AJ, Coley S, Sneade M, et al. **Cerecyte Coil Trial: clinical outcome of endovascular coiling in patients with ruptured and unruptured intracranial aneurysms treated with Cerecyte coils compared with bare platinum coils. Results of a prospective randomised trial.** *American Society of Neuroradiology Annual Meeting,* Boston, Massachusetts; May 15–20, 2010
7. Roy D, White PM, Fiorella D, et al. **A randomized trial comparing platinum and hydrogel-coated coils in patients prone to recurrence after endovascular treatment (the PRET trial).** *Interven Neuroradiol* 2008;14:73–84
8. Darsaut TE, Raymond J, for the STAT collaborative group. **The design of the Stenting in the Treatment of Aneurysms (STAT) trial.** *J NeuroIntervent Surg*; epub 2011, June 23. doi:10.1136/neurintsurg-2011–010065
9. Bankier AA, Levine D, Halpern EF, et al. **Consensus interpretation in imaging research: is there a better way?** *Radiology* 2010;257:14–17
10. Feinstein AR, Cicchetti DV. **High agreement but low kappa: I. The problems of two paradoxes.** *J Clin Epidemiol* 1990;43:543–49
11. Feinstein AR. **Misguided efforts and future challenges for research on "diagnostic tests."** *J Epidemiol Community Health* 2002;56:330–32
12. Roy D, Milot G, Raymond J. **Endovascular treatment of unruptured aneurysms.** *Stroke* 2001;32:1998–2004
13. Landis JR, Koch GG. **The measurement of observer agreement for categorical data.** *Biometrics* 1977;33:159–74
14. Campi A, Ramzi N, Molyneux AJ, et al. **Retreatment of ruptured cerebral aneurysms in patients randomized by coiling or clipping in the International Subarachnoid Aneurysm Trial (ISAT).** *Stroke* 2007;38:1538–44
15. Molyneux AJ, **International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group. International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial.** *The Lancet* 2002;360:1267–74
16. Higashida RT, Lahue BJ, Torbey MT, et al. **Treatment of unruptured intracranial aneurysms: a nationwide assessment of effectiveness.** *AJNR Am J Neuroradiol* 2007;28:146–51
17. Raymond J, Roy D. **Safety and efficacy of endovascular treatment of acutely ruptured aneurysms.** *Neurosurgery* 1997;41:1235–45, discussion 1245–46
18. Raymond J, Darsaut TE. **An approach to the problem of recurrent aneurysms following endovascular coiling.** *J NeuroIntervent Surg*; epub 2011, February 25. doi:10.1136/ jnis.2011.004788
19. Raymond J, Roy D, Bojanowski M, et al. **Endovascular treatment of acutely ruptured aneurysms of the basilar bifurcation.** *J. Neurosurg* 1997;86:211-19.
20. Cognard C, Weill A, Spelle L, et al. **Long-term angiographic follow-up of 169 intracranial berry aneurysms occluded with detachable coils.** *Radiology* 1999;212:348–56
21. Johnston SC, Dowd CF, Higashida RT, et al. **Predictors of rehemorrhage after**

treatment of ruptured intracranial aneurysms: the Cerebral Aneurysm Rerupture After Treatment (CARAT) study. *Stroke* 2008;39:120–25

22. Meyers PM, Schumacher HC, Higashida RT, et al. **Reporting standards for endovascular repair of saccular intracranial cerebral aneurysms.** *AJNR Am J Neuroradiol* 2010;31:E12–24

23. Pierot L, Cognard C, Anxionnat R, et al. **Ruptured intracranial aneurysms: factors affecting the rate and outcome of endovascular treatment complications in a series of 782 patients (CLARITY study).** *Radiology* 2010;256:916–23

24. Gönner F, Heid O, Remonda L, et al. **MR angiography with ultrashort echo time in cerebral aneurysms treated with Guglielmi detachable coils.** *AJNR Am J Neuroradiol* 1998;19:1324–28

25. Anzalone N, Righi C, Simionato F, et al. **Three-dimensional time-of-flight MR angiography in the evaluation of intracranial aneurysms treated with Guglielmi detachable coils.** *AJNR Am J Neuroradiol* 2000;21:746–52

26. Yamada N, Hayashi K, Murao K, et al. **Time-of-flight MR angiography targeted to coiled intracranial aneurysms is more sensitive to residual flow than is digital subtraction angiography.** *AJNR Am J Neuroradiol* 2004;25:1154–57

27. Boulin A, Pierot L. **Follow-up of intracranial aneurysms treated with detachable coils: comparison of gadolinium-enhanced 3D time-of-flight MR angiography and digital subtraction angiography.** *Radiology* 2001;219:108–13

28. Kähärä VJ, Seppänen SK, Ryymin PS, et al. **MR angiography with three-dimensional time-of-flight and targeted maximum-intensity-projection reconstructions in the follow-up of intracranial aneurysms embolized with Guglielmi detachable coils.** *AJNR Am J Neuroradiol* 1999;20:1470–75

29. Leclerc X, Navez JF, Gauvrit JY, et al. **Aneurysms of the anterior communicating artery treated with Guglielmi detachable coils: follow-up with contrast-enhanced MR angiography.** *AJNR Am J Neuroradiol* 2002;23:1121–27

30. Costalat V, Lebars E, Sarry L, et al. **In vitro evaluation of 2D-digital subtraction angiography versus 3D-time-of-flight in assessment of intracranial cerebral aneurysm filling after endovascular therapy.** *AJNR Am J Neuroradiol* 2006;27:177–84

31. Gauvrit JY, Leclerc X, Caron S, et al. **Intracranial aneurysms treated with Guglielmi detachable coils: imaging follow-up with contrast-enhanced MR angiography.** *Stroke* 2006;37:1033–37

32. Cottier JP, Bleuzen-Couthon A, Gallas S, et al. **Intracranial aneurysms treated with Guglielmi detachable coils: is contrast material necessary in the follow-up with 3D time-of-flight MR angiography?** *AJNR Am J Neuroradiol* 2003;24:1797–803

33. Majoie CB, Sprengers ME, van Rooij WJ, et al. **MR angiography at 3T versus digital subtraction angiography in the follow-up of intracranial aneurysms treated with detachable coils.** *AJNR Am J Neuroradiol* 2005;26:1349–56

34. Sprengers ME, Schaafsma J, van Rooij WJ, et al. **Stability of intracranial aneurysms adequately occluded 6 months after coiling: a 3T MR angiography multicenter long-term follow-up study.** *AJNR Am J Neuroradiol* 2008;29:1768–74

35. Cloft HJ, HEAL Investigators. **HydroCoil for Endovascular Aneurysm Occlusion (HEAL) study: 3–6 month angiographic follow-up results.** *AJNR Am J Neuroradiol* 2007;28:152–54

36. Vierkant, RA. **A SAS macro for calculating bootstrapped confidence intervals about a kappa coefficient.** Marshfield Medical Research Foundation, Marshfield, Wisconsin. Accessible online at http://www2.sas.com/proceedings/sugi22/STATS/PAPER295.PDF

37. Raymond J. **Counterpoint: has the last word been said?** *AJNR Am J Neuroradiol* 2009;30:1649–52