Neuroradiology
https://doi.org/10.1007/s00234-021-02676-0

INTERVENTIONAL NEURORADIOLOGY

Angiographic results of surgical or endovascular treatment of intracranial aneurysms: a systematic review and inter-observer reliability study

Anass Benomar[1] · Behzad Farzin[1] · David Volders[2] · Guylaine Gevry[1] · Justine Zehr[3] · Robert Fahed[4] · William Boisseau[1] · Jean-Christophe Gentric[5] · Elsa Magro[6] · Lorena Nico[7] · Daniel Roy[1] · Alain Weill[1] · Charbel Mounayer[8] · François Guilbert[1] · Laurent Létourneau-Guillon[1] · Gregory Jacquin[9] · Chiraz Chaalala[10] · Marc Kotowski[11] · Thanh N. Nguyen[12] · David Kallmes[13] · Phil White[14] · Tim E. Darsaut[15] · Jean Raymond[1]

Received: 5 November 2020 / Accepted: 15 February 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Purpose Results of surgical or endovascular treatment of intracranial aneurysms are often assessed using angiography. A reliable method to report results irrespective of treatment modality is needed to enable comparisons. Our goals were to systematically review existing classification systems, and to propose a 3-point classification applicable to both treatments and assess its reliability.

Methods We conducted two systematic reviews on classification systems of angiographic results after clipping or coiling to select a simple 3-category scale that could apply to both treatments. We then circulated an electronic portfolio of angiograms of clipped (n=30) or coiled (n=30) aneurysms, and asked raters to evaluate the degree of occlusion using this scale. Raters were also asked to choose an appropriate follow-up management for each patient based on the degree of occlusion. Agreement was assessed using Krippendorff's α statistics (α_K), and relationship between occlusion grade and clinical management was analyzed using Fisher's exact and Cramer's V tests.

Results The systematic reviews found 70 different grading scales with heterogeneous reliability (kappa values from 0.12 to 1.00). The 60-patient portfolio was independently evaluated by 19 raters of diverse backgrounds (neurosurgery, radiology, and neurology) and experience. There was substantial agreement (α_K=0.76, 95%CI, 0.67–0.83) between raters, regardless of background, experience, or treatment used. Intra-rater agreement ranged from moderate to almost perfect. A strong relationship was found between angiographic grades and management decisions (Cramer's V: 0.80±0.12).

Conclusion A simple 3-point scale demonstrated sufficient reliability to be used in reporting aneurysm treatments or in evaluating treatment results in comparative randomized trials.

Keywords Inter-rater reliability · Intracranial aneurysms · Classification system · Digital subtraction angiography

Introduction

The main clinical goal of surgical or endovascular treatments of intracranial aneurysms is to prevent future aneurysmal hemorrhage [1]. However, clinicians need a means to judge whether treatment has been successful before rupture occurs. The most commonly used clinical method to measure initial treatment success is angiography [2]. Angiographic results are also commonly used as a surrogate outcome to evaluate endovascular innovations in case series or to compare treatments in randomized trials [3–8]. Angiographic results are also a focal point in the debate that opposes endovascular and surgical treatment. Although endovascular treatment of ruptured aneurysms was shown to lead to better 1-year clinical outcomes in the ISAT trial, it has been associated with suboptimal aneurysm occlusions, long-term recurrences, and a small but still greater risk of rebleeding as compared to surgical

✉ Jean Raymond
jean.raymond@umontreal.ca

Extended author information available on the last page of the article

Published online: 24 February 2021

clipping [9, 10]. For unruptured aneurysms, the purported better long-term protection against rupture that surgical clipping may provide is related to better initial angiographic results, but this has never been proven [11]. A randomized assessment of the angiographic results of the two treatments has yet to be completed [12–14]. To properly compare the angiographic results of both clipping and coiling requires a standardized, reliable classification, applicable to both treatments [15]. Many classification systems have been published, but most are dedicated to one treatment or the other. Previous studies reported 21 different angiographic grading scales for aneurysm assessment following coiling, and 37 different nomenclatures of angiographic results after surgical clipping, but only a few scales have been tested, and even fewer shown to be reliable [11, 16–18]. A frequently used classification system of the angiographic results of coiling initially included 4 categories, but was later simplified to 3 categories [19–21]. The system has also been adapted to post-clipping angiographic results [11]. There is currently no classification system designed to apply to both surgical and endovascular treatments.

We sought to systematically review all of the published classification systems for angiographic results and their reported reliabilities. Our next aim was to precisely define a simple grading scale applicable to both coiling and clipping. We then performed an inter and intra-rater agreement study to assess the reliability of this grading scale in adjudicating the results of both treatments and tried to estimate the clinical significance of the various scores in terms of management decisions.

Material and methods

This paper was written in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [22] and the PRISMA guidelines for systematic reviews [23], and approval was obtained from our institutional review board.

Systematic review

Two electronic searches, one for the angiographic results of each treatment modality, were performed using CINAHL, EBM, COCHRANE, EMBASE, and MEDLINE databases with no starting date specification and included publications in English and French up to June 25, 2019. (Online resources 1, 2). One author (BF) defined and tested the search strategy and extracted data from all articles. Articles were reviewed in detail by two authors (BF, AB), and discrepancies were resolved in consensus with a senior author (JR). Clinical studies of adult patients that proposed a classification system to describe the degree of occlusion of intracranial aneurysms after coiling or clipping were included, with special attention paid to studies reporting inter-observer reliability.

Case selection

We selected and assembled an electronic portfolio of catheter angiograms of 60 patients with aneurysms treated by clipping (n=30) or coiling (n=30). The number of cases per treatment group was chosen to be >24, to achieve sufficient power to assess reliability for each treatment modality according to recommendations [24, 25]. Cases were selected to include an approximately balanced number of presumed completely occluded, residual aneurysms and "gray zone" cases in a 1:1:1 ratio to reduce the chance of kappa paradoxes [26, 27]. Seventeen aneurysms (28%; 10 coiled and 7 clipped) were actually retreated. To facilitate the participation of various interpreters from multiple centers and countries, for each case, a video was prepared, which scrolled slowly through an angiographic series in the projection that best showed the post-treatment residuum, when present. Cases were patients with ruptured or unruptured aneurysms treated from January 1999 to January 2019 at one tertiary care center (Centre Hospitalier de l'Université de Montréal). Patient and aneurysm characteristics included in the portfolio are presented in Table 1.

Classification system

Three senior authors (2 interventional neuroradiologists and 1 neurosurgeon) reviewed all of the published scales and multiple clinical cases in a consensus session to propose a classification with a few categories applicable to surgical and endovascular results, as assessed by any angiographic imaging modality (DSA, CTA or MRA). They agreed on a 3-item grading scale (based on the most frequently used endovascular classification [20, 21]) but modified slightly in order to improve the reliability of the judgment between residual necks

Table 1 True characteristics* of patients and aneurysms included in the portfolio

Characteristics	Clipped (n=30)	Coiled (n=30)	Total (n=60)
Male	12 (40%)	9 (30%)	21 (35%)
Age (yr)	50.5 ± 12.4	48.3 ± 9.8	49.4 ± 11.1
Ruptured aneurysms*	16 (53%)	20 (67%)	36 (60%)
Retreated aneurysms	7 (23%)	10 (33%)	17 (28%)
Anterior circulation	27 (90%)	24 (80%)	51 (85%)
Aneurysm size (mm)			
<4 mm	6 (20%)	2 (7%)	8 (13%)
4-10 mm	17 (57%)	19 (63%)	36 (60%)
> 10 mm	7 (23%)	9 (30%)	16 (27%)
Wide neck (>4 mm)	9 (33%)	13 (43%)	22 (39%)

Data displayed as *n* (%), or mean ± SD

*In the portfolio, all aneurysms were considered as having been ruptured in a 65-year-old patient

and aneurysms. Grade 1 represented complete aneurysm occlusion; Grade 2, a residual neck (less than 2 mm using visual estimation [eyeballing] only); and Grade 3, a residual aneurysm (Fig. 1). The 2-mm cut-off value defining grade 2 as opposed to grade 3 results was inspired from a previously validated post-clipping classification system [11].

Raters

Twenty-nine clinicians were invited to participate; 19 raters (9 interventional neuroradiologists, 4 neurosurgeons, 4 interventional neurologists, and 2 diagnostic neuroradiologists) from 5 different countries responded (66%). Nine raters were considered senior (more than 10 years of practice), and 4 had previously served as reviewers in a core lab (Online Resource 3). Fifteen raters (79%) agreed to review the same cases again in a permutated order at least two months apart for the intra-rater reliability portion of the study.

Agreement study

The raters, who were not formally trained to use the classification, were asked to independently classify each of the 60 cases using the 3-grade system described above. To estimate the clinical significance of the various grades, raters were also asked, for each case, to choose the best follow-up management for that patient, assuming the angiographic results in question concerned a 65-year-old patient with no other medical problems who had a ruptured aneurysm and had attained a good clinical outcome. Available choices were the following: "Follow-up imaging in 3 to 5 years (or none at all);" "close follow-up (6-18 months);" "immediate retreatment by endovascular means;" or "immediate retreatment by surgical means." The last two choices were then combined as "immediate retreatment (either by clipping or coiling)" for analyses.

Statistical analysis

Agreement between raters for the grading scale and for the clinical management question was estimated using Krippendorff's α (α_K) statistics. Calculations were performed using R 3.5.3 (R Core Team, Vienna, Austria), SPSS 25.0 (SPSS, Chicago, Illinois) and Stata 16.0 (StataCorp, College station, Texas), and 95% confidence intervals estimated using 1000 bootstrap iterations. Interpretation of α_K values was according to Landis and Koch [28]. Comparisons of proportions of ratings between relevant aneurysm and rater subgroups, as well as the strength of association between the raters' classification and the corresponding management of the patient, were evaluated using Fisher's exact test followed by Cramer's V test.

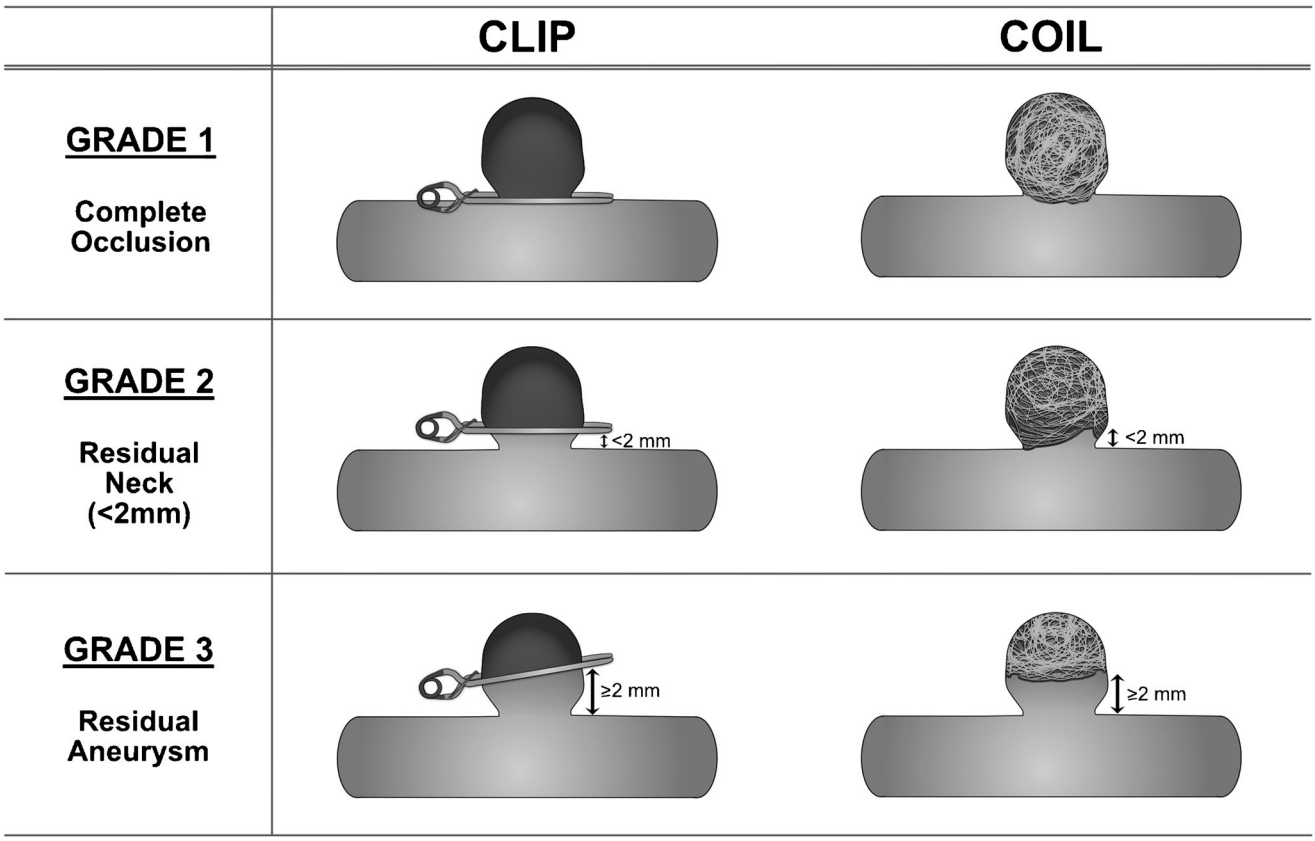

Fig. 1 Proposed grading scale to evaluate the angiographic outcome of aneurysms treated with surgical or endovascular management

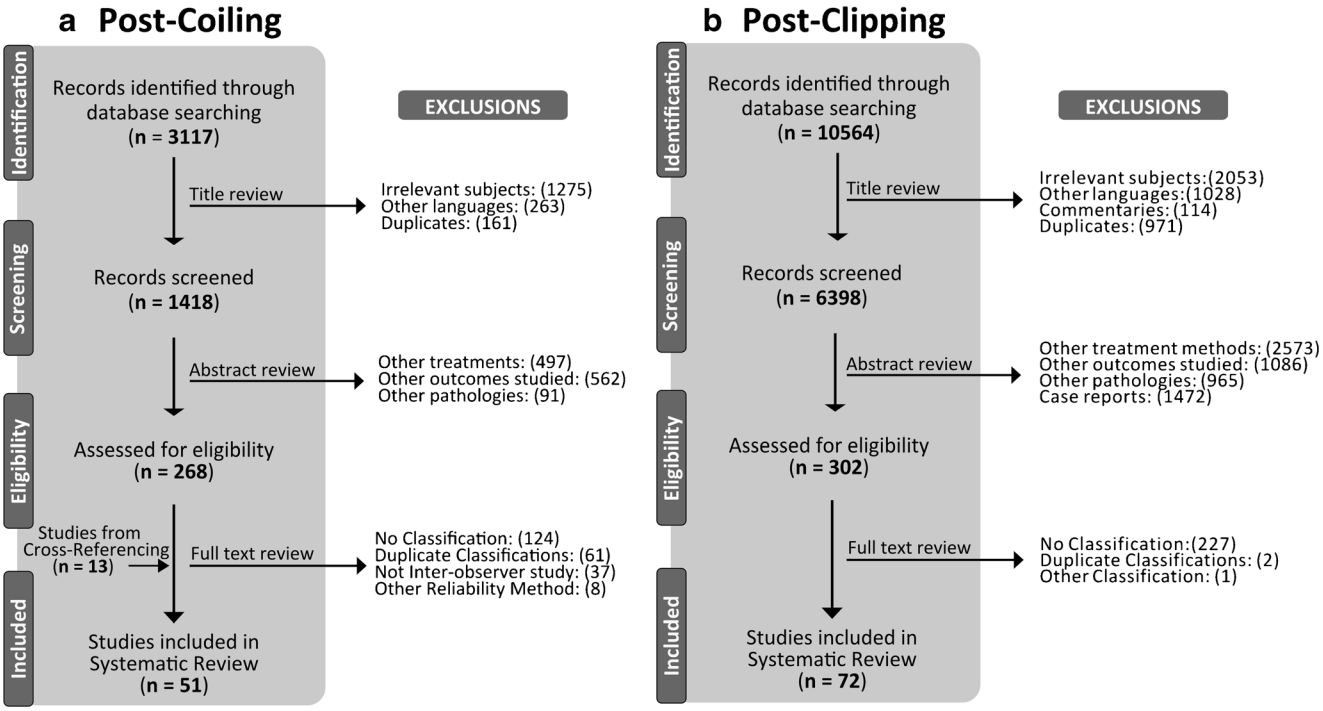

Fig. 2 Flowcharts for the systematic reviews of angiographic results after aneurysm treatment with coiling and clipping

Results

Systematic reviews

Fifty-one studies reporting 26 different grading scales (with 2-6 categories) to classify the angiographic results of endovascular coiling were identified (Fig. 2a). Inter-rater reliability was reported in 27 studies, with Kappa values ranging from 0.12 to 1.00. Twenty-three of the studies had only 2 raters, one had 3, and the other two each had 10 and 15 raters. The sample size varied from 20 to 434 patients (Online Resources 4 and 10).

The systematic review of the angiographic results of surgery yielded 72 studies, which proposed 44 different grading scales using 2-6 categories (Fig. 2b). Inter-rater reliability was reported in 9 studies, with Kappa values ranging from 0.17 to 0.90. Seven of 9 studies had only 2 raters, while the two others had 3 and 10 raters. The sample size ranged from 22 to 84 patients (Online Resources 5 and 10). The reliability of the same classification system, as used by the same raters, on the results of both clipping and coiling has never been studied.

Reliability study

Inter-rater agreement for all patients and raters was substantial when using 3 categories (α_K=0.76; 95%CI, 0.67–0.83), and remained so after dichotomization (α_K=0.74; 95%CI, 0.63–0.83). Agreement ranged from substantial to almost perfect for all subgroups of raters. Senior observers (α_K=0.77; 95%CI, 0.68–0.83) and core lab readers (α_K=0.83; 95%CI, 0.74–0.89) were not significantly more reliable than juniors (α_K=0.75; 95%CI, 0.65–0.82). Overall agreement was substantial for

Table 2 Inter-rater agreement for the classification (3-grade and dichotomized) according to aneurysm treatment, years of experience, and training background

	Inter-rater agreement (n_{Raters}=19)	
	3-grade (1 vs. 2 vs. 3)	Dichotomized (1+2 vs. 3)
Overall	0.76 (0.67–0.83)	0.74 (0.63–0.83)
Treatment		
Surgical	0.75 (0.57–0.85)	0.79 (0.63–0.90)
Endovascular	0.76 (0.65–0.84)	0.69 (0.52–0.82)
Experience		
Junior (0-10 yrs)	0.75 (0.65–0.82)	0.73 (0.62–0.82)
Senior (>10 yrs)	0.77 (0.68–0.83)	0.73 (0.62–0.82)
Background		
Core Lab	0.83 (0.74–0.89)	0.83 (0.71–0.94)
INR	0.75 (0.65–0.82)	0.74 (0.63–0.84)
DNR	0.74 (0.56–0.88)	0.72 (0.48–0.92)
NSX	0.78 (0.66–0.85)	0.72 (0.58–0.84)
INL	0.83 (0.72–0.89)	0.75 (0.60–0.87)

Data displayed as Krippendorff's α (α_K) (95% Confidence Interval) for ordinal 3-grade and dichotomized classification. Classification grades: 1, Complete occlusion; 2, Residual neck (<2 mm); 3, Residual aneurysm

INR interventional neuroradiology, *DNR* diagnostic neuroradiology, *NSX* neurosurgery, *INL* interventional neurology

the clipped (α_K=0.75; 95%CI, 0.57–0.85) and the coiled (α_K=0.76; 95%CI, 0.65–0.84) aneurysm subgroups (Table 2).

Intra-rater agreement ranged from substantial (α_K=0.78; 95%CI, 0.61–0.90) to almost perfect (α_K=0.95; 95%CI, 0.89–0.99), with no major differences between the reliability of judgments made by seniors or juniors, nor between raters with different training backgrounds, for both clipped and coiled aneurysms (Online Resource 6).

Clinical management

When the choice of follow-up clinical management was dichotomized into whether to perform follow-up imaging or retreat the aneurysm, perfect agreement existed among all 19 raters for 30/60 cases (50%). Choices for retreatment (n=323 choices out of 1130) were more frequently by endovascular (84%) than surgical means (16%).

Inter-rater agreement for the 3 choices of follow-up management was substantial (α_K=0.64; 95%CI, 0.52–0.72 with 3 choices; α_K =0.71; 95%CI, 0.60–0.79 after dichotomization into follow-up or retreatment). Agreement was similar for clipped (α_K=0.65; 95%CI, 0.45–0.76) and coiled aneurysms (α_K=0.58; 95%CI, 0.40–0.70), for senior raters (α_K=0.69; 95%CI, 0.57–0.78) and junior raters (α_K=0.58; 95%CI, 0.48–0.66), and between subspecialties. Core lab readers had almost perfect agreement (α_K=0.80; 95%CI, 0.70–0.88) (Table 3). The mean overall intra-rater agreement regarding follow-up was substantial (α_K=0.77±0.09), and again did not vary according to years of experience or training background (Online Resource 7).

Table 3 Inter-rater agreement for the clinical follow-up management (3-grade and dichotomized) according to aneurysm treatment, years of experience, and training background

	Inter-rater agreement (n_{Raters}=19)	
	3-grade (1 vs. 2 vs. 3)	Dichotomized (1+2 vs. 3)
Overall	0.64 (0.52–0.72)	0.71 (0.60–0.79)
Treatment		
Surgical	0.65 (0.45–0.76)	0.75 (0.61–0.86)
Endovascular	0.58 (0.40–0.70)	0.65 (0.47–0.79)
Experience		
Junior (0-10 yrs)	0.58 (0.48–0.66)	0.63 (0.52–0.73)
Senior (>10 yrs)	0.69 (0.57–0.78)	0.77 (0.65–0.87)
Background		
Core Lab	0.80 (0.70–0.88)	0.87 (0.75–0.96)
INR	0.65 (0.51–0.75)	0.76 (0.65–0.87)
DNR	0.57 (0.32–0.74)	0.68 (0.43–0.88)
NSX	0.68 (0.56–0.76)	0.61 (0.45–0.74)
INL	0.54 (0.38–0.66)	0.67 (0.50–0.82)

Data displayed as Krippendorff's α (α_K) (95% Confidence Interval) for ordinal 3-grade and dichotomized clinical management. Clinical management grades: 1: Delayed FU (3-5 years, or none at all); 2: Close FU (6-18 months); 3: Retreatment (endovascular or surgical)

INR interventional neuroradiology, *DNR* diagnostic neuroradiology, *NSX* neurosurgery, *INL* interventional neurology

Relationship between angiographic results and clinical management

For the theoretical ruptured aneurysm patient, raters had a strong tendency to opt for delayed follow-up with grade 1 (complete) occlusions, close follow-up for grade 2, and immediate retreatment for grade 3 aneurysm remnants (Fig. 3). However, choices were not equivalent for clipped and coiled aneurysms: clipped grade 1 aneurysms were more likely to be managed with delayed follow-up, while coiled grade 1 aneurysms were more closely followed (p<0.001) (Online Resource 8, Figure 8A). This asymmetry was also seen for grade 2 aneurysm remnants (p=0.03). Recommended management for grade 3 aneurysm residuals did not differ according to the initial type of treatment (p=0.06). Rater training background did not seem to make a difference. Individual raters all had a significant association (p<0.001) between the choice of grade of residual (i.e., complete occlusion, residual neck less than 2 mm, or residual aneurysm) and clinical management (delayed follow-up, close follow-up, or immediate retreatment). This correlation was strong for all raters (mean Cramer's V of 0.80±0.12) (Online Resource 8, Table 8A). As for the angiographic results of aneurysms which were actually retreated in real life (n=17), the vast majority of raters graded them as residual aneurysms (grade 3) and opted for a retreatment (Online Resource 9).

Discussion

The two systematic reviews showed substantial variability in the reporting and classifications of angiographic results of aneurysm treatments. Scales were specifically applied to one type of treatment using one imaging modality. Most grading scales had not been evaluated for reliability; when they were, the number of raters was generally too small (2 or 3) for reliability to be generalizable. The end result is that we found that the angiographic results of aneurysm treatments published in our literature cannot be validly compared.

The classification we propose proved to be reliable and reproducible among various clinicians of diverse backgrounds and expertise regarding patients treated with clipping or coiling. This result contrasts with previous studies where reliability could only achieve a substantial level when classification systems were dichotomized or used by expert core lab raters [16, 18]. This better reproducibility was obtained by defining residual necks in a more precise fashion (<2 mm). This 2 mm threshold was not meant to be precisely measured. A larger value (such as 3 mm) was felt to be too permissive, while 1 mm

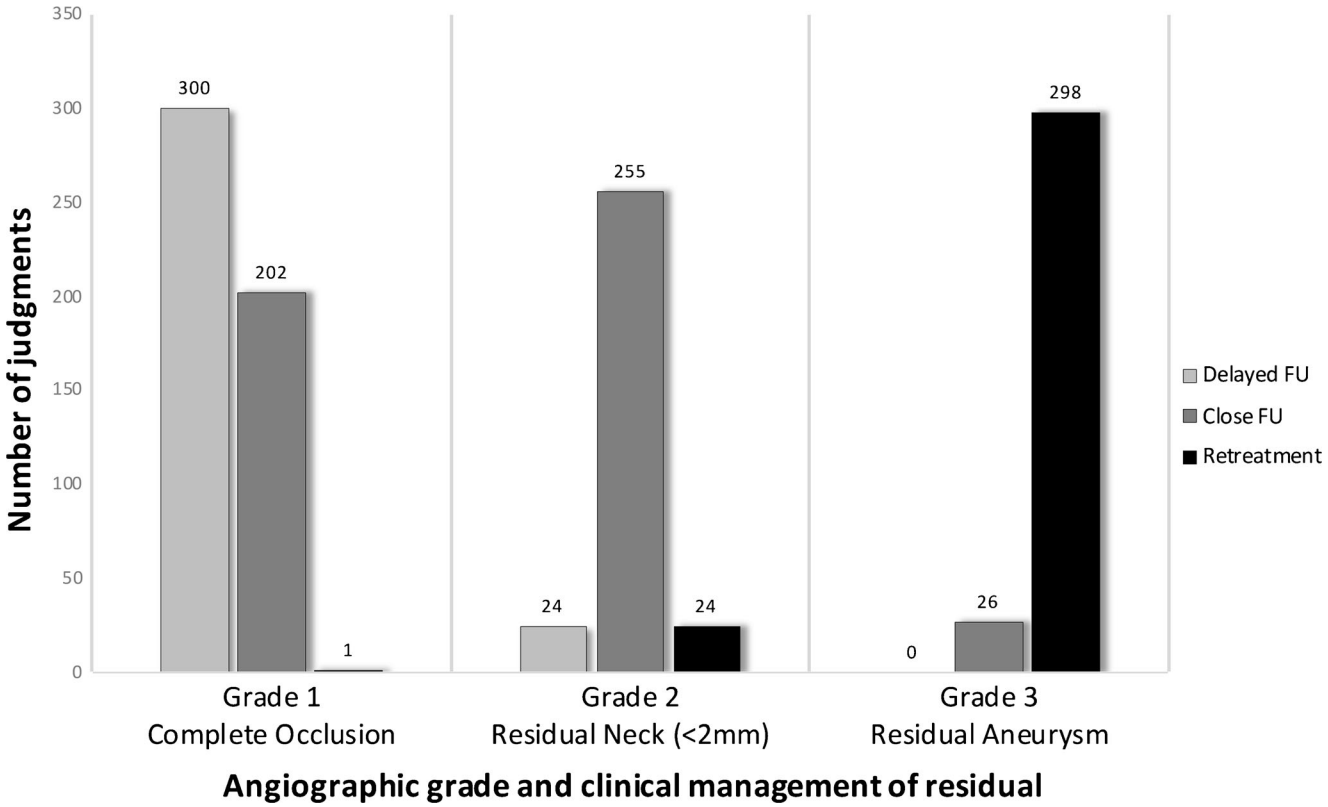

Fig. 3 Total number of follow-up management choices (n=1130) per chosen grade of aneurysm residuum. A judgment is defined as a rater's choice of grade and clinical follow-up management for a selected treated aneurysm. *FU* follow-up

would be difficult to apply in practice given the limitations of vascular imaging technologies in the presence of metallic foreign bodies such as clips or coils. This arbitrary cut-off may be intuitively problematic for small aneurysms (<4 mm for example). In addition, on an individual basis, we do not expect clinicians to evaluate or manage a 3 mm residuum the same way for initial aneurysms that measure 4 or 24 mm in diameter. This shortcoming may be an acceptable consequence of a classification system that is repeatable for most patients treated with clipping or coiling when rated by observers of various backgrounds and expertise. The merit of this classification is that it could allow comparisons between surgical clipping and endovascular coiling in randomized trials, and also permit comparisons being made between future publications.

The 3 categories of the proposed scale were shown to be clinically meaningful, as there existed a strong correlation between the grade of occlusion and clinical follow-up management selected by each rater, whether they had a surgical or endovascular training background. However, the clinical meaning of a complete occlusion (grade 1) or a residual neck (grade 2) differed depending on whether the patient had been treated with surgery or with coiling. This finding may be in accordance with the widely held beliefs of many practitioners, but could be problematic if angiographic results are to serve as endpoints for randomized trials that compare surgical and endovascular treatments. This contrast between clipped and coiled patients was not true for residual aneurysms (grade 3), suggesting that this grade may be the appropriate cut-off for angiographic results to serve as a surrogate outcome measure for comparing clipping to coiling [4, 12, 14]. This concept is further reinforced by the high agreement between raters for cases which were actually retreated in real life, the vast majority of which were rated to be residual aneurysms (grade 3) with retreatment recommended (Online Resource 9). This finding seems to confirm a recent publication linking angiographic residual aneurysms with clinical outcomes such as retreatments or rebleeding [29].

Our study has several limitations. First, although we tried to include a diversity of cases representative of real practice, case selection remained artificial and arbitrary. Results might have been different with a different selection of treated aneurysms. We did not construct the portfolio using a random sample, because while it may have preserved statistical inferences to the population of interest, a random sample from a routine case series would have included few residual aneurysms and few re-treatments (lower than 5-10%). Using small proportions increases the risk of paradoxical results when reporting kappa indices [26, 27, 30–32]. The context of evaluation also differed from normal practice, and we only provided selected angiographic projections. Access to the multiple angiographic series on a PACS system would have been more realistic and might have introduced more variability.

However, a previous study showed similar results when raters had access to all series on a PACS system or when raters reviewed only selected images [18]. Another research design choice that weakens generalizability to normal practice is that raters were artificially asked to consider the "same ruptured aneurysm patient" to study the clinical significance of the three angiographic grades in isolation from other clinical factors that could influence clinical decisions. The endovascular patients in the portfolio only included patients who had been treated with coil embolization. Whether the classification we propose would show the same reliability if other devices, such as stents, flow diverters, or intra-saccular flow diverters had been used, remains to be studied. A common classification system that could apply to all treatments is needed if we are to reliably compare results and identify best practice [8]. Finally, we only evaluated reliability using catheter angiography. Non-invasive vascular imaging modalities are increasingly used to follow aneurysm patients. An additional problem in comparing results is that patients treated with surgical clipping are commonly followed by CTA, while patients treated with coiling are commonly followed by MRA. The 2 mm definition of a residual neck may be found to be particularly problematic for aneurysms occluded with multiple clips when followed by CTA due to artifacts. Whether the classification we propose will turn out to be reliable when applied to CTA of clipped or MRA of coiled aneurysms remains to be studied.

Conclusion

A simple classification system to grade post-treatment angiographic results was shown to be reliable and clinically relevant for aneurysm patients treated by clipping or coiling. This classification could be used to adjudicate results of comparative trials as well as to standardize the reporting of clipped or coiled aneurysms.

Abbreviations DSA, digital subtraction angiography; MRA, magnetic resonance angiography; CTA, computed tomography angiography; FU, follow-up

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00234-021-02676-0.

Funding This research was funded by a grant (#271994) from the *Fonds de Recherche du Québec – Santé* (FRQS).

Data availability Upon request.

Code availability Upon request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval This is a retrospective research using medical files which received a formal approval from our institutional review board.

Consent to participate For this type of study formal consent is not required.

Consent for publication For this type of study formal consent is not required.

References

1. Macdonald RL, Schweizer TA (2017) Spontaneous subarachnoid haemorrhage. Lancet 389(10069):655–666. https://doi.org/10.1016/S0140-6736(16)30668-7
2. Johnston SC, Dowd CF, Higashida RT, Lawton MT, Duckwiler GR, Gress DR, CARAT Investigators (2008) Predictors of rehemorrhage after treatment of ruptured intracranial aneurysms: the Cerebral Aneurysm Rerupture After Treatment (CARAT) study. Stroke 39(1):120–125. https://doi.org/10.1161/STROKEAHA.107.495747
3. White PM, Lewis SC, Nahser H, Sellar RJ, Goddard T, Gholkar A, on behalf of the HELPS Trial Collaboration (2008) HydroCoil Endovascular Aneurysm Occlusion and Packing Study (HELPS trial): procedural safety and operator-assessed efficacy results. AJNR Am J Neuroradiol 29(2):217–223. https://doi.org/10.3174/ajnr.A0936
4. Darsaut TE, Findlay JM, Raymond J, Group CC (2011) The design of the Canadian UnRuptured Endovascular versus Surgery (CURES) trial. Can J Neurol Sci 38(2):236–241. https://doi.org/10.1017/s0317167100011392
5. Raymond J, Roy D, White PM, Fiorella D, Chapot R, Bracard S, Kallmes DF, THE ICONE COLLABORATIVE GROUP (2008) A randomized trial comparing platinum and hydrogel-coated coils in patients prone to recurrence after endovascular treatment (the PRET trial). Interv Neuroradiol 14(1):73–83. https://doi.org/10.1177/159101990801400110
6. Becske T, Kallmes DF, Saatci I, McDougall CG, Szikora I, Lanzino G et al (2013) Pipeline for uncoilable or failed aneurysms: results from a multicenter clinical trial. Radiology 267(3):858–868. https://doi.org/10.1148/radiol.13120099
7. Pierot L, Moret J, Barreau X, Szikora I, Herbreteau D, Turjman F, Holtmannspötter M, Januel AC, Costalat V, Fiehler J, Klisch J, Gauvrit JY, Weber W, Desal H, Velasco S, Liebig T, Stockx L, Berkefeld J, Molyneux A, Byrne J, Spelle L (2018) Safety and efficacy of aneurysm treatment with WEB in the cumulative population of three prospective, multicenter series. J Neurointerv Surg 10(6):553–559. https://doi.org/10.1136/neurintsurg-2017-013448
8. Darsaut TE, Chapot R, Raymond J (2020) Changing the rules of the game: the problem of surrogate angiographic outcomes in the evaluation of aneurysm treatments. AJNR Am J Neuroradiol 41:2174–2175. https://doi.org/10.3174/ajnr.A6825
9. Molyneux AJ, Kerr RS, Yu LM, Clarke M, Sneade M, Yarnold JA et al (2005) International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients

with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. Lancet 366(9488):809–817. https://doi.org/10.1016/S0140-6736(05)67214-5
10. Burkhardt JK, Chua MHJ, Weiss M, Do ASS, Winkler EA, Lawton MT (2017) Risk of aneurysm residual regrowth, recurrence, and de novo aneurysm formation after microsurgical clip occlusion based on follow-up with catheter angiography. World Neurosurg 106:74–84. https://doi.org/10.1016/j.wneu.2017.06.110
11. Kotowski M, Farzin B, Fahed R, Guilbert F, Chagnon M, Darsaut TE, Daniel RT, Raymond J (2019) Residual cerebral aneurysms after microsurgical clipping: a new scale, an agreement study, and a systematic review of the literature. World Neurosurg 121:e302–e321. https://doi.org/10.1016/j.wneu.2018.09.100
12. Darsaut TE, Jack AS, Kerr RS, Raymond J (2013) International subarachnoid aneurysm trial - ISAT part II: study protocol for a randomized controlled trial. Trials 14:156. https://doi.org/10.1186/1745-6215-14-156
13. Darsaut TE, Findlay JM, Magro E, Kotowski M, Roy D, Weill A, Bojanowski MW, Chaalala C, Iancu D, Lesiuk H, Sinclair J, Scholtes F, Martin D, Chow MM, O'Kelly CJ, Wong JH, Butcher K, Fox AJ, Arthur AS, Guilbert F, Tian L, Chagnon M, Nolet S, Gevry G, Raymond J (2017) Surgical clipping or endovascular coiling for unruptured intracranial aneurysms: a pragmatic randomised trial. J Neurol Neurosurg Psychiatry 88(8):663–668. https://doi.org/10.1136/jnnp-2016-315433
14. Darsaut TE, Desal H, Cognard C, Januel AC, Bourcier R, Boulouis G, Shiva Shankar JJ, Findlay JM, Rempel JL, Fahed R, Boccardi E, Valvassori L, Magro E, Gentric JC, Bojanowski MW, Chaalala C, Iancu D, Roy D, Weill A, Diouf A, Gevry G, Chagnon M, Raymond J (2020) Comprehensive aneurysm management (CAM): an all-inclusive care trial for unruptured intracranial aneurysms. World Neurosurg 141:e770–e777. https://doi.org/10.1016/j.wneu.2020.06.018
15. Cloft HJ, Kallmes DF (2011) Scaling back on scales with a scale of scales. AJNR Am J Neuroradiol 32(2):219–220. https://doi.org/10.3174/ajnr.A2432
16. Tollard E, Darsaut TE, Bing F, Guilbert F, Gevry G, Raymond J (2012) Outcomes of endovascular treatments of aneurysms: observer variability and implications for interpreting case series and planning randomized trials. AJNR Am J Neuroradiol 33(4):626–631. https://doi.org/10.3174/ajnr.A2848
17. Ernst M, Yoo AJ, Kriston L, Schonfeld MH, Vettorazzi E, Fiehler J (2015) Is visual evaluation of aneurysm coiling a reliable study end point? Systematic review and meta-analysis. Stroke 46(6):1574–1581. https://doi.org/10.1161/STROKEAHA.114.008513
18. Jamali S, Fahed R, Gentric JC, Letourneau-Guillon L, Raoult H, Bing F, Estrade L, Nguyen TN, Tollard E, Ferre JC, Iancu D, Naggara O, Chagnon M, Weill A, Roy D, Fox AJ, Kallmes DF, Raymond J (2016) Inter- and intrarater agreement on the outcome of endovascular treatment of aneurysms using MRA. AJNR Am J Neuroradiol 37(5):879–884. https://doi.org/10.3174/ajnr.A4609
19. Raymond J, Roy D, Bojanowski M, Moumdjian R, L'Esperance G (1997) Endovascular treatment of acutely ruptured and unruptured aneurysms of the basilar bifurcation. J Neurosurg 86(2):211–219. https://doi.org/10.3171/jns.1997.86.2.0211
20. Roy D, Milot G, Raymond J (2001) Endovascular treatment of unruptured aneurysms. Stroke 32(9):1998–2004. https://doi.org/10.1161/hs0901.095600
21. Raymond J, Guilbert F, Weill A, Georganos SA, Juravsky L, Lambert A et al (2003) Long-term angiographic recurrences after selective endovascular treatment of aneurysms with detachable coils. Stroke 34(6):1398–1403. https://doi.org/10.1161/01.STR.0000073841.88563.E9
22. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A et al (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 64(1):96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002
23. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097
24. Donner A, Rotondi MA (2010) Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. Int J Biostat 6(1):Article 31. https://doi.org/10.2202/1557-4679.1275
25. Fahed R, Darsaut TE, Farzin B, Chagnon M, Raymond J (2020) Measuring clinical uncertainty and equipoise by applying the agreement study methodology to patient management decisions. BMC Med Res Methodol 20(1):214. https://doi.org/10.1186/s12874-020-01095-8
26. Feinstein AR, Cicchetti DV (1990) High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 43(6):543–549. https://doi.org/10.1016/0895-4356(90)90158-l
27. Cicchetti DV, Feinstein AR (1990) High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 43(6):551–558. https://doi.org/10.1016/0895-4356(90)90159-m
28. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174
29. McDougall CG, Johnston SC, Hetts SW, Gholkar A, Barnwell SL, Vazquez Suarez JC et al (2020) Five-year results of randomized bioactive versus bare metal coils in the treatment of intracranial aneurysms: the Matrix and Platinum Science (MAPS) Trial. J Neurointerv Surg:neurintsurg-2020-016906. https://doi.org/10.1136/neurintsurg-2020-016906
30. Byrt T, Bishop J, Carlin JB (1993) Bias, prevalence and kappa. J Clin Epidemiol 46(5):423–429. https://doi.org/10.1016/0895-4356(93)90018-v
31. Quarfoot A, Levine RA (2016) How robust are multirater interrater reliability indices to changes in frequency distribution? Am Stat 70(4):373–384. https://doi.org/10.1080/00031305.2016.1141708
32. Gwet KL (2014) The kappa coefficient and its paradoxes. In: handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. 4th edn. Advanced Analytics, LLC, Gaithersburg, MD, pp 57–62

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Anass Benomar[1] · Behzad Farzin[1] · David Volders[2] · Guylaine Gevry[1] · Justine Zehr[3] · Robert Fahed[4] · William Boisseau[1] · Jean-Christophe Gentric[5] · Elsa Magro[6] · Lorena Nico[7] · Daniel Roy[1] · Alain Weill[1] · Charbel Mounayer[8] · François Guilbert[1] · Laurent Létourneau-Guillon[1] · Gregory Jacquin[9] · Chiraz Chaalala[10] · Marc Kotowski[11] · Thanh N. Nguyen[12] · David Kallmes[13] · Phil White[14] · Tim E. Darsaut[15] · Jean Raymond[1]

Anass Benomar
anass.benomar@umontreal.ca

Behzad Farzin
behzadfarzin@yahoo.com

David Volders
david.volders@gmail.com

Guylaine Gevry
guylaine.gevry@crchum.qc.ca

Justine Zehr
Justine.zehr@umontreal.ca

Robert Fahed
robert.fahed@hotmail.fr

William Boisseau
boisseau.william@gmail.com

Jean-Christophe Gentric
jcgentric@gmail.com

Elsa Magro
elsa.magro@chu-brest.fr

Lorena Nico
lorena.nico@gmail.com

Daniel Roy
daniel.roy.chum@ssss.gouv.qc.ca

Alain Weill
alain.weill.chum@ssss.gouv.qc.ca

Charbel Mounayer
cmounayer@hotmail.com

François Guilbert
francois.guilbert@gmail.com

Laurent Létourneau-Guillon
Laurent.letourneau-guillon.1@umontreal.ca

Gregory Jacquin
gregory.jacquin.chum@ssss.gouv.qc.ca

Chiraz Chaalala
cchaalala@gmail.com

Marc Kotowski
Kotowski80@gmail.com

Thanh N. Nguyen
Thanh.Nguyen@bmc.org

David Kallmes
Kallmes.david@mayo.edu

Phil White
phil.white@ncl.ac.uk

Tim E. Darsaut
tdarsaut@ualberta.ca

[1] Department of Radiology, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, QC H2X 0C1, Canada

[2] Department of Diagnostic Radiology, Queen Elizabeth II Health Sciences Centre, Halifax, NS, Canada

[3] Department of Mathematics and Statistics, Université de Montréal, Montreal, QC, Canada

[4] Division of Neurology, The Ottawa Hospital, Ottawa, ON, Canada

[5] Service de neuroradiologie interventionnelle, Hôpital de la Cavale Blanche, CHRU de Brest, Brest, France

[6] Service de neurochirurgie, Hôpital de la Cavale Blanche, CHRU de Brest, Brest, France

[7] Service de radiologie, CHU de Caen Normandie, Caen, France

[8] Service de neuroradiologie interventionnelle, Hôpital Dupuytren, CHU Limoges, Limoges, France

[9] Department of Medicine, Division of Neurology, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, QC, Canada

[10] Division of Neurosurgery, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, QC, Canada

[11] Service de neurochirurgie, Hôpital de la Providence, Neuchâtel, Switzerland

[12] Department of Neurology, Neurosurgery, and Radiology, Boston Medical Center, Boston, MA, USA

[13] Department of Radiology, Mayo Clinic, Rochester, MN, USA

[14] Neuroradiology, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle Upon Tyne, UK

[15] Department of Surgery, Division of Neurosurgery, University of Alberta Hospital, Edmonton, AB, Canada