

1 **Subarachnoid Hemorrhage Trials:**

2 **Cutting, Sliding, or Keeping mRS Scores and WFNS Grades**

3
4 Akshittkumar M. Mistry, MD;¹ Jeffrey Saver, MD;² William Mack, MD;³ Hooman Kamel, MD;⁴
5 Jordan Elm, PhD;⁵ Jonathan Beall, PhD⁵

6
7 *¹Department of Neurosurgery, University of Louisville, Louisville, KY; ²Department of Neurology*
8 *and Comprehensive Stroke Center, David Geffen School of Medicine, University of California,*
9 *Los Angeles, CA; ³Department of Neurosurgery, Keck School of Medicine, University of*
10 *Southern California, Los Angeles, CA; ⁴Clinical and Translational Neuroscience Unit, Feil Family*
11 *Brain and Mind Research Institute and Department of Neurology, Weill Cornell Medicine, New*
12 *York, NY; ⁵Department of Public Health Sciences, Medical University of South Carolina,*
13 *Charleston, SC.*

14
15 **Short title:** Outcomes, Eligibility, & Sample Size in SAH Trials

16 **Total word count:** 2957

17
18 **Corresponding Author:**

19 Address: Department of Neurological Surgery
20 University of Louisville
21 220 Abraham Flexner Way, 15th Floor
22 Louisville, KY 40202

23 Telephone: 502-588-2329

24 Fax: 502-407-3256

25 Email: axitamm@gmail.com

26 **X (Twitter) tags:** @UofLNeurosurg ; @HoomanKamel ; @NeurosurgryUSC ; @UCLASroke

27 **Abstract**

28 Rigorous evidence generation with randomized controlled trials (RCTs) has lagged for
29 aneurysmal subarachnoid hemorrhage (SAH) compared to other forms of acute stroke. Besides
30 its lower incidence compared to other stroke subtypes, the presentation and outcome of SAH
31 patients also differ. This must be considered and adjusted for in designing pivotal RCTs of SAH
32 patients. Here, we show the effect of the unique expected distribution of the SAH severity at
33 presentation (World Federation of Neurological Surgeons, WFNS, grade) on the outcome most
34 used in pivotal stroke RCTs (modified Rankin Scale, mRS) and consequently on the sample
35 size. Further, we discuss the advantages and disadvantages of different options to analyze the
36 outcome and control the expected distribution of WFNS grades in addition to showing their
37 effects on the sample size. Last, we offer methods that investigators can adapt to more
38 precisely understand the effect of common mRS analysis methods and trial eligibility pertaining
39 to the WFNS grade in designing their large-scale SAH RCTs.

40 **Purpose**

41 The generation of rigorous evidence to inform the management of patients with aneurysmal
42 subarachnoid hemorrhage (SAH) has lagged other types of acute strokes. The reason for this
43 lag is multifactorial—one being that SAH has the lowest incidence of all forms of stroke.
44 However, the paucity of SAH randomized controlled trials (RCTs) can also be self-exacerbating.
45 Rather than adopting existing trial designs and biostatistical methods, it forces new investigators
46 to craft these anew. Here, we provide a basic biostatistical guide for investigators to navigate
47 two foundational dilemmas in designing large-scale SAH RCTs.

48

49 **Measures of SAH Severity and Outcome**

50 The severity of presentation universally predicts patients' functional outcomes in stroke. The
51 ordinal World Federation of Neurological Surgeons (WFNS) scale¹ is one scale used to grade
52 the SAH severity, increasing from 1 to 5. Although the Glasgow Outcome Scale has been used
53 for SAH, the archetypal outcome used to measure neurological morbidity and mortality in stroke
54 patients is the ordinal modified Rankin Scale (mRS), ranging from 0 (no neurological disability)
55 to 6 (death). For a biostatistically-informed trial, it is critical to estimate the expected distribution
56 of WFNS grades, mRS scores, and correlation between the two.

57

58 **Expected WFNS Distribution is Skewed**

59 The expected distribution of WFNS grades is largely skewed toward grade 1. For exemplary
60 purposes, we retrieved a few large RCTs with broad eligibility criteria and cohort studies from
61 different regions of the world (which should possess less selection bias common to research
62 studies) and estimated the proportion of WFNS grade 1 to 5 (in %): 41, 22, 7, 15, 15 (Figure 1A;
63 Supplemental Figure 1), with grade 2 being the median. For each WFNS grade, we acquired the
64 outcome on the full mRS scale from the ULTRA² RCT (Figure 1B). We chose ULTRA because it
65 is the most recent (2021) pivotal SAH RCT of 955 patients treated with the current standard of

66 care. Collectively, the skewed expected WFNS distribution and the correlation between WFNS
67 grade and mRS outcome in SAH patients may pose challenges. The effect of any intervention in
68 WFNS grade 1 patients will be diluted because 75% of them will already have a good outcome
69 (commonly accepted as mRS 0-2; Figure 1B). Conversely, unless an intervention has strong
70 therapeutic effects sufficient to improve substantial brain injury in WFNS grade 5 patients, the
71 effect size will also be diluted because 76% of them will have a bad outcome (mRS 4-6). Since
72 interventions confirmed to show a therapeutic benefit for SAH patients in RCTs are scarce,³ the
73 design of large-scale SAH RCTs must be optimized to detect a therapeutic effect considering
74 variable effects based on WFNS grade.

75

76 **Outcome: Cut, Slide, or Keep mRS Scores?**

77 Deciding how to analyze the mRS scale is a frequent dilemma. Common options are 1) fixed
78 dichotomization, focusing on either good outcomes (mRS 0-2 vs. mRS 3-6) or bad outcomes
79 (mRS 4-6 vs. 0-3); 2) “responder” analysis with a sliding dichotomy, prespecifying what
80 constitutes a good outcome depending on the presenting SAH severity (for example, mRS 0-1
81 for WFNS grade 1, mRS 0-2 for grade 2-4, and mRS 0-4 for grade 5); and 3) “shift analysis”
82 across all mRS scores analyzed with an ordinal regression.

83

84 We studied the impact of these options on the sample size for a two-group pivotal SAH trial
85 (80% power; two-tailed α 0.05) using the baseline (control) mRS distribution dictated by
86 expected WFNS distribution (Figure 1B-C). While fixed dichotomy at mRS 0-2 and mRS 4-6 is
87 widely accepted, there are no precedents or acceptable, validated schemes based on WFNS
88 grades for sliding dichotomy. Therefore, we generated all reasonable permutations of sliding
89 dichotomy schemes by limiting a good outcome of WFNS grade 1 to mRS ≤ 2 and grade 5 to
90 mRS 3-4. An mRS 0-2 for WFNS grade 5 is not only ambitious but also numerically constricts
91 the “slide” and thus greatly limits the sliding options because the fundamental principle of a

92 sliding dichotomy is cutting (or dichotomizing) the mRS scale at the same or higher mRS score
93 for sequentially higher WFNS grades. For example, allowing dichotomization of WFNS grade 5
94 at mRS 2 will limit WFNS grades 1-4 to be dichotomized at mRS ≤ 2 only (i.e., a “quicker” slide)
95 as opposed to if WFNS grade 5 is dichotomized at mRS 4, then WFNS grades 1-4 can be
96 dichotomized broadly along mRS scores ≤ 4 (see examples in Supplemental Figure 2). With
97 these rules, 99 sliding dichotomy permutations or schemes were generated. Thus, including the
98 two commonly used fixed dichotomy schemes, we tested a total of 101 dichotomy schemes.

99
100 For exemplary purposes, we considered the common effect size of 10% absolute change in the
101 dichotomized outcomes. (For reference, Supplemental Figure 3 reviews the difference between
102 absolute and relative effect sizes.) Using the expected mRS distribution from Figure 1C as the
103 control group, a fixed dichotomy at mRS 0-2 requires 14.6% more patients than at mRS 4-6
104 (754 vs. 658; see Figure 2A blue versus purple lines at 10% effect size). Sliding dichotomy
105 schemes generally lower the required sample size compared to fixed dichotomy for a 10%
106 absolute effect size. The sample size required with 49 of the 99 sliding dichotomy schemes is
107 less than that required for mRS 4-6 fixed dichotomy, and 11 sliding dichotomy schemes require
108 a sample size greater than that required for mRS 0-2 fixed dichotomy. Their sample sizes
109 depend on how quickly the “slide” occurs from WFNS grade 1 down to 5 (blue dots in
110 Supplemental Figure 4). The lowest sample size (340) is for a scheme setting a good outcome
111 for grades 1-4 at mRS 0 and grade 5 at mRS ≤ 3 , denoted as (0,0,0,0,3) for each sequential
112 WFNS grade, while the highest (778) is for a scheme setting a good outcome for grade 1 at
113 mRS ≤ 1 and grades 2-5 at mRS ≤ 4 , denoted as (1,4,4,4,4). Last, if one intends to use all the
114 information across the full mRS scale rather than dichotomizing it, then the sample size can only
115 be estimated if the control mRS distribution and the expected mRS distribution for the
116 intervention are known. The sample size may be greater or lesser than the dichotomy schemes,

117 being entirely dependent on the two distributions, and the latter is often unknown. If a constant
118 therapeutic effect is expected across all mRS scores (the practical validity of this assumption is
119 highly arguable), then one may generate an expected mRS distribution for the intervention using
120 a constant log odds shift (i.e., effect size; Figure 2B). With this assumption, a constant log odds
121 shift generally yields a lower sample size than fixed dichotomies (Figure 2C).

122
123 We focused on the influence of these outcome constructions on the sample size because
124 sample size is concrete and often drives resources and cost. However, the decision should
125 ultimately be multifactorial. The advantages and disadvantages of these approaches have been
126 discussed in this reference.⁴ For the unique expected control distribution of mRS scores for
127 SAH, it is critical to consider where is the greatest numerical and/or theoretical potential to
128 detect a therapeutic effect based on the biological effect of therapy (i.e., further improving the
129 outcomes of the large portion of mRS 2 patients, or decreasing mortality, the second largest
130 portion of patients—or both; Figure 1C). Last, utility-weighted mRS is an option, but it is not
131 well-studied or measured in aSAH patients,⁵ who receive different medical and surgical
132 treatments compared to ischemic stroke patients.

133

134 **Eligibility: Cut, Slide, or Keep WFNS Grades?**

135 Eligibility criteria can influence a trial's sensitivity to detect a therapeutic effect. As discussed
136 above, interventions could be less therapeutic on either end of the WFNS grade spectrum, thus
137 including all WFNS grades may produce a diluted effect size of the intervention. While outright
138 excluding WFNS grade 1 and/or 5 patients in pivotal SAH RCTs (which make up ~55% of SAH
139 patients) is an option, we caution against it to maintain the generalizability of the intervention
140 given the lack of therapies with proven benefits in SAH patients and to not prolong the study
141 with a lower the enrollment rate. A compromising option is to limit the enrollment of patients on
142 either end of the WFNS grade spectrum. Although this reduces the enrollment rate and may

143 dilute the treatment effect, it maintains some level of generalizability. Given that ~40% of the
144 SAH patients are WFNS grade 1, their enrollment could be limited to a percent that is deemed
145 substantial for applicability of the results in clinical practice that is deemed to not drastically slow
146 enrollment. Similarly, WFNS grade 5 patients can be limited, for example, using a different
147 approach. Recent multicenter studies have shown that WFNS grade 5 patients can be further
148 prognostically subclassified using the ‘herniated WFNS’ scale.⁶ Herniated WFNS grade 5 make
149 up ~30% of the WFNS grade 5 patients, and an overwhelming 88% of them have a poor
150 outcome (mRS 4-6). The primary disadvantage of limiting WFNS grade 5 patients to only non-
151 herniated WFNS grade 5 is the uncaptured applicability and missed opportunity to show the
152 benefit of the intervention in this subpopulation with high mortality. We evaluated the effect on
153 the sample size of limiting WFNS grade 1 from ~40% to 0 and excluding herniated WFNS grade
154 5, comparing them to the default option of keeping all patients. We estimated the sample size of
155 a two-group pivotal trial testing a 10% absolute effect size (80% power; two-tailed α 0.05).

156
157 Counterintuitively, in most cases, limiting WFNS grade 1 and excluding herniated WFNS grade
158 5 patients each increases the sample size compared to keeping all patients (Figure 2D).
159 Limiting WFNS grade 1 decreases the sample size (max by -56) across the spectrum down to
160 0% in only 6 of the total 101 fixed or sliding dichotomy schemes. The maximum increase is 126.
161 Excluding herniated WFNS grade 5 patients does not change the previously established
162 dependence of sample size of the sliding dichotomy schemes on how quickly the “slide” occurs
163 from WFNS grade 1 down to 5 (Supplemental Figure 4). The “quicker” and “slower” sliding
164 dichotomy schemes experience a greater change in sample size. The largest increase in
165 sample size (70) by excluding herniated WFNS grade 5 patients occurs in a scheme setting a
166 good outcome as mRS 0 for grades 1-4 and mRS \leq 3 for grade 5 (0,0,0,0,3), while the largest
167 decreases occur in schemes setting mRS \leq 1 for grade 1 and mRS \leq 4 for grades 2-5 (1,4,4,4,4;

168 -64) and the fixed dichotomy at mRS 0-3 vs. mRS 4-6 (-68). Excluding herniated WFNS grade 5
169 patients in 30 of the 101 dichotomy schemes decreases the sample size across the range of
170 WFNS grade 1 proportions. The combined effect of excluding both WFNS grade 1 and
171 herniated WFNS grade 5 is an increase in the sample size (max +172) in 87 schemes, including
172 the fixed dichotomies, and a decrease (max -92) in 14 schemes. Of the fixed dichotomies, mRS
173 0-2 vs. mRS 3-6 is less sensitive to any of these single or combination of these approaches
174 (min 732, max 784). Last, the sample size for a test (Wilcoxon-Mann-Whitney or proportional
175 odds regression) utilizing all the information across the full mRS scale is relatively independent
176 of the two WFNS eligibility approaches we discuss simply because the sample size is entirely
177 dependent on the control and expected distributions. The latter often is unknown. Even
178 predicting the expected distribution with a constant log odds shift in the control distribution will
179 have minimal effect on the sample size.

180
181 Overall, these results seem counterintuitive. By limiting WFNS grade 1 and 5 patients, one may
182 expect a lower sample size with a greater effect size of the intervention. However, we did not re-
183 estimate the sample size with a higher expected effect size, which would decrease the sample
184 size, because often preliminary estimates of the latter are unavailable. Second, these are purely
185 mathematical results. The sample size needed to test an absolute 10% decrease (i.e., effect
186 size) with the two-sample test for proportions is greatest when the control proportion is 55%
187 (Supplemental Figure 5). Because of the skewed WFNS distribution and its dependent control
188 mRS distribution, limiting WFNS grade 1 and 5 patients results in the control proportion
189 approaching 55% in most dichotomy schemes. If the effect size of the intervention being tested
190 has not been estimated with a preliminary study limiting WFNS grade 1 and 5 patients but is
191 expected to be higher, we suggest an interim analysis that best fits the study to save resources.

192

193

194

195 **Conclusion**

196 The expected distribution of WFNS scores skewed to the lowest grade dictates the sample size
197 based on the primary analysis approach. There are advantages and disadvantages of different
198 analysis approaches (fixed or sliding dichotomy and using the full mRS scale) and options to
199 modify the expected WFNS distribution in a trial. Greater sample sizes are required for fixed
200 dichotomies (especially mRS 0-2 vs. 3-6) and while most of the sliding dichotomies yield a lower
201 sample size, there is no established precedent for picking one scheme over the other. Although
202 sample size estimates are more precise when using all information across the entire mRS scale,
203 this requires that both the control and expected distributions are known. While estimating the
204 latter with a constant therapeutic effect across all mRS scores can be done and offers the
205 greatest power compared to dichotomous schemes, the practical validity of this approach is
206 arguable. Generally, limiting WFNS grade 1 enrollment and excluding justified subpopulations of
207 WFNS grade 5 to increase the effect size of the intervention being tested will also increase the
208 sample size. We provide the R code (Supplemental Content) that investigators can adapt to
209 their expected WFNS and mRS distributions to navigate the two foundational dilemmas more
210 precisely in designing large-scale SAH RCTs.

211

212 **Acknowledgments:** We thank René Post, MD, PhD, Department of Neurosurgery,
213 Neurosurgical Centre Amsterdam, Amsterdam University Medical Centres, Amsterdam, the
214 Netherlands, for providing aggregated data from ULTRA.

215

216 **Sources of funding:** NINDS U01NS087748 (J.E. and J.B.)

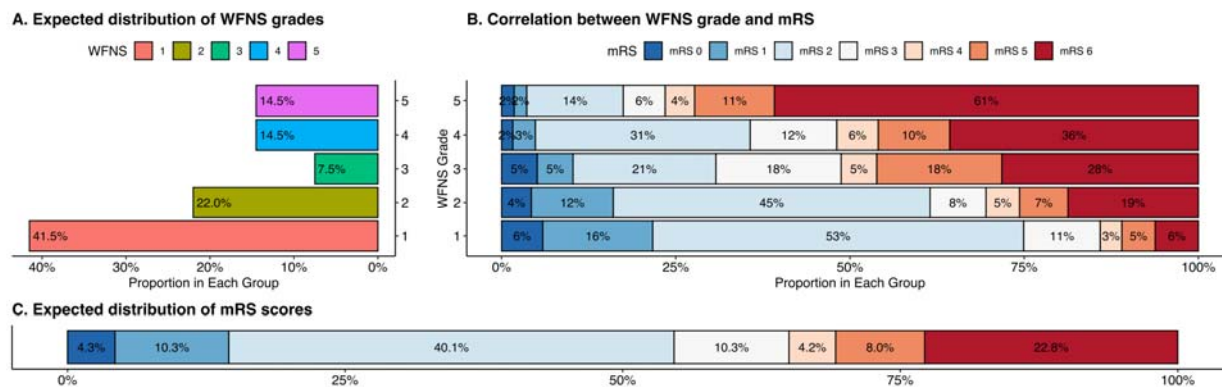
217

218 **Disclosures:** None

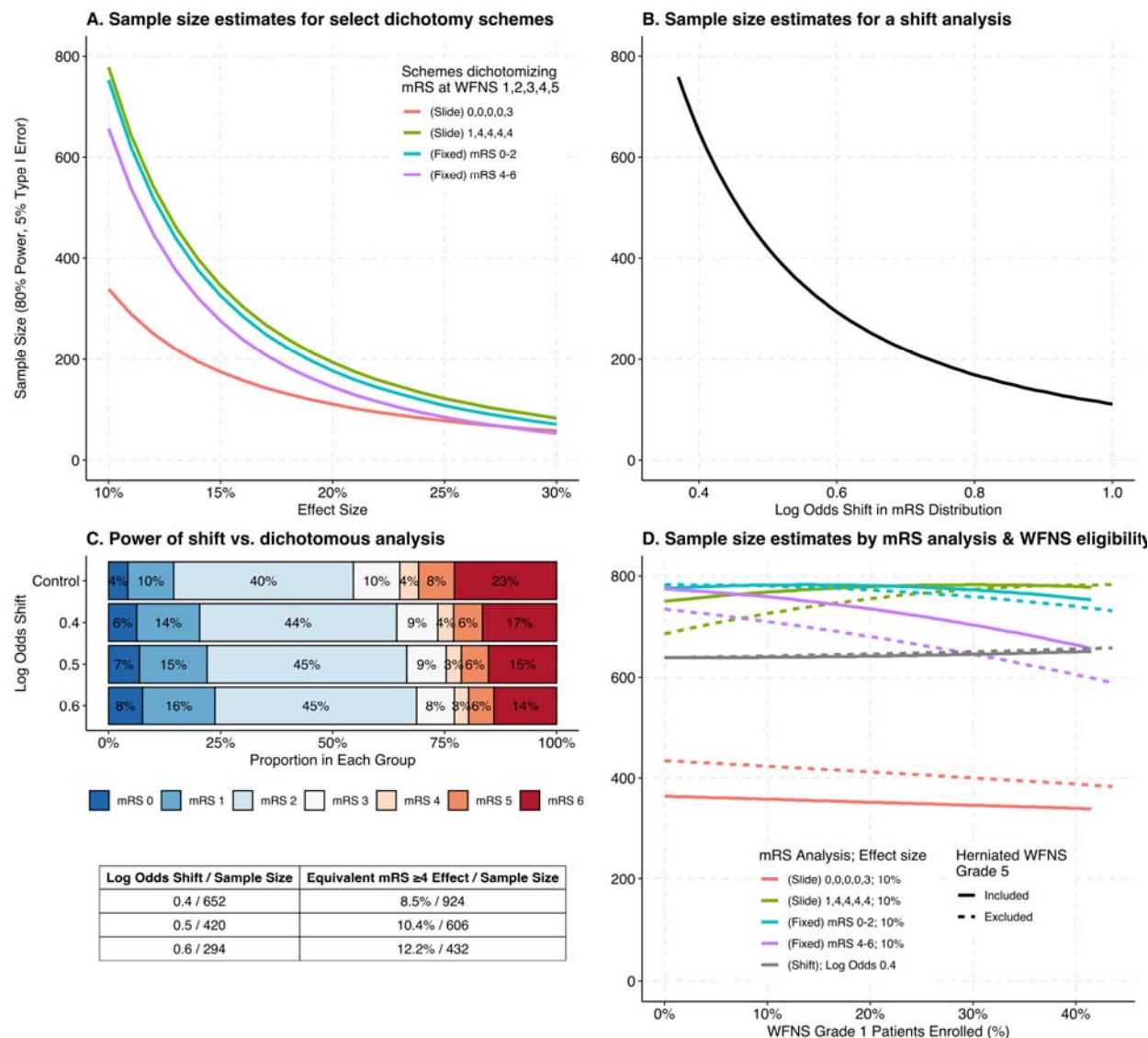
219 **References**

- 220 1. Report of world federation of neurological surgeons committee on a universal
221 subarachnoid hemorrhage grading scale. *J Neurosurg.* 1988;68:985-986.
- 222 2. Post R, Germans MR, Tjerkstra MA, Vergouwen MDI, Jellema K, Koot RW, et al. Ultra-
223 early tranexamic acid after subarachnoid haemorrhage (ultra): A randomised controlled
224 trial. *Lancet.* 2021;397:112-118.
- 225 3. Hoh BL, Ko NU, Amin-Hanjani S, Hsiang-Yi Chou S, Cruz-Flores S, Dangayach NS, et
226 al. 2023 guideline for the management of patients with aneurysmal subarachnoid
227 hemorrhage: A guideline from the american heart association/american stroke
228 association. *Stroke.* 2023;54:e314-e370.
- 229 4. Saver JL. Optimal end points for acute stroke therapy trials: Best ways to measure
230 treatment effects of drugs and devices. *Stroke.* 2011;42:2356-2362.
- 231 5. Custal C, Koehn J, Borutta M, Mrochen A, Brandner S, Eyupoglu IY, et al. Beyond
232 functional impairment: Redefining favorable outcome in patients with subarachnoid
233 hemorrhage. *Cerebrovasc Dis.* 2021;50:729-737.
- 234 6. Raabe A, Beck J, Goldberg J, WJ ZG, Branca M, Marbacher S, et al. Herniation world
235 federation of neurosurgical societies scale improves prediction of outcome in patients
236 with poor-grade aneurysmal subarachnoid hemorrhage. *Stroke.* 2022;53:2346-2351.
- 237

238 **Figure Legends**

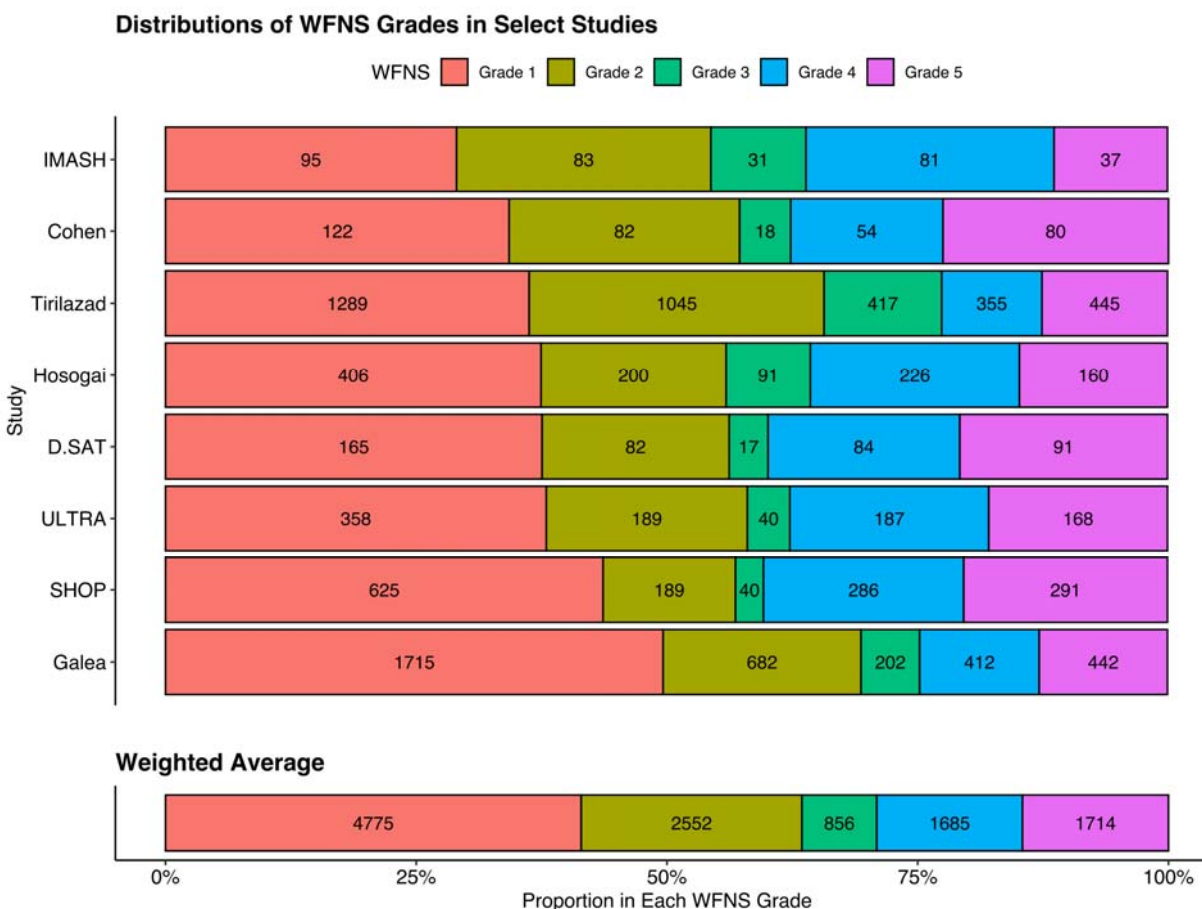


239
 240 **Figure 1. Expected distribution of World Federation of Neurological Surgeons (WFNS)**
 241 **grades (A) and modified Rankin Scale (mRS) scores for subarachnoid hemorrhage**
 242 **patients (B,C).** Data in (A) are derived from a weighted average of 8 large studies in
 243 Supplemental Figure 1 and (B) from the ULTRA trial.



244
 245 **Figure 2. Samples size calculations for a two-group pivotal subarachnoid hemorrhage**
 246 **trial.** WFNS=World Federation of Neurological Surgeons; mRS=modified Rankin Scale. Sample
 247 size estimates for (A) two examples of sliding and two fixed dichotomized outcomes using a
 248 two-sample test of proportions and (B) shift analysis using the Wilcoxon-Mann-Whitney test and
 249 control distribution in Figure 1C and (C). The sliding notation, for example, 1,4,4,4,4 designates
 250 the dichotomizing mRS score (\leq) for each WFNS grade sequentially. (C) compares sample
 251 sizes of shift analyses by equating them to the effect sizes in a dichotomous analysis of mRS.
 252 (D) Sample size estimates based on WFNS grade 1 enrollment and grade 5 eligibility.

253 **SUPPLEMENTAL CONTENT**



254

255 **Supplemental Figure 1.** Estimation of expected World Federation of Neurological Surgeons

256 (WFNS) grade in SAH patients. WFNS grade distributions from selected 8 studies are depicted.

257 The weighted average is calculated and also plotted in manuscript Figure 1A. The references

258 for these studies are listed below.

259

260 References:

261 IMASH - Wong GKC, et al., IMASH Investigators. Intravenous magnesium sulphate for

262 aneurysmal subarachnoid hemorrhage (IMASH): a randomized, double-blinded, placebo-

263 controlled, multicenter phase III trial. *Stroke* 2010 May;41(5):921-6. doi:

264 10.1161/STROKEAHA.109.571125; PMID:20378868

265

266 Cohen – Cohen, et al., Dysnatremia and 6-Month Functional Outcomes in Critically Ill Patients
267 With Aneurysmal Subarachnoid Hemorrhage: A Prospective Cohort Study. Crit Care Explor
268 2021 Jun 8;3(6):e0445. doi: 10.1097/CCE.0000000000000445; PMID: 20378868

269

270 Tirilazad / D-SAT / SHOP – Data obtained from Jaja, et al., Development and validation of
271 outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational
272 cohort study. BMJ 2018 Jan 18;360:j5745. doi: 10.1136/bmj.j5745. PMID: 29348138

273

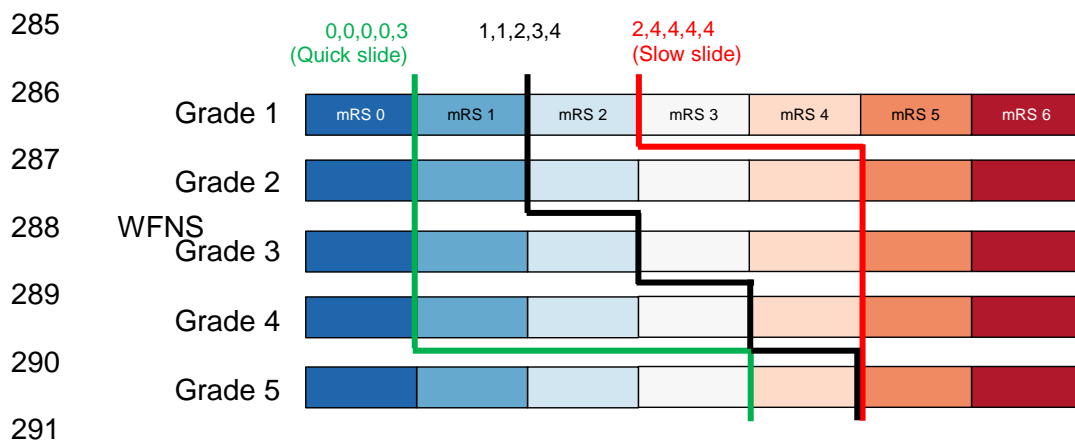
274 Hosogai – Hosogai, et al., Changes in Short-Term Outcomes After Discharge in Patients with
275 Aneurysmal Subarachnoid Hemorrhage: A Multicenter, Prospective, Observational Study. World
276 Neurosurg 2022 Aug;164:e1214-e1225; PMID: 35688375

277

278 ULTRA – Post, et al., Ultra-early tranexamic acid after subarachnoid haemorrhage (ULTRA): a
279 randomised controlled trial. Lancet 2021 Jan 9;397(10269):112-118. doi: 10.1016/S0140-
280 6736(20)32518-6; PMID: 33357465

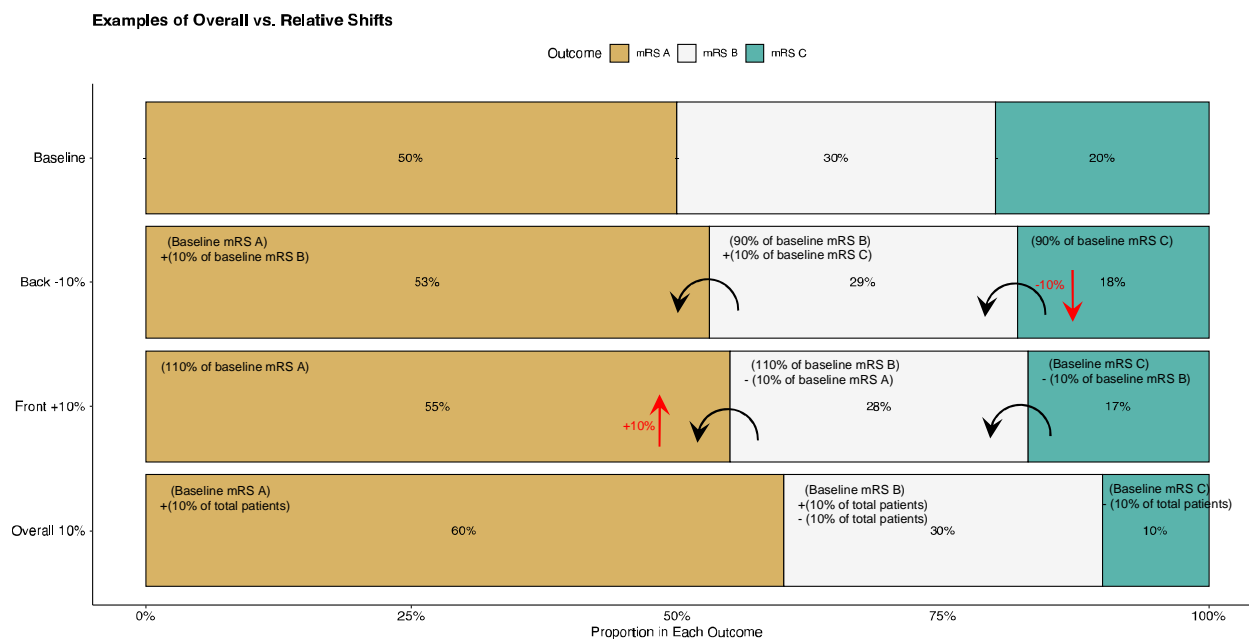
281

282 Galea – Galea, et al., Predictors of Outcome in Aneurysmal Subarachnoid Hemorrhage
283 Patients: Observations From a Multicenter Data Set. Stroke 2017 Nov;48(11):2958-2963. doi:
284 10.1161/STROKEAHA.117.017777; 28974630



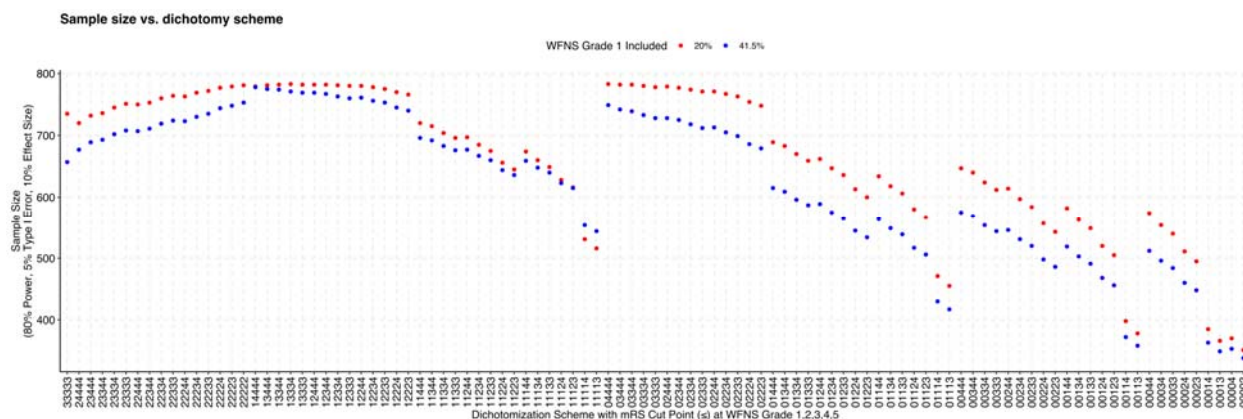
292 **Supplemental Figure 2.** Examples of sliding dichotomies. Showing in green and red are quick
293 and slow slides. The notation, for example, 1,1,2,3,4 designates the dichotomizing mRS score
294 (\leq) for each WFNS grade sequentially.

295

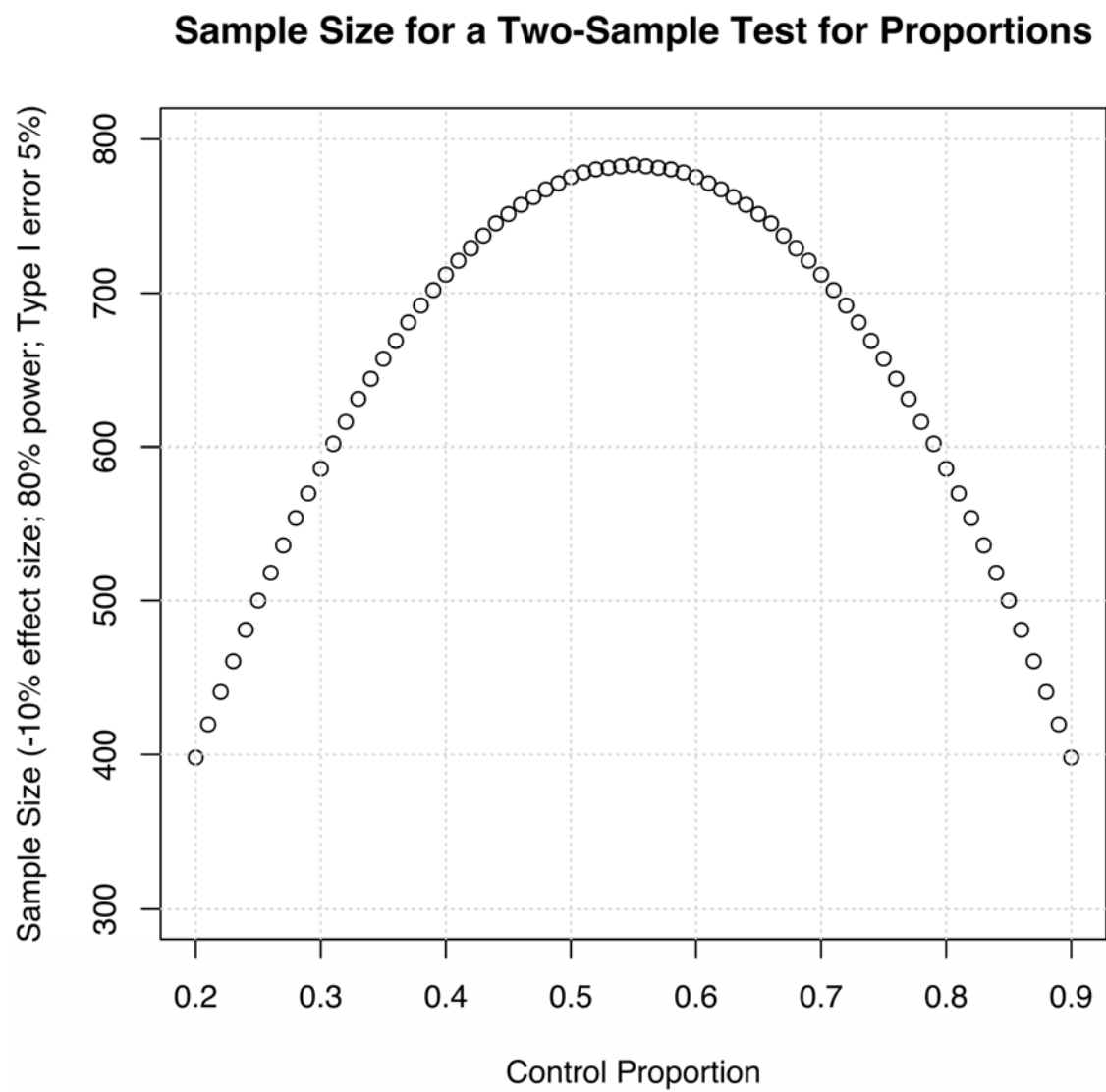


296

297 **Supplemental Figure 3.** Examples of overall (absolute) vs. relative shifts. For exemplary
 298 purposes, we crafted a 3-scale ordinal mRS outcome (A, B, C). The control (baseline)
 299 distribution is shown with 50% of the patients having outcome A, 30% B, and 20% C. “Back -
 300 10%” depicts the resulting distribution from the baseline by starting to decrease the proportion of
 301 patients with outcome C by 10%. “Front +10%” depicts the resulting distribution from the
 302 baseline by starting to increase the proportion of patients with outcome A by 10%. “Overall 10%”
 303 depicts the resulting distribution of a 10% absolute shift.



304
 305 **Supplemental Figure 4.** Sample size estimates of 101 dichotomization schemes. The notation,
 306 for example, 11234 designates the dichotomizing mRS score (\leq) for each WFNS grade
 307 sequentially: mRS ≤ 1 for grade 1, mRS ≤ 1 for grade 2, mRS ≤ 2 for grade 3, mRS ≤ 3 for grade
 308 4, and mRS ≤ 4 for grade 5. There are two fixed dichotomy schemes: 22222, which equates to
 309 mRS ≤ 2 , and 33333, which equates to mRS ≥ 4 (which is the same as mRS ≤ 3). Red dots are
 310 sample sizes when WFNS grade 1 patient enrollment is limited to 20% from what is expected
 311 (41.5%).



312 **Supplemental Figure 5.** Sample size estimates for a two-sample test for proportions.