Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Nogueira RG, Jadhav AP, Haussen DC, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. N Engl J Med 2018;378:11-21. DOI: 10.1056/NEJMoa1706442

Contents

List of Investigators and Administrative Staff

Enrolling Clinical Centers (number recruited)

UPMC Presbyterian/Shadyside Hospital (46) -A. Jadhav (PI), A. Aghaebrahim, A. Al-Bayati, D. Campbell, A. Ducruet, D. Giurgiutiu, B. Gross, D. Gulati, M. Hammer, A. Horev, B. Jankowitz, T. Jovin, C. Kenmuir, K. Limaye, B. Molyneaux, V. Reddy, M. Rocha, H. Shoirah, M. Starr, C. Streib, P. Tadi, V. Totoraitis, G. Walker, L. Wechsler; **Emory University School of Medicine at Grady Memorial Hospital (38) -**D. Haussen (PI), L. Amerson, A. Anderson, S. Belagaje, N. Bianchi, M. Bouslama, R. Dharia, M. Frankel, M. Gebru, M. Grigoryan, J. Grossberg, L. Hernaiz, J. Hinton, V. Javalkar, A. Katema, S. Lahoti, A. Lima, N. Mahdi, R. Nogueira, V. Patel, S. Rangaraju, R. Vardaman, M. Winningham; **Riverside Methodist Hospital/ Ohio Health Research Institute (14) -**R. Budzik (PI), T. Davis, G. Eubank, W. Hicks, B. Katz, P. Pema, N. Vora; **Gui de Chauliac - CHRU de Montpellier (14) -**A. Bonafe (PI), C. Arquizan, X. Ayrignac, B. Carlander, M. Charif, V. Costalat, N. Gaillard, G. Gascou, C. Gény, I. Mourand, E. Omer, C. Riquelme Bareiro; **Texas Stroke Institute at HCA North Texas (10) -**P. Bhuva (PI), R. Gianatasio, P. Hansen, V. Janardhan, A. Venizelos, A. Yoo; **Jackson Memorial Hospital (9) -**D. Yavagal (PI), N. Asdaghi, N. Bhatt, J. Caplan, P. Khandelwal, A. Malik, E. Peterson, V. Saini, R. Starke, L. Torres; **Valley Baptist Medical Center - Harlingen (7) -**A. Hassan (PI), T. Wondwossen; **University Hospitals Case Medical Center (7) -**C. Sila (PI), D. Abdul-Rahman, K. Asi, N. Bambakidis, K. Blackham, M. DeGeorgia, A. Furlan, Y. Hu, V. Jadhav, Y. Mao, B. Miller, P. Murty, J. Nelson, J. Pace, C. Ramos Estebanez, A. Ray, J. Singer, S. Sundararajan, J. Sunshine, M. Talhama, R. Tarr, W. Xiong; **Baptist Jacksonville/Lyerly Neurosurgery (7) -**R. Hanel (PI), A. Aghaebrahim, E. Sauvageau; **Hospital Vall d'Hebron (7) -**M. Ribo (PI), S. Boned, P. Coscojuela, J. Juega, O. Maisterra, C. Molina, M. Muchada, J. Pagola, D. Rodriguez-Luna, Á. Rovira, M. Rubiera, E. Santamarina, A. Tomasello, C. Vert; **Hopital Purpan Toulouse (7) -**C. Cognard (PI), J. Albucher, F. Bonneville, F. Calvas, L. Calviere, M. Galitzky, A. Januel, C. Michelozzi, J. Olivot, N. Raposo, P. Tall, A. Tricoire, A. Viguier; **Buffalo General Medical Center (5) -**E. Levy (PI), J. Davies, A. Siddiqui, K. Snyder; **Rush University Medical Center (5) -**M. Chen (PI), R. Crowley, D. Lopes, R. Moftakhar; **California Pacific Medical Center (5) -**J. English (PI), W. Kim, N. Telischak; **The Royal Melbourne Hospital (5) -**P. Mitchell (PI), B. Campbell, S. Davis, R. Dowling, A. Meretoja, D. Shah, T. Wu, H. Zhao; **Hospital Germans Trias i Pujol (5) -**M. MillanTorne (PI), L. Abraira, J. Broto, T. Canento, C. Castaño, J. Ciurans, A. Crespo, P. Cuadras, L. Dorado, M. Gomis, M. Hernandez, E. López-Cancio, L. Martín, K. Mekkaoui, J. Munuera, E. Palomeras, M. Paré, A. Planas, P. Puyalto, N. Pérez de la Ossa, S. Remollo, D. Samaniego, A. Sorrentino, N. Zhu; **Abington Memorial Hospital (3) -**Q. Shah (PI), H. Choe, N. Herial, O. Kozak, L. MacKenzie; **Erlanger Health System/Chattanooga Center for Neurologic Research (2) -**T. Devlin (PI), B. Sapkota (PI), B. Baxter, J. Calvert, H. Hawk, S. Quarfordt, A. Sirelkhatim; **Memorial Healthcare System (2) -**B. Mehta (PI), G. Dimartini, H. Duong; **Toronto Western Hospital - University Health Network (2) -**F. Silver (PI), R. Agid, D. Alrasheed, L. Casaubon, M. Del Campo, R. Farb, C. Jaigobin, S. Jalini, T. Krings, V. Pereira, A. Pikula, J. Schaafsma; **Capital Health - Regional Medical Center (1) -**E. Veznedaroglu (PI), M. Binning, Z. Hakma, K. Liebman; **Kaiser Permanente Los Angeles Medical Center (1) -**L. Feng (PI), Z. Ajani, K. Chao, J. Kim, P. Mathena, J. Noroozi, N. Sangha; **JFK Neuroscience Institute at JFK Medical Center (1) -**J. Kirmani (PI), A. Kulhari, S. Mehta, M. Moussavi, S. Panezai, A. Patel; **Florida Hospital Neurosciences Institute (1) -**F. Hellinger (PI), I. Acosta, E. Allen, M. Bellew, R. Gandhi, S. Manda, W. Morgan; **The University of Kansas Medical Center Research Institute, Inc. (1) -**M. Abraham (PI), S. Billinger, C. Costa, K. Ebersole, G. Gronseth, K. Husman, A. LacKamp, C. Lechtenberg, E. Madarang, M. Mittol, A. Reeves, M. Rippee, L.

Rosterman, M. Rymer, C. Shipley, Y. Wong; **Hospital Universitari de Bellvitge (1) -**P. Cardona (PI), S. Aixut, L. Aja, R. Barranco, M. De Miquel, B. Lara, A. Luttich, P. Mora, H. Quesada;

Non-Enrolling Clinical Centers

Baptist Health Lexington (0) -C. Given (PI), B. Newman, C. Ramsey; **Christiana Hospital (0) -**S. Satti (PI), B. Albani, D. Black, V. Dechant, T. Eden, E. Kahn, P. Moftakhar, J. Nomura, J. Raser-Schramm, T. Sivapatham, W. Tai, G. Zoarski; **University of California Los Angeles (0) -**S. Starkman (PI), A. Archer, M. Bahr Hosseini, A. Chhabra, G. Duckwiler, L. Frischmann, N. Gonzalez, J. Hinman, R. Jahan, K. Malhotra, M. Nour, N. Rao, J. Saver, L. Sharma, P. Singh, V. Szeder, X. Tang, S. Tareshima, A. Tipirneni, B. Yoo; **Wellstar Kennestone Hospital (0) -**R. Gupta (PI), S. Dhungana, R. Hamilton, A. Johnson, A. Khaldi, A. Koganti, K. Owada; **St. Joseph Mercy - Oakland (0) -**A. Xavier (PI), A. Majjhoo; **Hospital Clínic - Barcelona (0) -**X. Urra (PI), S. Amaro, J. Blasco, Á. Chamorro, O. Chirife, L. Llull, A. López, J. Macho, N. Macías, V. Obach, A. Renú, S. Rudilosso, L. San Roman, M. Werner; **Hôpital Pierre Wertheimer - Hôpital Neurologique Lyon () -**F. Turjman (PI), R. Ameli, Y. Berthezene, L. Chamard, T. Cho, L. Derex, B. Gory, M. Hermier, P. Labeyrie, G. Louis-Tisserand, L. Mechtouff, N. Nighoghossian, E. Ong, R. Riva; **University of Calgary () -**M. Hill (PI), M. Goyal;

Steering Committee

T.G. Jovin (Co-Principal Investigator), R. Nogueira (Co-Principal Investigator), B. Baxter, D. Lopes, A. Bonafe, A. Furlan, R. Gupta, O. Jansen, V. Pereira, M. Ribo, J. Saver

Data Safety Monitoring Committee

W. Smith (Chair), D. Gress, S. Hetts, R. Lewis

Clinical Events Committee

T. Malisch, A. Rai, K. Sheth

Independent Statisticians

S. Berry (Berry Consultants), T. Graves (Berry Consultants)

Core Lab

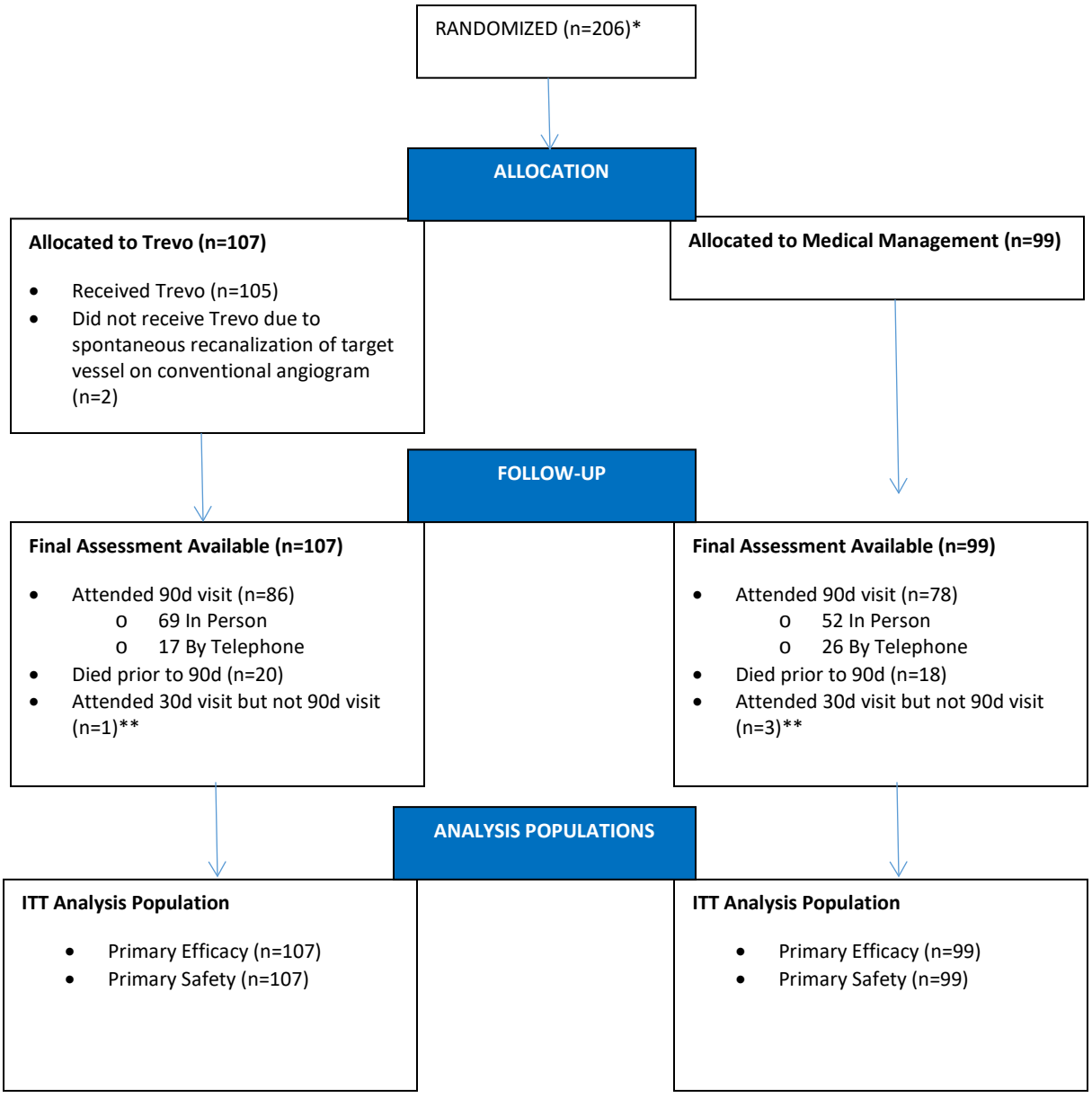
Neurovascular Imaging Research Core -D. Liebeskind

Inclusion and Exclusion Criteria

General Inclusion Criteria	1. Clinical signs and symptoms consistent with the diagnosis of an acute ischemic stroke, and subject belongs to one of the following subgroups: a. Subject has failed IV t-PA therapy (defined as a confirmed persistent occlusion 60 min after administration) b. Subject is contraindicated for IV t-PA administration 2. Age ≥18 3. Baseline NIHSS ≥10 (assessed within one hour of measuring core infarct volume) 4. Subject can be randomized between with 6 to 24 hours after time last known well 5. No significant pre-stroke disability (pre-stroke mRS must be 0 or 1) 6. Anticipated life expectancy of at least 6 months 7. Subject willing/able to return for protocol required follow up visits 8. Subject or subject's Legally Authorized Representative (LAR) has signed the study Informed Consent form* * If approved by local ethics committee and country regulations, the investigator is allowed to enroll a patient utilizing emergency informed consent procedures if neither the patient nor the representative or person of trust is available to sign the informed consent form. However, as soon as possible, the patient is informed and his/her consent is requested for the possible continuation of this research. (Not applicable to U.S. Sites.)
General Inclusion Criteria (additional information)	1. Subjects receiving heparin or low molecular weight (LMW) heparin e.g. Fragmin® (Dalteparin Sodium) or an intravenous direct thrombin inhibitor such as Angiomax® (Bivalirudin), or Argatroban within the last 24 hours from screening are eligible for participation if their coagulation profile remains acceptable. 2. Subjects on factor Xa inhibitors (e.g. apixaban) or direct thrombin inhibitors are eligible for participation
Imaging Inclusion Criteria	1. < 1/3 MCA territory involved, as evidenced by CT or MRI 2. Occlusion of the intracranial ICA and/or MCA-M1 as evidenced by MRA or CTA 3. Clinical Imaging Mismatch (CIM) defined as one of the following on MR-DWI or CTP-rCBF maps: a. 0-<21 cc core infarct and NIHSS ≥ 10 (and age ≥ 80 years old) b. 0-<31 cc core infarct and NIHSS ≥ 10 (and age < 80 years old) c. 31 cc to <51 cc core infarct and NIHSS ≥ 20 (and age < 80 years old)
General Exclusion Criteria	1. History of severe head injury within past 90 days with residual neurological deficit, as determined by medical history 2. Rapid improvement in neurological status to an NIHSS <10 or evidence of vessel recanalization prior to randomization 3. Pre-existing neurological or psychiatric disease that would confound the neurological or functional evaluations, e.g. dementia with prescribed anti-cholinesterase inhibitor (e.g. Aricept) 4. Seizures at stroke onset if it makes the diagnosis of stroke doubtful and precludes obtaining an accurate baseline NIHSS assessment 5. Baseline blood glucose of <50mg/dL (2.78 mmol) or >400mg/dL (22.20 mmol) 6. Baseline hemoglobin counts of <7 mmol/L 7. Baseline platelet count < 50,000/uL 8. Abnormal baseline electrolyte parameters as defined by sodium concentration <130 mmol/L, potassium concentration <3 mEq/L or >6 mEq/L 9. Renal failure as defined by a serum creatinine >3.0 mg/dL (264 µmol/L)

	NOTE: subjects on renal dialysis may be treated regardless of serum creatinine levels 10. Known hemorrhagic diathesis, coagulation factor deficiency, or on anticoagulant therapy with INR > 3.0 or PTT > 3 times normal. Patients on factor Xa inhibitor for 24-48 hours ago must have a normal PTT. 11. Any active or recent hemorrhage within the past 30 days 12. History of severe allergy (more than rash) to contrast medium 13. Severe, sustained hypertension (Systolic Blood Pressure >185 mmHg or Diastolic Blood Pressure >110 mmHg) NOTE: If the blood pressure can be successfully reduced and maintained at the acceptable level using medication the subject can be enrolled 14. Female who is pregnant or lactating at time of admission 15. Current participation in another investigational drug or device study 16. Presumed septic embolus, or suspicion of bacterial endocarditis 17. Treatment with any cleared thrombectomy devices or other intra-arterial (neurovascular) therapies prior to randomization
Exclusion Criteria (additional information)	1. The "correction" of baseline glucose or coagulation laboratory values to meet inclusion criteria will not be allowed. 2. Subjects who have taken Clopidogrel, aspirin, or both within the last 24 hours from screening for the trial should not be excluded if their coagulation profile remains acceptable. 3. Subjects with a questionable seizure at onset of stroke should not be excluded if CTA/MRA confirms the presence of intracranial ICA and/or M1 occlusion, and accurate NIHSS can be obtained.
Imaging Exclusion Criteria	1. Evidence of intracranial hemorrhage on CT/MRI 2. CTA or MRA evidence of flow limiting carotid dissection, high-grade stenosis, or complete cervical carotid occlusion requiring stenting at the time of the index procedure (i.e., mechanical thrombectomy). 3. Excessive tortuosity of cervical vessels on CTA/MRA that would likely preclude device delivery/deployment 4. Suspected cerebral vasculitis based on medical history and CTA/MRA 5. Suspected aortic dissection based on medical history and CTA/MRA 6. Intracranial stent implanted in the same vascular territory that would preclude the safe deployment/removal of the Trevo device 7. Occlusions in multiple vascular territories (e.g., bilateral anterior circulation, or anterior circulation/vertebrobasilar system) as confirmed on CTA/MRA, or clinical evidence of bilateral strokes or strokes in multiple territories 8. Significant mass effect with midline shift as confirmed on CT/MRI 9. Evidence of intracranial tumor (except small meningioma) as confirmed on CT/MRI

Figure S1 Consort Diagram



*Stratified by:

- Clinical Infarct Mismatch (CIM) Group A: Age ≥ 80, Core <21cc, NIHSS ≥ 10
- Clinical Infarct Mismatch (CIM) Group B: Age < 80, Core <31cc, NIHSS ≥ 10
- Clinical Infarct Mismatch (CIM) Group C: Age < 80, Core <31 to <51cc NIHSS ≥ 20

**In Person mRS and Used to impute 90 day outcomes with Bayesian multiple imputation

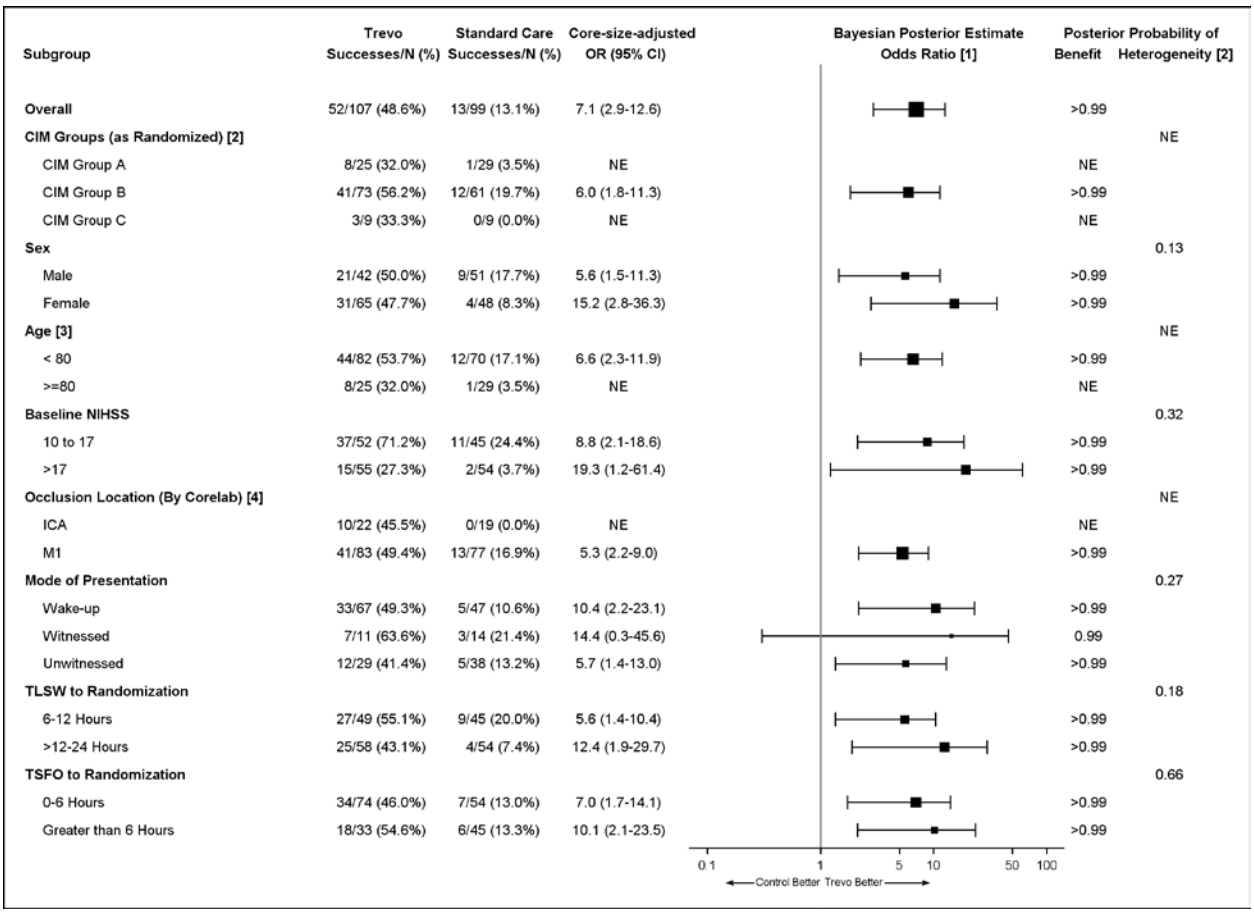
Table S1 Additional Patient Presentation and Workflow Data

Variable	Intervention (n=107)	Control (n=99)
Systolic blood pressure– mm Hg –mean ± SD	147 ± 22.5	152.0 ± 23.7
Glucose level, mg/dL – mean ± SD	125.4 ± 29.1	147.7 ± 65.0
Premorbid mRS 0 (no symptoms) – no. (%)	99 (92.5)	96 (97.0)
Antiplatelet use during hospitalization – no (%)	85(79)	86(87)
Baseline Core by Site RAPID (cc) - Median [IQR]	7.6 (2.0, 18.0)	8.9 (3.0, 18.1)
Qualifying RAPID volume obtained by CTP– no. (%)	67 (62.6)	64 (64.6)
Qualifying RAPID volume obtained by DWI MRI– no. (%)	40 (37.4)	35 (35.4)
Cervical carotid stenosis– no. (%)		
0-50%	80 (74.8)	72 (72.7)
51-99%	12 (11.2)	14 (14.1)
100% (occlusion)	15 (14.0)	13 (13.1)
Workflow times		
TLSW to eligibility imaging – hours ∫ Median [IQR] 6-12 hours – no. (%) 12.1-24 hours – no. (%)	 11.9 (9.2, 15.8) 55 (51.4) 52 (48.6)	 12.8 (8.5, 15.2) 46 (46.5) 53 (53.5)
TLSW to randomization – hours Median [IQR] 6-12 hours – no. (%) 12.1-24 hours – no. (%)	 12.2 [10.2, 16.3] 50 (46.7) 57 (53.3)	 13.3 [9.4, 15.8] 46 (46.5) 53 (53.5)
TSFO to randomization – hours, Median [IQR] TSFO to randomization ≤6 h - no. (%) TSFO to randomization >6 h - no. (%)	4.8 [3.6, 6.2] 74 (69.2) 33 (30.8)	5.6 [3.6, 7.8] 54 (54.6) 45 (45.5)
Time from hospital arrival to randomization-min, Median [IQR]	84 [52, 123]	81 [56, 111]
Time from hospital arrival to arterial puncture-min, Median [IQR]	109 [76, 150]	NA
Time from qualifying imaging to arterial puncture -min, Median [IQR]	57[36,84]	NA
Time from randomization to arterial puncture –min, Median [IQR]	16[9,29]	
Time last seen well to arterial puncture-hours, Median [IQR]	12.8 (10.6, 16.7)	NA
Time last seen well to revascularization (TICI 2b-3) – hours		

Variable	Intervention (n=107)	Control (n=99)
Median [IQR]	13.6 [11.3,18.0]	NA

Additional Efficacy Results

Figure S2 Bayesian estimated odds ratios for functional independence as measured by 90 day mRS overall and by subgroup.



†Rates of 90-day functional independence provided for descriptive purposes.

[1] NE represents not estimable. Due to small subgroup sample size and lack of functional independence in the control group (e.g. 0/9 in CCM Group C) the mathematical estimates were not clinically meaningful.

[2] Probability of benefit is the posterior probability of treatment benefit of Trevo thrombectomy. Probability of Heterogeneity is the probability of an interaction between the sub-groups and the treatment benefit. A probability greater than 0.975 or less than 0.025 can be considered a significant interaction.

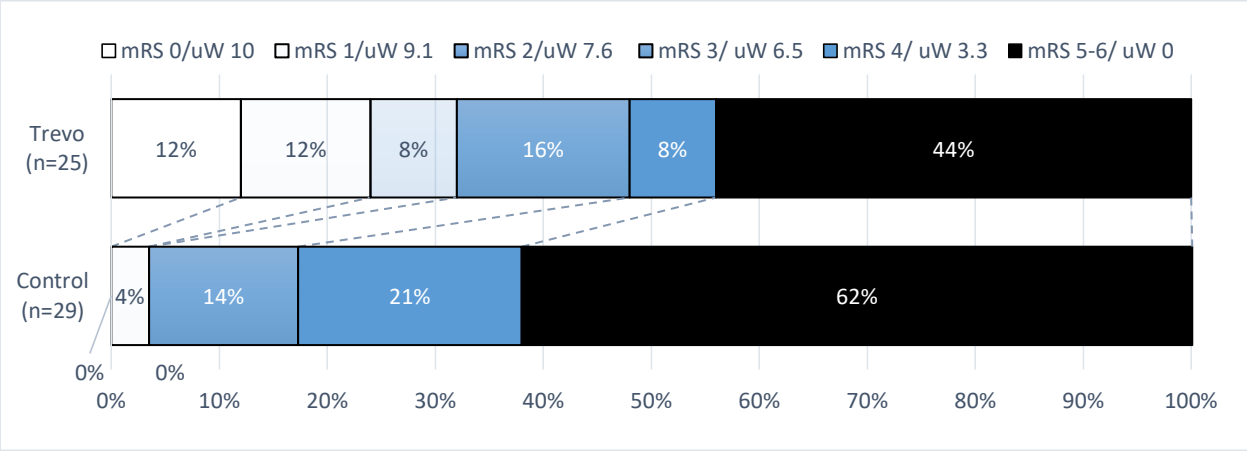
[3] Protocol defined Clinical Infarct Mismatch (CIM) groups as follows: CIM Group A: Age ≥ 80, NIHSS ≥ 10, Core <21cc, CIM Group B: Age < 80, NIHSS ≥ 10, Core <31cc, CIM Group C: Age < 80, NIHSS ≥ 20, Core <51cc.

[4] Age Greater than or equal to 80 and Clinical Infarct Mismatch (CIM) group A are identical.

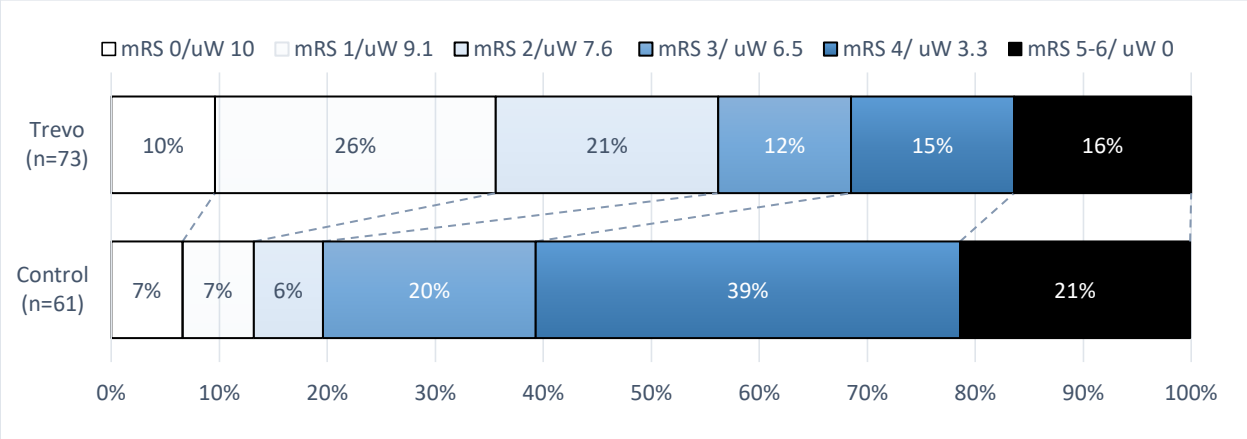
TLSW = Time last seen well ; TSFO = Symptom First Observed to Randomization

Figure S3 Disability Outcomes at 90 Days in the Two Treatment Arms, in Patients Stratified by Clinical Infarct Mismatch (CIM) subgroup. There is no evidence of heterogeneity of treatment effect between these subgroups (prob =0.47, Bayesian Posterior Probability of Heterogeneity). Each stacked bar in the figure is also a choropleth map: the degree of blue of each bar segment indicates the precise value of that segment's UW-mRS score.

Clinical Infarct Mismatch (CIM) Group A (Age ≥80, NIHSS ≥10, Core <21cc)



Clinical Infarct Mismatch (CIM) Group B (Age< 80, NIHSS ≥10, Core <31cc)



Clinical Infarct Mismatch (CIM) Group C (Age< 80, NIHSS ≥20, Core <31-51cc)

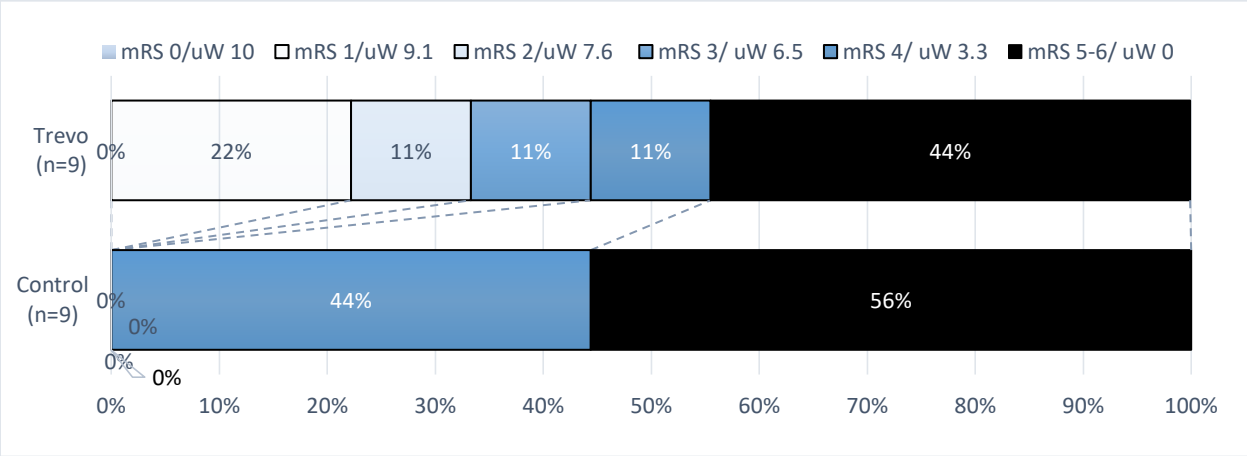
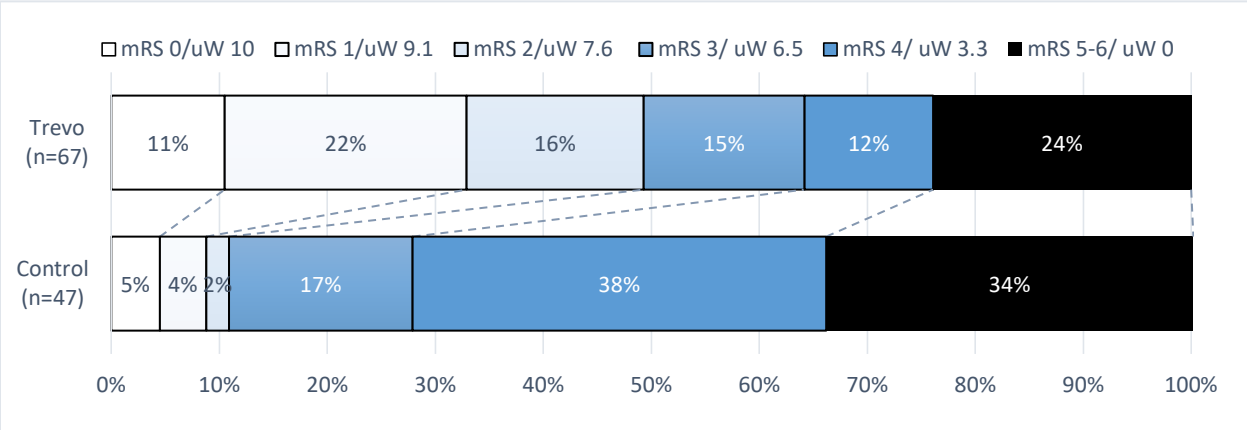
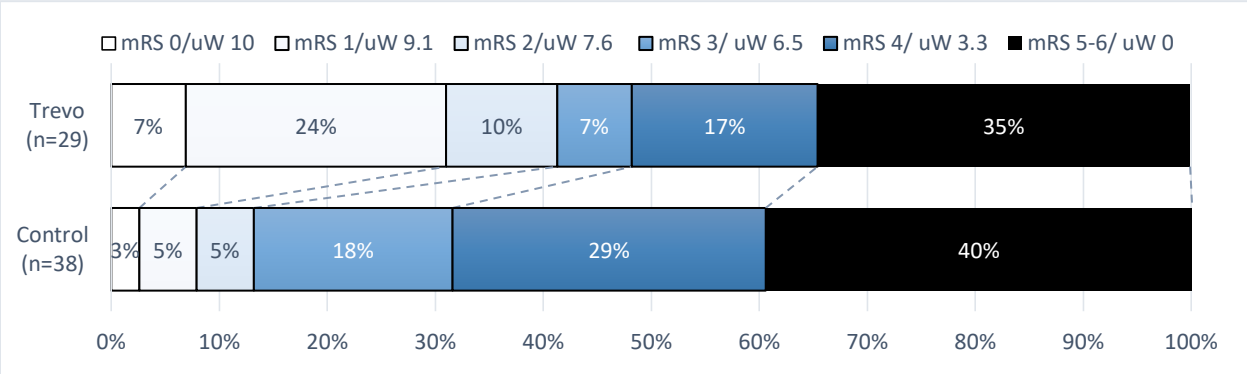


Figure S4 Disability Outcomes at 90 Days in the Two Treatment Arms, in Patients Stratified by Mode of Presentation. There is no evidence of heterogeneity of treatment effect between these subgroups (prob =0.21, Bayesian Posterior Probability of Heterogeneity). Each stacked bar in the figure is also a choropleth map: the degree of blue of each bar segment indicates the precise value of that segment's UW-mRS score.

Wake up



Unwitnessed



Witnessed

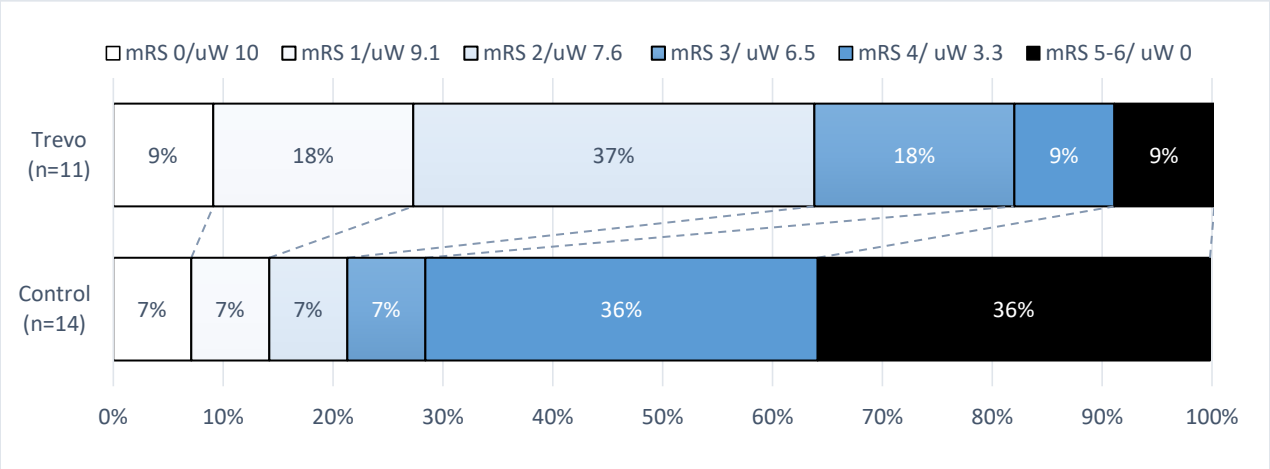
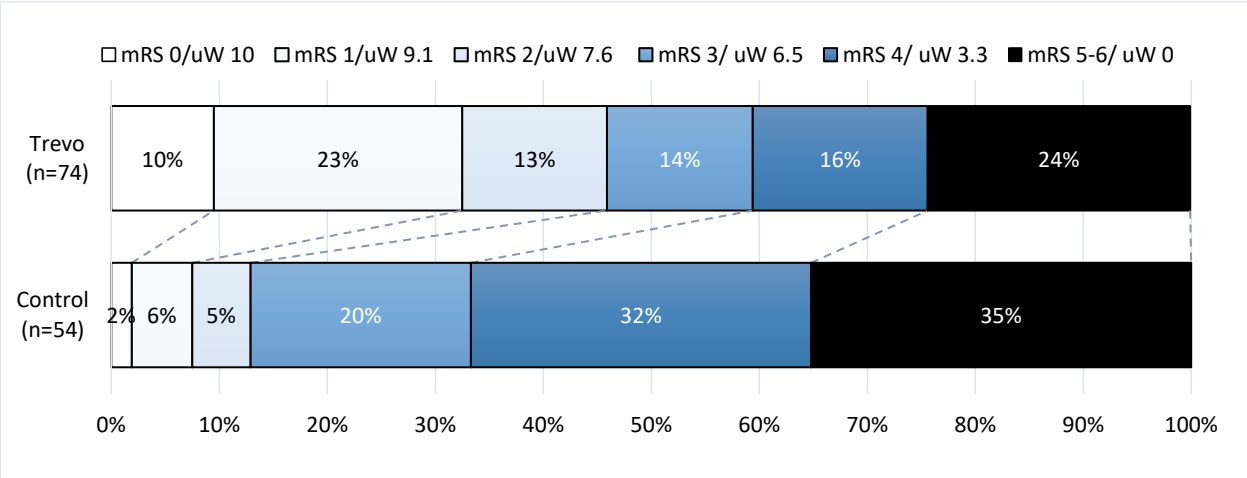


Figure S5 Disability Outcomes at 90 Days in the Two Treatment Arms, in Patients Stratified by Time Symptom First Observed to Randomization. There is no evidence of heterogeneity of treatment effect between these subgroups (prob =0.70, Bayesian Posterior Probability of Heterogeneity). Each stacked bar in the figure is also a choropleth map: the degree of blue of each bar segment indicates the precise value of that segment's UW-mRS score.

Symptom First Observed ≤ 6 hours



Symptom First Observed > 6 hours

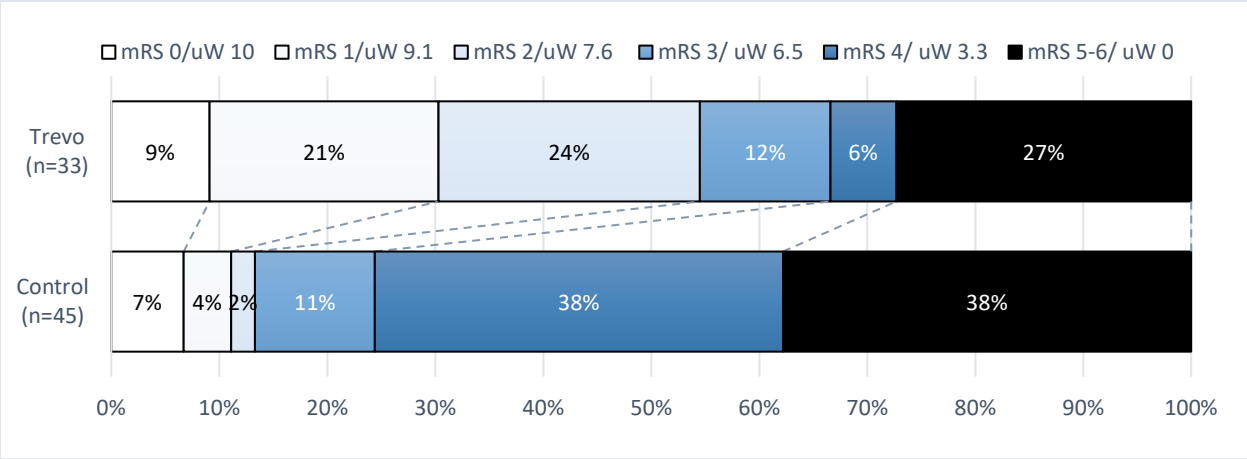
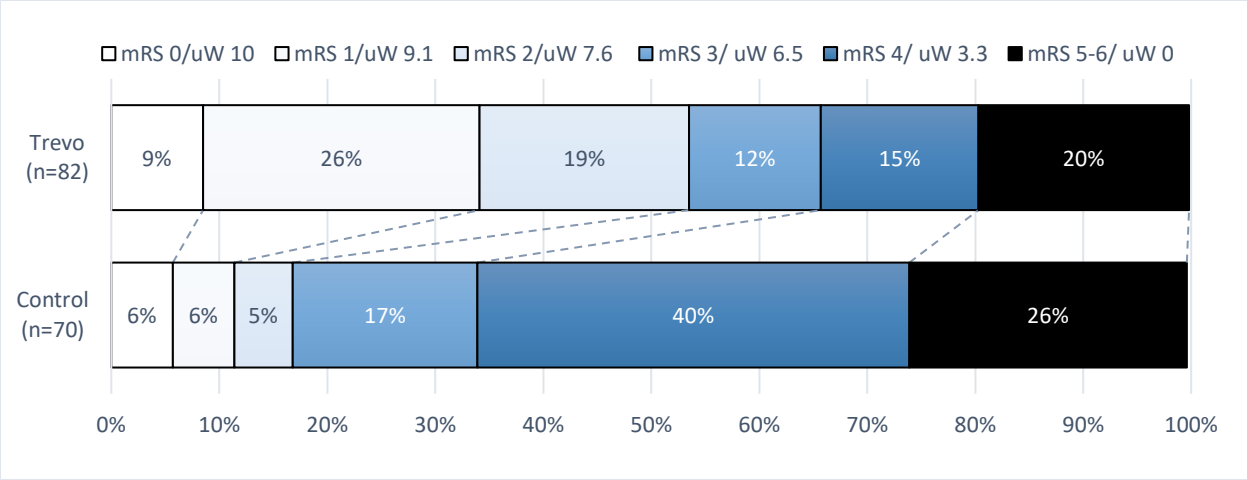


Figure S6 Disability Outcomes at 90 Days in the Two Treatment Arms, in Patients Stratified by Age. There is no evidence of heterogeneity of treatment effect between these subgroups (prob =0.42, Bayesian Posterior Probability of Heterogeneity). Each stacked bar in the figure is also a choropleth map: the degree of blue of each bar segment indicates the precise value of that segment's UW-mRS score.

Age <80 years



Age ≥ 80 years

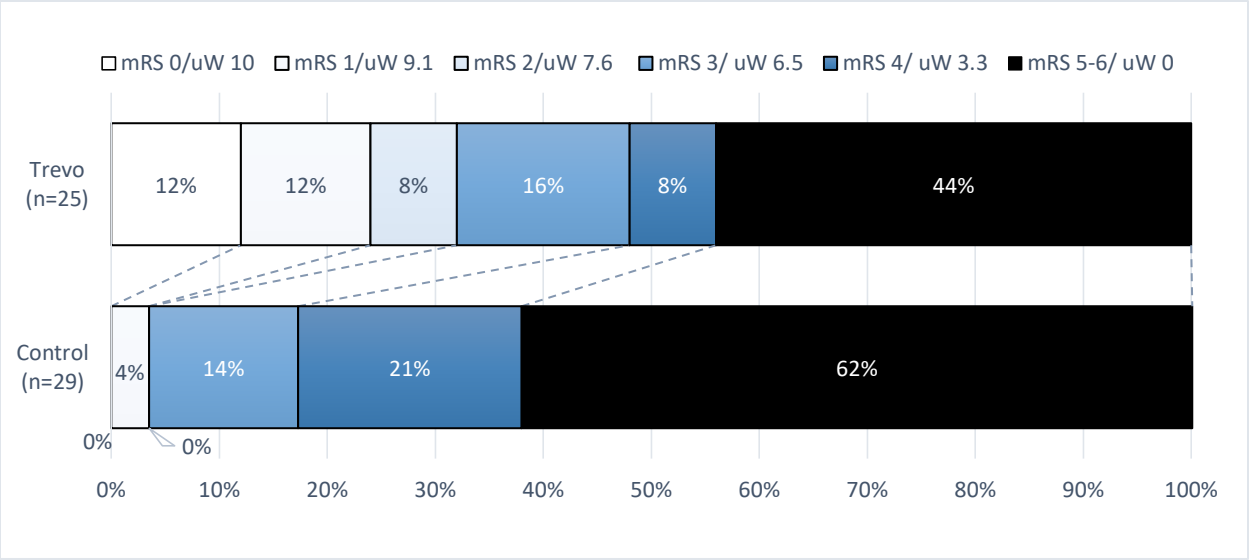
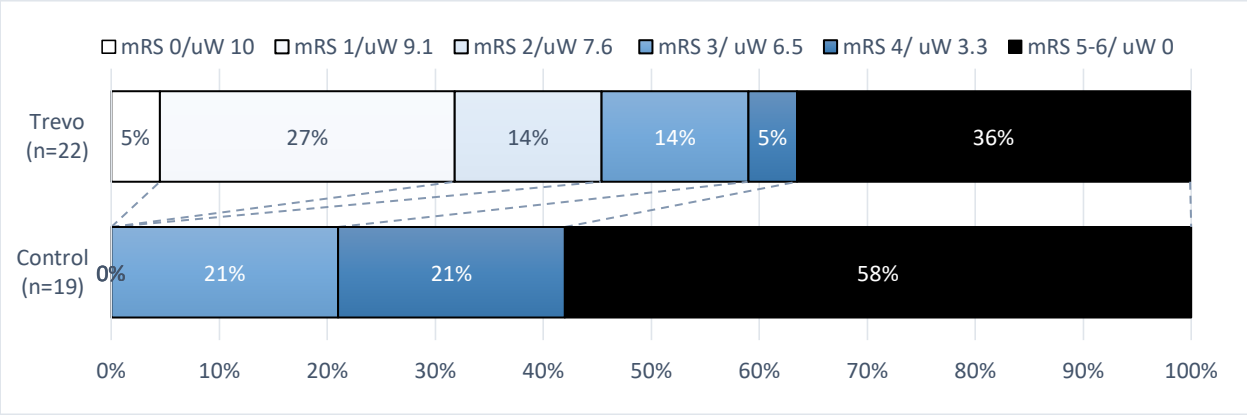
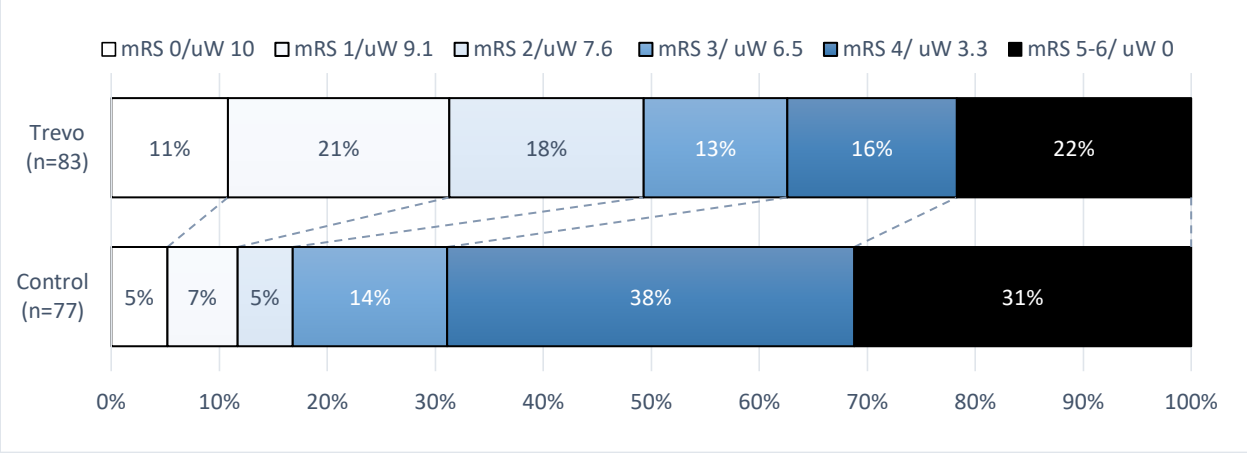

mRS 0/uW 10
mRS 1/uW 9.1
mRS 2/uW 7.6
mRS 3/ uW 6.5
mRS 4/ uW 3.3
mRS 5-6/ uW 0
Trevo (n=25)
12%
12%
8%
16%
8%
44%
Control (n=29)
4%
14%
21%
62%
0%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

Figure S7 Disability Outcomes at 90 Days in the Two Treatment Arms, in Patients Stratified by Imaging Core Laboratory Reported Target Occlusion Location. There is no evidence of heterogeneity of treatment effect between these subgroups (prob =0.26, Bayesian Posterior Probability of Heterogeneity). Each stacked bar in the figure is also a choropleth map: the degree of blue of each bar segment indicates the precise value of that segment's UW-mRS score.

Internal Carotid Artery (ICA)



Middle Cerebral Artery M1 Segment



Middle Cerebral Artery M2 Segment

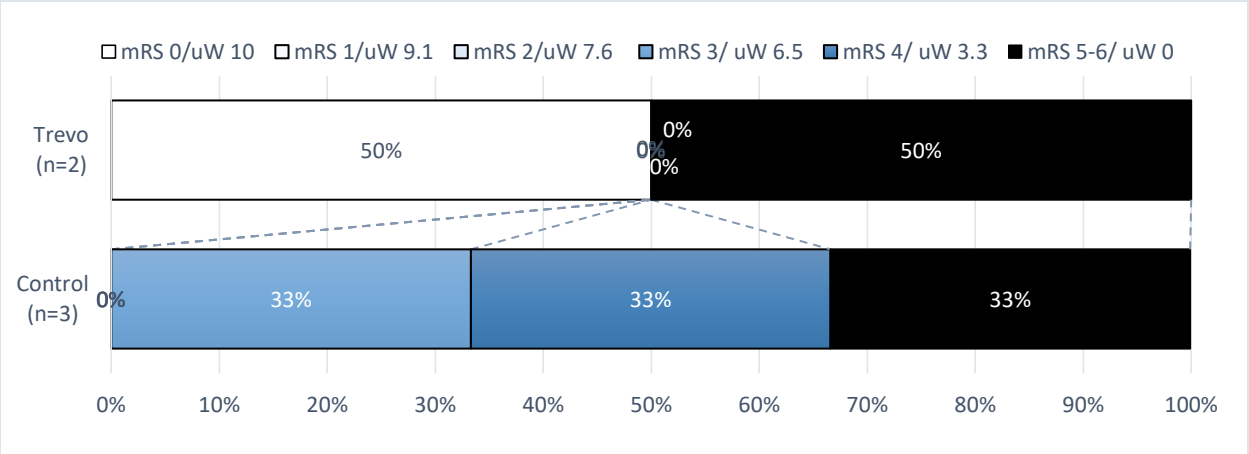
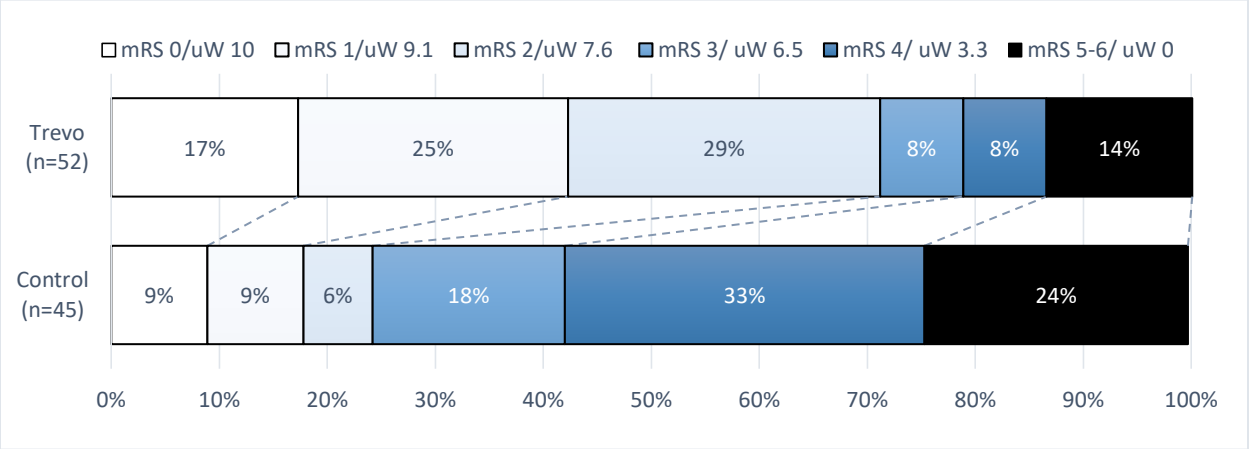
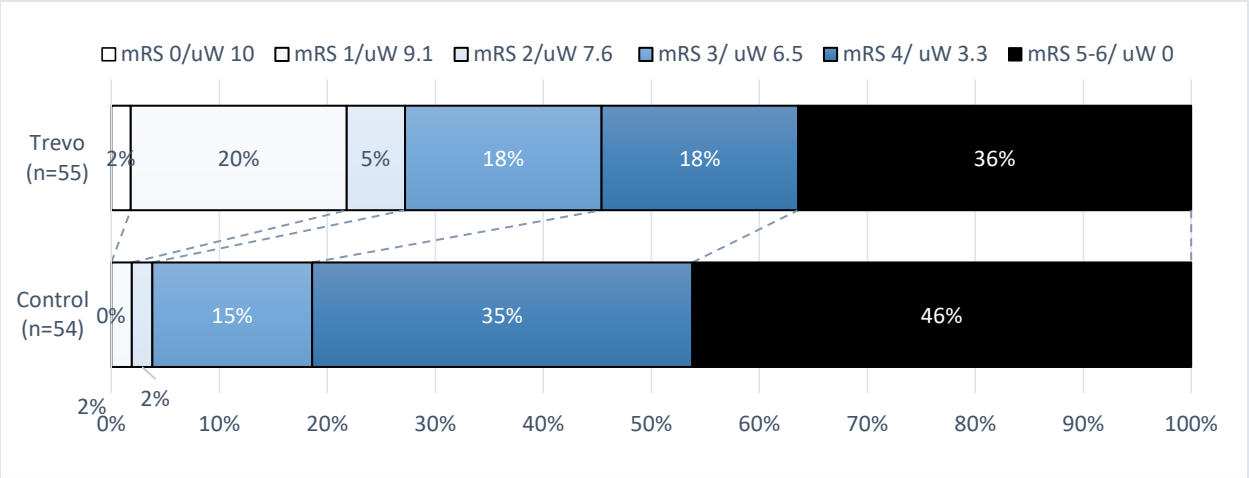


Figure S8 Disability Outcomes at 90 Days in the Two Treatment Arms, in Patients Stratified by Stroke Severity. There is no evidence of heterogeneity of treatment effect between these subgroups (prob =0.71, Bayesian Posterior Probability of Heterogeneity). Each stacked bar in the figure is also a choropleth map: the degree of blue of each bar segment indicates the precise value of that segment's UW-mRS score.

Baseline NIHSS 10-<17



Baseline NIHSS ≥ 17



Additional Safety Results

Table S2 Site Reported Device and Procedure –Related Serious Adverse Events (SAEs) and Non Serious Adverse Events (AEs) in the Investigational Arm (n=107).

	Treatment arm N= 107			
System Organ Class/Preferred Term	**SAEs**	**Subjects with SAEs**	**Non SAEs**	**Subjects with Non SAEs**
Any Adverse Event(AE)	5	5 (4.7%)	31	26 (24.3%)
Blood and lymphatic system disorders	0	0	1	1 (0.9%)
Anaemia	0	0	1	1 (0.9%)
Cardiac disorders	1	1 (0.9%)	0	0
Pulmonary oedema	1	1 (0.9%)	0	0
Injury, poisoning and procedural complications	1	1 (0.9%)	1	1 (0.9%)
Nephropathy toxic	0	0	1	1 (0.9%)
Vascular pseudoaneurysm	1	1 (0.9%)	0	0
Nervous system disorders	3	3 (2.8%)	20	18 (16.8%)
Brain oedema	1	1 (0.9%)	0	0
Cerebral artery embolism	0	0	3	3 (2.8%)
Cerebral ischaemia	0	0	1	1 (0.9%)
Haemorrhage intracranial	0	0	3	3 (2.8%)
Haemorrhagic transformation stroke	1	1 (0.9%)	8	8 (7.5%)
Headache	0	0	2	2 (1.9%)
Hydrocephalus	0	0	1	1 (0.9%)
Neurological decompensation	0	0	1	1 (0.9%)

	Treatment arm N= 107			
System Organ Class/Preferred Term	**SAEs**	**Subjects with SAEs**	**Non SAEs**	**Subjects with Non SAEs**
Subarachnoid haemorrhage	1	1 (0.9%)	1	1 (0.9%)
Renal and urinary disorders	0	0	1	1 (0.9%)
Renal failure	0	0	1	1 (0.9%)
Respiratory, thoracic and mediastinal disorders	0	0	1	1 (0.9%)
Lung disorder	0	0	1	1 (0.9%)
Vascular disorders	0	0	7	7 (6.5%)
Hypotension	0	0	2	2 (1.9%)
Vascular dissection	0	0	1	1 (0.9%)
Vasospasm	0	0	3	3 (2.8%)
Vessel puncture site haemorrhage	0	0	1	1 (0.9%)

[1] Inclusive of site reported Serious and Non-Serious Adverse Events. Device or procedure relatedness per site reported

Table S3 Core Lab Adjudicated Intracranial Hemorrhages at 24 hours Post Stroke

24 hour Intracranial Hemorrhage Type	Treatment	Control	Difference
HI-1	43.0% (46/107)	24.2% (24/99)	18.8% [5.8%, 30.6%]
HI-2	9.4% (10/107)	8.1% (8/99)	1.3% [-7.0%, 9.3%]
PH-1	0.0% (0/107)	0.0% (0/99)	0.0% [0.0%, 0.0%]
PH-2	1.9% (2/107)	1.0% (1/99)	0.9% [-3.8%, 5.6%]
RIH	0.0% (0/107)	0.0% (0/99)	0.0% [0.0%, 0.0%]
IVH	0.0% (0/107)	0.0% (0/99)	0.0% [0.0%, 0.0%]

24 hour Intracranial Hemorrhage Type	Treatment	Control	Difference
Subdural	0.0% (0/107)	0.0% (0/99)	0.0% [0.0%, 0.0%]
Epidural	0.0% (0/107)	0.0% (0/99)	0.0% [0.0%, 0.0%]
SAH	0.9% (1/107)	0.0% (0/99)	0.9% [-2.9%, 5.9%]

Hemorrhagic Transformations (HT) was categorized according to radiological criteria used in the European Cooperative Acute Stroke Study (ECASS) II trial.[1] In cases of >1 hemorrhagic lesion on CT scan, the worst possible HT category was assumed.

Hemorrhagic Infarction type 1 (HI-1): small petechiae along the margins of the infarct.
Hemorrhagic Infarction type 2 (HI-2): more confluent petechiae within the infarcted area but without space-occupying effect.
Parenchymal Hematoma type 1 (PH-1): hematoma in ≤30% of the infarcted area with some slight space-occupying effect.
Parenchymal Hematoma type 2 (PH-2): dense hematoma >30% of the infarcted area with substantial space-occupying effect or as any hemorrhagic lesion outside the infarcted area.
Remote Intracranial Hemorrhage (RIH)
Intraventricular Hemorrhage (IVH)
Subarachnoid Hemorrhage (SAH)

Supplemental Text Section S1: Number Needed to Treat Derivation Methods and Results

Appropriate applications of clinical trial results to clinical practice are facilitated by effectively communicating to clinicians, patients, and payors the magnitude and characteristics of the treatment effect observed in the study. Ideally, a trial's primary endpoint should capture all of the effects of the intervention in proportion to the degree they are valued by the patient. Because acute stroke is a condition that incapacitates as well as kills, clinically important outcome states extend over a continuum from fully normal, over the entire range of impairment and disability, to death. Consequently, in acute ischemic stroke, outcomes are intrinsically non-binary and most commonly are assessed using: 1) ordinal scales that partition patients among levels that are ordered, but of uncertain distance from one another, or 2) quantitative scales that assign patients precise positions along an outcome dimension, using continuous or discrete numeric values.[2-4] In the DAWN trial, the primary measure of neurologic outcome for patients was captured by the widely-accepted modified Rankin Scale (mRS) that allocates patients to

7 disability levels, and the numerical disability value of the each of the mRS disability outcome states was quantified in a patient-centered manner by assigning each disability level a discrete value, derived from both: 1) the disability-weighting method of the World Health Organization Global Burden of Disease Project, indexing health provider assessments of disability state value, and 2) the utility-weighting method, a standard health outcome measure, indexing patient assessments of disability state value.[5,6]

The resulting score on the utility-weighted modified Rankin Scale (UW-mRS) indicates the value to patients, families, and health providers of the long-term disability outcome, across a broad range from fully normal, through varying degrees of disability, to death.[5,6] In contrast, the original ordinal mRS scale only ranks patients in levels of disability, but does not capture or quantify the differential values of transitions from one mRS level to the next.[7] For example, the difference in degree of disability for a patient of a 1 level change in the final mRS outcome varies substantially for different steps in the mRS. The value of a 1 step change from an mRS level of 4 to an mRS level of 3 is marked, 3.2 on a scale of 0-10, representing a change of 32% of the entire disability range from normal to death, whereas the value of a 1 step change from an mRS level of 2 to an mRS level of 1 is of less value, 1.5 on a scale of 0-10, representing 15% of the entire disability range from normal to death.[7]

In this Supplement section, to facilitate clinically meaningful interpretations of the benefit associated with thrombectomy, we provide methods and results for the derivation of number needed to treat (NNT) indices of the benefit observed in DAWN on the primary UW-mRS endpoint, analogous to the traditional number needed to treat to benefit (NNTB) for coarser binary endpoints. The analyses provide two types of NNT values. The first is the number needed to treat to benefit (NNTB), derived for the quantified UW-mRS outcomes over the entire disability range, rather than just for a single, binary, disability level transition. The NNTB is commonly used to communicate the magnitude of benefit of a new therapy to clinicians.[8,9] The NNTB is the number of patients on average who must receive a superior

therapy to confer one additional favorable outcome. Deriving an NNTB value from a parallel group clinical trial, such as DAWN, with a continuous outcome, reflecting the full range of disability transitions important to patients, rather than a binary outcome reflecting only one among the many disability transitions import to patients, requires making plausible assumptions regarding the pattern of benefit within the treated population. In addition, we describe methods and results for a related metric, the number needed to treat to save one life (NNTSOL). The NNTSOL is the number of patients who, on average, must receive a superior therapy for the accrued benefit to equal saving one life with return to fully normal function.

The Proportion of Patients Who Benefit from Thrombectomy

The NNTB is commonly used to communicate the magnitude of benefit of a new therapy to clinicians, and indicates the average number of patients who must receive a superior therapy to confer benefit to a single additional patient. However, derivation of a specific NNTB based on a continuous scale like UW-mRS, or an ordinal scale like the unweighted, 7-level mRS, and the results of a parallel group design clinical trial requires knowing granular aspects of the distribution of treatment effects at the individual patient level that are not fully specified by the group-level distribution of outcomes available from the trial. Specifically, one must know the proportion of responders and non-responders among treated patients, and the relative frequencies of small, intermediate, and large treatment benefits among the responders (Supplemental Figure S1). Technically, this individual patient – level aspect of the treatment distribution is the "within-patient correlation," meaning the correlation between the outcome the patient would have experienced with the control treatment and the outcome the same patient would have experienced with the experimental treatment. In parallel group trials, individual outcomes of patients are known only for the single treatment arm to which they were assigned, and not for the treatment arm to which they were not allocated. In contrast, in crossover

design trials, when there is no carryover effect, individual patient-level outcomes are known under both experimental and control therapies. But for single episode, acute conditions, like stroke, crossover trials cannot be conducted. Parallel group trial data indicate the total amount of benefit accruing to the treatment group, but do not specify if this benefit arises from many individuals each improving a little bit, from a few individuals each improving a lot, or from an intermediate pattern of benefit.

For a treatment like thrombectomy, it is a biologic near-certainty that the benefit seen in the intervention arm at the group level is not experienced by every intervention patient equally. For example, in the DAWN intervention arm, 16% of patients did not have substantial reperfusion (mTICI 2b-3) achieved by thrombectomy; in these patients, the biological mechanism by which thrombectomy confers benefit was reduced or completely absent. In another subset of intervention arm patients, pre-treatment core ischemic injury, although modest in volume, likely was strategically located in eloquent brain regions, precluding substantial functional recovery despite reperfusion and salvage of other threatened areas. In a small subset of intervention patients, thrombectomy was associated with symptomatic hemorrhage, occurring nominally 1.5% more often in interventional patients in DAWN, which would tend to produce a worse rather than better outcome.

Although exact specification of the number needed to treat to benefit for continuous and ordinal outcomes is not possible in parallel design trials, several techniques are available that allow informative estimates of the NNTB values, based directly upon the available trial results. These techniques each make simplifying assumptions and the results must be interpreted with these assumptions in mind.

For a quantified, continuous or discrete, outcome measure, the first step for NNTB derivation is to define the minimal clinically important difference (MCID), the least amount of change in the measure that is of value to patients.[10] Changes in the quantified measure below this threshold will not change the health state outcome of the patient in a clinically important manner, and so patients with such changes

will not be counted as benefitting from intervention. Conversely, changes at or above this threshold are clinically meaningful, and so patients with such changes will be counted as experiencing treatment benefit. For the outcomes using utility weighting, the MCID has been established as 0.03 on a scale from 0 to 1, equivalent to 0.3 on the DAWN UW-mRS scale of 0-10.[11,12] Patients whose UW-mRS score at 3 months is 0.3 or more higher with thrombectomy, compared to what would have been achieved without, have benefitted from therapy.

With the MCID defined, the next step for NNTB derivation is to determine the proportion of patients whose benefit from treatment exceeded the MCID. The moderate granularity of the UW-mRS affects this analysis. Other, more detailed, disability and utility outcome measures can provide much finer gradations of outcome among patients, including generic disability measures (e.g. Academic Medical Center Linear Disability Scale),[13] generic utility measures (e.g. SF-36, ED-5D),[6,13,14] and neurologic disease-specific utility measures (e.g NINDS-QoL).[15] But on the UW-mRS, individual patients can only move in a discrete, quantized manner along the continuous utility dimension, and distances between steps have different magnitudes. One of the 6 mRS state transitions does not rise to the MCID threshold once the UW-mRS level weights are applied: the transition from mRS 6 to mRS 5 has a net disability change value of 0. All other single-step changes on the mRS surpass the MCID threshold after level weight application, with the lowest being the transition from mRS 1 to mRS 0, with a net disability change value of 0.9, and the highest being the transition from mRS 5 to mRS 4, with a net disability change value of 3.3. As a result, calculating the proportion of patients whose final outcomes on the UW-mRS are better in a clinically meaningful way, requires deriving the proportion of patients who improve by one or more health state transitions on the original mRS, but combining mRS levels 5 and 6 into a single worst outcome category so as to not count a one-step mRS 6 to 5 transition as being clinically meaningful. The resulting derivation is mathematically equivalent to deriving an NNTB for a 6-level ordinal scale.

There are three classes of techniques for such derivations: (1) automated derivation of an ordinal NNTB; (2) expert-informed derivation of an ordinal NNTB; and (3) automated derivation of a dichotomous NNTB. Each requires assumptions. The third approach, automated derivation of a dichotomous NNTB, requires assuming that only a single health state transition along the entire spectrum of disability is of value to patients and providers. As this assumption is obviously incorrect, dichotomous approaches to NNTB estimation are the least useful approach (Supplementary Table S1). Expert-informed deviation of an ordinal NNTB is superior to other approaches when benefits cluster in extremely small or extremely large increments, but draws upon content-expert input in addition to observed group outcome constraints, so is not solely data-driven. Accordingly, automated derivation of an ordinal NNTB is an important alternative approach.

Techniques for automated derivation of an ordinal NNTB include: (1) joint outcome table multiple resampling;[16,17] (2) joint outcome table algorithmic min-max specification;[16,17]and (3) the permutation test method.[18] All of these ordinal NNTB derivation techniques are both fully automated and handle the benefits of therapy as occurring in only a proportion of treatment group patients and being variable in degree amongst them, rather than being uniformly distributed among all treatment group patients. However, all these techniques assume that, among the universe of possible distributions of the group benefit to individual patients, the actual distribution falls toward the middle of the possible range. Among the automated techniques, the joint outcome table multiple resampling approach has the advantage of employing the least assumptions, and is preferred when its greater computational intensity is not an obstacle.

Accordingly, for DAWN, the joint outcome multiple resampling derivation method was applied to derive the NNTB for the primary UW-mRS endpoint, using automated random sampling as previously described.[17] Within the 6 by 6 joint distribution matrix with fixed marginal frequencies adding to 1000, there are 21 unknown frequencies subject to 10 constraints (5 columns, 5 rows). The constraints are

defined by the observed results of the parallel group clinical trial. The vector "x" is a 21-dimension vector that satisfies the 10 constraints. This is a linear inequality system of the form Ex=f, where x>0 and the 10 by 21 E matrix and the 10 by 1 f vector are obtained from the constraints. Because computing all possible values of x is not practical, we used the R function "xsample" in the R program library "limSolve" to randomly sample values of the vector "x" from all possible joint outcome table solutions, under the assumption that all possible solutions are equally likely. Using the mirror algorithm, 3000 samples were taken from the large population of all possible solutions, without replacement. The mean and range NNTB values of these random samples from all possible NNTB values under the constraints were analyzed.

The resulting NNTB was 2.0 (SD±0.3) (Supplementary Table S1). This value indicates that, assuming an intermediate dispersion of benefit degree among responding patients in DAWN, for every 2.0 patients treated with thrombectomy rather than medical care alone, 1 additional patient would have a clinically meaningful improvement in degree of disability at 90 days.

Table S2. Number Needed to Treat to Benefit [NNTB] Values Derived from DAWN Results

Outcome	Weights Applied to mRS Levels							Number Needed to Treat for One Patient to Benefit
	mRS 0	mRS 1	mRS 2	mRS 3	mRS 4	mRS 5	mRS 6	
Dichotomized 0 vs 1-6	1	0	0	0	0	0	0	18.8
Dichotomized 0-1 vs 2-6	1	1	0	0	0	0	0	4.4
Dichotomized 0-2 vs 3-6*	1	1	1	0	0	0	0	2.8
Dichotomized 0-3 vs 4-6	1	1	1	1	0	0	0	3.1
Dichotomized 0-4 vs 5-6	1	1	1	1	1	0	0	9.0
Dichotomized 0-5 vs 6	1	1	1	1	1	1	0	No benefit
Utility-Weighted mRS**	1	0.91	0.76	0.65	0.33	0	0	2.0 (SD±0.3)
Full 7-Level Ordinal mRS	A	B	C	D	E	F	G	2.1 (SD±0.4)

*Nested Co-Primary Analysis

**Lead Co-Primary Analysis, with NNTB determined using joint outcome table multiple resampling as described above

The Average Expected Benefit of Thrombectomy

A uniquely defined index of the magnitude of treatment effect in DAWN, the NNTSOL, can be derived directly from the mean observed disability change in the on the UW-mRS. The mean UW-mRS disability change associated with thrombectomy in DAWN was 2.04 (95% credible interval of 1.10 to 2.98) on a scale of zero to 10, where 10 corresponds to being neurologically asymptomatic and free of all disability and 0 corresponds to death. Thus, the expected benefit of 2.04 represents 20.4% of the total value of preventing a single death and, instead, achieving a normal neurologic outcome for that individual patient. Accordingly, the NNTSOL is 10/2.04 = 4.9 (95% credible interval 3.4-9.1). This value indicates that 4.9 patients need to be treated with thrombectomy rather than medical care alone for the accrued benefit to equal saving one life with return to normal function.

The numerical value of the NNTSOL for ordinal or continuous outcomes is fully specified by parallel trial group data. The NNTSOL thus is free from the limitation of requiring additional assumptions that affect all three of the NNTB derivation techniques for ordinal or continuous outcomes (automated derivation of an ordinal/continuous NNTB, expert-informed derivation of an ordinal/continuous NNTB, and resort to ignoring important outcomes in order to derive a dichotomous NNTB instead). Parallel trial data do completely specify the numerical value of the NNTSOL, and the NNTSOL captures all, not just one, of the many effects of treatment.

Accordingly, the NNTSOL is a better index of all patient-centered effects of an intervention than is the binary NNTB, and is an assumption-free index of patient-centered effects of an intervention, unlikely the ordinal/continuous NNTB. Interpretation of the NNTSOL should take into account that the effect magnitude value it provides is an average over all patients, including some who did not benefit at all from thrombectomy or were even harmed, and others who received a benefit much greater than the average benefit. The resulting value may be particularly helpful for those considering the effects of thrombectomy on a population basis, by indexing much more accurately than the binary NNTB the

magnitude of the treatment benefit, considering all valued effects of treatment rather than just one. The NNTSOL thereby complements the NNTB derived for an ordinal-continuous outcome (ocNNTB). The ocNNTB quantifies much more accurately than the bNNTB the proportion of patients who are responders – who experience clinically worthwhile clinical benefit, of greater or lesser degree, from a therapy. The NNTSOL quantifies much more accurately than the bNNTB the average benefit in the treatment group.

Together, the ocNNTB and the NNTSOL capture the two perspectives from which patients, physicians, regulators, and payors consider the benefit of therapy when making a treatment decision, before treatment exposure. One framework that decision-makers use is to consider the likelihood that the patient will benefit at all from treatment. This perspective is captured by the ocNNTB. A complementary framework that decision-makers use is to consider the average expected benefit across all treated patients. This perspective is captured by the NNTSOL. Both of these perspectives are important pretreatment, when patients are in the "the original position," and do not yet know if, were they treated, they would end up among the responders or non-responders. When decisions must be made from the original position, in situations of uncertainty regarding the distribution of benefits, decision-makers must, and do, consider both the "response probability" (the chance of being a responder) and the "expected utility" (the benefit averaged across all treated individuals).[19-21]

Supplemental Text Section S2: Rationale for the Clinical-Imaging Paradigm Utilized in the DAWN Trial

Current evidence suggests that benefit of thrombectomy rapidly decays over time and may no longer exist beyond 7.3 hours from stroke onset.[22] However, this relation between clinical benefit from thrombectomy and time to treatment initiation was established based on patient populations largely

unselected with respect to mismatch between extent of brain tissue irreversibly damaged (core) and extent of total brain tissue at risk. Indeed, only 141 (the first 71 SWIFT Prime patients and the all the 70 EXTEND-IA patients) out of the 1,757 patients included in the recent trials (MR CLEAN, ESCAPE, SWIFT Prime, EXTEND-IA, REVASCAT, THRACE, and PISTE) were required to have a target-mismatch penumbral profile prior to enrollment.[23-29] An exclusive time-based treatment paradigm disregards individual variations in compensatory mechanisms for ischemia including collateral flow. Therefore, it remains possible that a physiologic (e.g. Perfusion Imaging Mismatch or Clinical-Core Mismatch) rather than a purely time-based selection approach may demonstrate a benefit from reperfusion even at later time windows.[30,31]

The optimal patient selection methods for thrombectomy in large vessel occlusion stroke (LVOS) have yet to be defined. Specifically, there is no data to suggest that perfusion imaging mismatch is superior to clinical-core mismatch as a selection modality for endovascular reperfusion in LVOS. In fact, Nogueira et al. compared clinical diffusion mismatch (CDM, defined as NIHSS ≥8 and DWI volume ≤25 cc) with mean transit time-diffusion (MTT-DWI) mismatch as predictors of infarct growth in patients with proximal middle cerebral artery (MCA) occlusion and small infarct core on presentation and demonstrated that CDM was the most powerful predictor of infarct growth in this population.[32] The authors also showed that most of these patients will have a significant oligemic MTT lesion regardless of admission NIHSS score. Similarly, Bouslama et al. investigated a large cohort (n=384) of anterior circulation LVOS with baseline CT perfusion (CTP) and moderate or severe strokes (NIHSS ≥10) and assessed their thrombectomy eligibility by 4 mismatch criteria: Perfusion-Imaging Mismatch (PIM): between CTP-derived perfusion defect (Tmax >6 seconds) and ischemic core volumes (rCBF <30%); Clinical-Core Mismatch (CCM): between age-adjusted NIHSS and CTP core; Clinical-ASPECTS Mismatch(CAM-1): between age-adjusted NIHSS and ASPECTS; Clinical-ASPECTS Mismatch (CAM-2): between NIHSS and ASPECTS.[33] CAM-2 and CCM had higher inclusion (89.3 and 82.3%) vs. CAM-1 (67.7%) and PIM (63.3%).

Proportions of selected patients were statistically different except for PIM and CAM-1 (p = 0.19), with PIM having the highest disagreement. There were no differences in good outcome rates between PIM(+)/PIM(-) (52.2 vs. 48.5%; p = 0.51) and CAM-2(+)/CAM-2(-) (52.4 vs. 38.5%; p = 0.12). CCM(+) and CAM-1(+) had higher rates of good outcome compared to non-selected counterparts (53.4 vs. 38.7%, p = 0.03; 56.6 vs. 38.6%; p = 0.002). The abilities of PIM, CCM, CAM-1, and CAM-2 to predict outcomes were similar according to the c-statistic, Akaike and Bayesian information criterion. The authors concluded that for patients with NIHSS ≥10, PIM appears to disqualify more patients without improving outcomes and that CCM may potentially improve selection, combining a high inclusion rate with optimal outcome discrimination across (+) and (-) patients. In a different analysis based on the SWIFT Prime population, Nogueira et al. demonstrated that CCM had a non-significantly larger treatment effect gradient across mismatch+ vs. mismatch- patients as compared to PIM and CAM but the overall clinical outcomes and endovascular treatment effects in SWIFT PRIME did not differ based on clinical-imaging and perfusion-imaging mismatch selection criteria.[34]

Our inclusion criteria were based on the principle of age-adjusted clinical-core mismatch. The specific cut-offs were derived by the investigators based on their interpretation of the best available data on this subject at the time of the study design. Zaidi et al. demonstrated that at 40 cc infarct volume cutoff, the prediction sensitivity and specificity for favorable outcome were 63% and 72%, respectively (area under curve 0.75, 95% CI 0.38–0.72).[35] Similarly, a FIV of approximately 50 cc demonstrated the greatest accuracy for distinguishing good versus poor outcome in the analysis conducted by Yoo et al.[36] In the Ribo et al. analysis, the target cut-off infarct volume that better predicted a good outcome for patients <70 years was 49 cc (sensitivity 80%, specificity 92.6%) and for patients 70-79 years was 32.5 cc (sensitivity 80%, specificity 81%).[37]

Lima et al. demonstrated that a FIV ≤16 cc had the greatest accuracy for identifying good outcomes for patients ≥80 years (sensitivity 75.0%, specificity 82.6%).[38] Similarly, the target cut-off infarct volume that

better predicted a good outcome for patients ≥80 years was 15.2 cc (sensitivity 81.3%, specificity 86.7%) in the Ribo et al. analysis.[37] This large interaction between age, final infarct volumes, and 90-day functional outcomes was also demonstrated in the Pittsburgh Outcomes After Stroke Thrombectomy score (POST = age + 0.5× final infarct volume + 15 × H; where H=1 if PH1/PH2 hematoma was present and 0 if absent).[39] Elderly (age ≥80 years) patients with POST scores of 80-89, 90-119, and ≥120 had 90-day functional independence rates of 44%, 19%, and 10%, respectively. Since our primary endpoint analysis did not exclusively assess the rates of good outcomes (90-day mRS 0-2) but rather primarily evaluated the absolute difference in the UW-mRS scores at 90 days it was decided to adopt slightly larger cutoff cores volumes in order to also capitalize on the degrees of improvement beyond the mRS 0-2 and by doing such improve generalizability and optimize trial enrollment.

Although not statistically significant for clinical benefit, the IMS-III trial suggested that patients with NIHSS ≥20 could potentially have greater benefit from endovascular therapy as compared to those in the 8-19 category (RR, 99%CI: 1.37 [0.63-2.99] vs. 1.01 [0.78-1.31]).[40] We therefore adjusted the core thresholds for patients younger <80 years according to clinical severity e.g. NIHSS 10-19: <31cc and NIHSS ≥20: <51cc. Given the high sensitivity of patients age ≥80 years to final infarct volumes (as demonstrated in the aforementioned analysis), we decided to restrict their maximum threshold to cores < 21cc regardless of any additional clinical severity.

We opted to adopt the RAPID automated software in order to optimized the standardization of the imaging inclusion criteria across the many different sites and increase the precision of the baseline core measurements. Notably, the other existent option at the time of the study design included local ASPECTS reading which would have theoretically added greater variability in the imaging assessment across sites.[41]

The overall data on the RAPID Software demonstrates good agreement between baseline CTP and MRI core estimates as well as good ability to follow-up infarct volumes in patients achieving near full or full reperfusion. Cereda et al performed a direct comparison of infarct core size between RAPID CTP and RAPID MRI.[42] The authors found that a 38% rCBF threshold gave the smallest mean absolute difference across the two methods but the rCBF threshold that optimized the median absolute difference between CTP and DWI lesion volumes was 30%. There were a few significant overcalls at 38% which did not occur at 30%. Therefore, the 30% threshold was considered optimal to avoid overcalls, and the mean absolute difference was very similar to 38%. Furthermore, since the DWI was always done after the CTP in that study, it was likely that there was some growth between the CTP and the MR, further arguing that having some "under-call" was appropriate and reasonable. For example, the rCBF<30% threshold, on average, underestimates the DWI lesion by 12 ml; however, it had greater specificity for predicting DWI positive voxels compared to the rCBF<38% threshold (95% vs 87%). Because of the bias towards underestimation, the rCBF<30% threshold is less likely to overestimate the ischemic core than the <38% threshold and when overestimation occurs, the volume by which it overestimates is smaller. As such, the rCBF<30% threshold was utilized in the recent endovascular trials including DAWN. In SWIFT Prime trial, Albers et al. demonstrated that the rCBF<30% CTP threshold gave a similar accuracy for predicting 27-hour infarct in patients with >90% reperfusion core prediction as did DWI with standard ADC thresholds.[43] Notably, de Champfleur et al. have also showed that the MRI selected patients in the SWIFT Prime trial had a nearly identical treatment response as the CTP selected patients.[44]

Supplemental Text Section S3: Additional Methods for Imaging and Angiography Analyses

All central imaging and angiography analyses for the DAWN trial were conducted at the Neurovascular Imaging Research Core at UCLA, the imaging core laboratory for the entire study. Central ratings of

baseline and follow up imaging findings and quantitative measures for ischemic lesions were conducted by a neuroimaging expert blinded to all other data. While real-time, volume assessments on baseline CT and MRI lesion performed at enrolling sites were carried out using fully automated image processing software (RAPID, IschemiaView, Menlo Park, CA) with limited local reader quality check, the central core lab CT and MRI lesion volume assessments were performed in a semi-automated fashion using manual lesion outlining in combination with automated image processing software.

Baseline CT Imaging

CT-selected patients underwent an imaging protocol that included: (1) non-contrast CT (NCCT), (2) head and neck CT angiography (CTA), and (3) CT perfusion (CTP). The CTP protocols for all sites were tuned to harmonize acquisition parameters.[43]Criteria included brain coverage of at least 8cm, temporal sampling resolution no more than 1.8 seconds, tube voltage = 80kVp, Volume CT dose index (CDTIvol) <360mGy, scan duration = 70-90 seconds, reconstructed slice thickness = 5mm and no gap or overlap, high iodine concentration contrast agent (e.g. Omnipaque 350 or Isovue 370), injection flow rate = 4-6 ml/s, and amount of injected contrast 40–50ml, with no scan delay after bolus injection. Allowed scan modes included: burst mode (GE, Toshiba, Philips), jog mode (Philips) or dynamic helical shuttle (Siemens). For CTs with a detector width <8cm, either 2 CT perfusion runs or dynamic helical shuttle mode were mandated. Iterative reconstruction methods were avoided to reduce variability between vendors.[43]

To ensure reliable enrollment in the trial based on uniformly measured core lesion size at randomization, images were evaluated on site with a fully automated software environment (RAPID; iSchemaView, Menlo Park, CA).

The ischemic core volume as assessed on perfusion CT, (which represents estimated infarcted tissue was defined as voxels showing relative cerebral blood flow of <30% of the contralateral tissues. Only the

ischemic core volume was used when assessing whether patients met clinical core mismatch entry criteria. For exploratory additional analyses, the total critically hypoperfused volume was defined by >6-second delay in the time to maximum of the tissue residue function (Tmax); and the penumbral volume, representing imminently threatened but not yet infarcted tissue, was defined as the total critically hypoperfused tissue volume minus ischemic core tissue volume.

Baseline MR Imaging

MRI-selected patients underwent an imaging protocol that included: (1) diffusion-weighted image (DWI), and (2) head and neck contrast-enhanced MR angiography (CE MRA). The acquisition of an MR perfusion (MRP) image was optional. The MR protocols (1.5T and 3T) were also adjusted across sites.[43] DWI sequences required diffusion encoding in 3 principal directions from which an isotropically diffusion-weighted image was computed for subsequent analysis. The b values were 0 and 1,000s/mm2. Parallel imaging was used to decrease geometric distortion, except on GE scanners, where residual aliasing causes erroneous apparent diffusion coefficient (ADC) artifacts. Other criteria included whole brain coverage, slice thickness 5mm, and scan duration < 90 seconds. When performed, MR perfusion was carried out with gradient-echo echo planar imaging sequences. The sequence repetition time and thus the temporal sampling resolution was 1.8 seconds. The number of 5mm slices that could be fit within this sampling time varied between scanners and their hardware from 14 to 25 slices. Flip angle was chosen to be approximately 80 degrees to maximize signal. Echo time was 45 milliseconds at 1.5T and 30 milliseconds at 3T.[43]

Follow-Up Brain Imaging

Follow-up imaging evaluation including MRI/MRA or CT/CTA was performed at 24 (-6/+24) hours post randomization in order to assess for hemorrhage, vessel patency, and follow-up infarct volume. Sites

were encouraged to use the same imaging modality at 24 hours as the one used at baseline, to increase comparability of infarct volume and vessel patency assessments between baseline and follow-up. At the investigator's discretion, and if in keeping with local clinical practice, additional repeat imaging between day 5 to 7 (or earlier if discharge was planned earlier), could be performed, ideally including MRI DWI, T2 and FLAIR with MRA, to assess the day 5-7 infarct volume and vessel patency. CT could be performed at the day 5-7-time point if MRI was contraindicated.

All cervicocerebral imaging studies (including baseline and follow up CT or MRI) were sent to the independent core lab for analyses which were performed blinded to treatment allocation. Imaging parameters evaluated by the core lab for the purposes of this report include follow-up infarct volume at 24h and presence and radiologic degree of ICH evaluated using European Cooperative Acute Stroke Study (ECASS) III criteria][45] For the purposes of this report, Infarct growth was expressed as 24 hour infarct size (core lab adjudicated) minus baseline infarct size (assessed pre-randomization by RAPID at each individual site) between the treatment and control groups. In cases where 24-hour infarct size was smaller than baseline infarct size, infarct growth was considered to be 0.

Infarct volumes at 24 hours (-6/+24) and at day 5-7 or prior to discharge (if available). were determined based on the imaging modalities acquired. For MRI's, infarct volume at 24 hours was assessed on DWI and infarct volume at all subsequent time points on FLAIR if available, on T2 if FLAIR not available and on DWI if neither FLAIR or T2 available. For CT based imaging, infarct volume was assessed on NCCT if available, by manually outlining the hypodense lesion. If NCCT was not available, CTP ischemic lesion volumes were calculated. If both CT and MRI were performed within a follow-up assessment window, then the volume from the MRI lesion was selected. Regions of hemorrhagic transformation were included in the infarct volume.

In addition, whenever available, recanalization of the qualifying occlusive lesion on the 24-h follow-up

CTA or MRA scans was also analyzed. Vessel occlusion status on CTA or MRA at 24 hours was characterized according to a scale ranging from 0 (no recanalization) to 1 (partial recanalization) to 2 (complete recanalization).[46] Revascularization at 24 hours was defined as the presence of partial or complete recanalization. If the initially occlude M1 MCA segment was found to be completely or partially recanalized on follow-up CTA or MRA but both M2 MCA divisions remained occluded, recanalization was considered as unsuccessful.

Supplemental Text Section S4: Statistical design considerations

1.0 Background

This is a plan for a confirmatory trial of the use of the Trevo device for treatment of stroke patients. The hypothesis is that randomization to the Trevo device leads to improved outcomes as measured by 90 day scores on the modified Rankin scale (mRS), after controlling for the severity of the patient's injury, as measured by infarct size evaluated by MRI or CTP.

The trial is a Bayesian adaptive design that includes multiple key features:

1) adaptive sample size ranging from 150 to 500 patients

   the trial can stop for futility once 150 patients are enrolled

   the trial can stop for predicted success once 200 patients are enrolled
2) an enrichment design that may adjust inclusion criteria to specific infarct sizes
3) a utility function on 90-day mRS scores to reflect patient and society valuation of outcome health states.

The overall goal of the trial is to identify whether the Trevo device improves 90-day mRS scores in patients with 0 to 50cc infarcts. *A priori* however there is reason to believe the device may not be efficacious for the entire range of infarct sizes. Therefore, during the trial, the effect size will be

measured as a function of infarct size and the inclusion criteria may adapt to enroll only patients for whom efficacy is likely.

The adaptive sample size is incorporated so that the trial will stop for futility if the device offers no efficacy at any infarct size or will stop accrual early for predicted success if it is highly likely to demonstrate efficacy with the currently enrolled sample size.

Finally, it is widely accepted that a 1-point improvement on the mRS scale has different consequences for different parts of the scale. For instance, an improvement from 6 to 5 offers little patient benefit, however improving from a 4 to a 3 reflects a dramatic improvement in a patient's quality of life and need for daily assistance. Therefore, we use a utility function to reflect different utilities offered by improvements on the mRS scale.

The mRS score can also be dichotomized, where a score of 0 to 2 is counted as a success and a score of 3 to 6 is counted as a failure. This analysis method does a poorer job of measuring benefit to patients, but it has been popular historically. Dichotomized mRS also plays a role in the design by contributing to the sample size decisions. Specifically, the trial cannot stop early for expected success unless there is a high probability that the final data will demonstrate an mRS benefit to patients as measured both by the utility measure and by the dichotomous measure.

2.0 Trial overview

The trial will enroll from 150 to a maximum of 500 patients. It can stop early for either futility or expected success. This is also an enrichment design that attempts to determine the subpopulation of patients that benefit from the device. When the trial begins, subjects are enrolled with infarcts ranging

from 0 to 50 cc. After 150 patients are enrolled, the first interim analysis occurs, and the trial can potentially reduce the largest acceptable infarct to 30, 35, 40, or 45 cc. Interim analyses are repeated every 50 patients up to 450 patients.

If an enrichment decision is made, the trial stops enrolling subjects with infarct sizes that are larger than the new upper limit, and at the end of the trial, the final analysis includes only the subjects in the enriched population. All enrolled subjects and groups will be followed and summarized in reports of the results of the trial, including those subjects belonging to subpopulations for which enrollment was stopped.

In this document, we first describe the primary outcome and primary statistical test to be used at the final analysis. Then we describe the Bayesian adaptive trial design, including decision rules for patient enrichment and early stopping, while offering example trials to illustrate how it may proceed. Next, we illustrate the trial's operating characteristics including Type I error, power, sample size distribution, and probability of enrichment being implemented for a range of possible scenarios. Finally, two appendices fully describe the statistical models used in the trial. Appendix A describes the model used in the final analysis. Appendix B describes the model used during interim analyses.

3.0 Study population, Primary Endpoint & Statistical Test

3.1 Entry Criteria

Patients with infarct volumes between zero and <51 cc will be accepted at the start of the trial if they meet one of the following three conditions:

- 0- <21 cc core infarct and NIHSS ≥ 10 (and ≥ 80 years old)
- 0- <31 cc core infarct and NIHSS ≥ 10 (and < 80 years old)

- 31 cc to $\leq$51 cc core infarct and NIHSS ≥ 20 (and < 80 years old)

3.2 Treatment arms

This is a randomized control trial. The treatment arm, patients randomized to the TREVO device, will be compared with a control arm, including patients randomized to receive the standard of care. Patients will be randomized 1:1 to each of the two treatment arms. Patients randomized to the TREVO device will be given an angiography to determine suitability for receiving the device. Patients determined to be unsuitable for the device will receive the standard of care, but they are still considered to be part of the device arm population, and endpoint analyses will be based on intent-to-treat populations.

3.3 Primary Endpoint

The primary endpoint for this trial is the 90-day mRS score. We choose to analyze this standard endpoint by converting the mRS scores into weights that directly reflect patient and society valuation of outcome health states. We then model a subject's weighted mRS score as normally distributed with expected value depending on infarct size and treatment assigned.

The weights assigned to the possible mRS scores are shown in Table S5 below. These weights were obtained through a synthesis of studies. [5,6] Both these studies assigned utility values and confidence intervals to mRS scores; these are also shown in Table1. We renormalized these utilities to a scale where an mRS of 6 implies a utility of 0 and mRS of 0 implies a utility of 10. The two scales are quite similar, and we take the mean of the renormalized utilities to obtain our own weights. The second study reported more precise estimates, so in some cases the consensus value is closer to its value.

Table 1 weights used for 90 day mRS scores.

mRS	0	1	2	3	4	5	6
Rivero-Arias et al	10	8.7	7.3	6.0	2.8	-0.1	0
Hong & Saver	10	9.5	7.9	6.7	3.5	0.1	0
This Trial	10	9.1	7.6	6.5	3.3	0	0

Relative to an approach that dichotomizes the 7 possible mRS scores into two possibilities, weighting the 7 Rankin levels by utilities improves the precision of the scale as a measure of disability. The weighted approach should also not be confused with an approach based on the raw mRS scores, which would erroneously treat each single-point increase in mRS as equally valuable to the subject.

We consider the device to offer a positive treatment effect if the device offers an increase in the expected utility among subjects who are randomized to the device.

Another possible method of analyzing the 90 day mRS endpoint involves using a dichotomized version of the utility function (weight of 10 for mRS of two or better, weight of zero for mRS of three or worse). We use this analysis method in early stopping decisions as described in section 4.3 and as an additional analysis method of the primary endpoint for the marketing application.

3.4 *Primary Analysis*

The final analysis is Bayesian and includes a flexible normal dynamic linear model (NDLM) to account for different expected outcomes as a function of infarct size. This is a flexible spline-like model that will capture that the average weighted mRS score in the control group as a (possibly non-linear) function of infarct size. Meanwhile the average treatment effect, θ, is assumed to be equal over the identified range of infarct sizes.

The overall treatment effect θ is given a vague prior, $\theta \sim N(0,2.5^2)$, and if there is a high posterior probability that the overall treatment effect θ is positive, the device is declared to be efficacious.

3.5 Threshold for a Successful Trial

The threshold for declaring success depends on the degree to which the population has been enriched with the thresholds adjusted to control Type I error probability to be no more than 0.025. If no adaptation occurs, the threshold is 0.986:

If $Pr(\theta > 0) \geq 0.986$ then conclude the device is efficacious.

This is similar to a one-sided frequentist test at the α=0.014 level. This critical value is smaller than 0.025 in order to account for multiple possible stopping times.

The threshold for a successful trial depends on whether the population was enriched and the numbers of enrolled subjects in the enriched population at the time of the enrichment decision, the number of enrolled subjects outside the enriched population, and the number of subjects enrolled after the enrichment decision. The formula is given in Appendix A; the threshold for the posterior probability of a

positive treatment effect assuming no enrichment is 0.986. The threshold is constructed to increase as enrichment becomes earlier and more aggressive.

The adaptive design does not utilize the specific values of the NIHSS scores used to define entry criteria, except to define discontinuities in the expected weight as a function of infarct size. Specifically, there is a discontinuity between 30 and 31 cc, and we expect subjects with infarct sizes of 30 and 31 cc to be less similar than other subjects separated by a single cc. For the purposes of this trial, infarct volumes are treated as integers between 0 and 50 inclusive; measured infarct volumes should be rounded to the nearest integer if necessary.

4.0 Prospectively Planned Interim Analyses

Interim analyses will take place after 150, 200, 250, 300, 350, 400, and 450 patients have been enrolled. Five possible decisions may be made at interim analyses:

1. The trial may stop for futility starting at the 150-patient analysis.
2. The trial may stop accrual for predicted success starting at the 200-patient analysis.
3. The trial may enrich and continue to accrue patients: exclude certain larger infarct sizes in the future and focus on a smaller subset of infarct sizes.
4. The trial will stop enrolling if it has enrolled 500 patients.
5. If none of Conditions 1-4 are met the trial proceeds as-is with another analysis occurring after 50 more patients are enrolled.

All interim decisions to alter the trial's study population or sample size will be made based upon predictive probability calculations described in greater detail in Appendix B. An overview is described here.

4.1 Interim Monitoring for Early Futility

The decision to stop the trial early for futility or to enrich is based on predictive probabilities that the trial would be a success if subjects were enrolled to the maximum sample size.

One predictive probability calculation is done for each possible maximum infarct size. Specifically, define π_{30} to be the Bayesian predictive probability that the trial would be successful if no further subjects were enrolled with infarct sizes larger than 30 cc but the trial continued to enroll subjects up to the maximum sample size of 500. Define π_{35}, π_{40}, π_{45}, and π_{50} analogously, and write e^* to be the maximum infarct size for the current population (e^*= 50 at the beginning of the trial). The trial stops early for futility if $\pi_e \leq 0.10$ for all $e \leq e^*$ (if no legal enrichment gives a probability of more than 10% of winning the trial).

This means the trial stops for futility even if enrolling only the best possible subset of patients, based on infarct size, offers less than or equal to a 10% chance of success – even if enrolling to the maximum sample size of 500 patients.

4.2 Enrichment

The trial may choose to focus on a subpopulation of infarct sizes, i.e. 'enrich', starting at the 150-patient interim analysis. The 400-patient interim analysis is the last opportunity to enrich.

The decision will be made to *enrich* the population if it appears that the device benefits one subset of the population considerably more than another. The candidate enriched populations that the trial considers are based on relatively smaller infarct sizes. The five possible subpopulations are

1. The full population of infarct sizes 0 to <51 cc;
2. Infarct sizes of 0 to 45 cc;
3. Infarct sizes of 0 to 40 cc;
4. Infarct sizes of 0 to 35 cc;
5. Infarct sizes of 0 to 30 cc.

If the population is enriched, subjects with larger infarct sizes are no longer enrolled, and the final efficacy analysis omits subjects with larger infarct sizes from consideration. (The upper limit of 500 total enrolled subjects applies whether or not the population is enriched, and the 500 include subjects enrolled early in the trial with infarct sizes larger than ultimately included in the enriched population.)

Enrichment decisions are irreversible, once an infarct size is dropped it can never be re-included. However, the trial can enrich the population further after it has already been enriched. Also, enrichment can only occur from larger to smaller, for example smaller infarct sizes (e.g. 0 to 20) cannot be excluded from future enrollment producing an enriched set of infarcts from 20 to 35.

If an enrichment occurs, the trial is required to enroll at least 100 more subjects after that point, unless the trial elects to stop for futility earlier.

The design will enrich if one of two conditions is met. First, we calculate the effect size in the highest 5 included infarct sizes. If the current maximum allowable infarct size is $e = 35$ or more, calculate the

posterior probability the average treatment effect is positive for infarct sizes between $e-4$ and e. If the probability is less than 40%, enrich by only including patients with infarct sizes $e-5$ or less. This rule is applied only one time per interim analysis.

Next, we seek to identify whether it's possible to increase the predictive probability of trial success by only enrolling a subset of infarct sizes. If it is possible to increase the predictive probability of winning the trial by at least 10% by enriching (i.e. if $\pi_e > \pi_{e^*} - 0.10$ for some $e < e^*$), the design enriches to e. If multiple enrichments meet this criterion, the least severe (largest e) enrichment is chosen, but this rule is applied iteratively. For example, if e = 50 at the current interim analysis (no enrichment has occurred yet and patients are being enrolled up to the maximum infarct size of 50cc) and $\pi_{50} = 0.4$, $\pi_{45} = 0.55$, and $\pi_{40} = 0.70$, the rule enriches from 50 to 45 and then from 45 to 40, although the effect is the same as a single-step enrichment from 50 to 40.

Step 1 was added because if patients with infarct sizes $e-4$ to e were rare and these rare patients saw no benefit with the Trevo device, π_e might not be 10 points less than π_{e-5}, nevertheless we would prefer to exclude these patients.

If an enrichment decision is made before the 400-patient analysis, these tests are performed again at the next interim analysis 50-patients later and further enrichment may occur. If enrichment occurs, the trial will never stop accrual for predicted success until at least 100 more patients are enrolled from just the enriched patient population.

One result of these enrichment rules is that we do not incentivize excluding rare infarct sizes from the study population. The predictive probability calculation factors in the relative probabilities of each infarct size, so that the predictive probability of winning the trial will not change if we exclude a group of subjects that we rarely expect to enroll by the end of the trial.

4.3 Interim Monitoring for Early Stopping for Predicted Success

Expected success stopping decisions are based on a different predictive probability calculation, also discussed in Appendix B. If at least 100 subjects have been enrolled since the last enrichment, we calculate the probability that, if all currently enrolled patients were followed to their 90-day outcomes, but no further subjects were enrolled, the trial would end in success.

At such an analysis, some enrolled patients will have reached their 90-day outcomes, some will have 30-day mRS scores, and others (the most recently enrolled) will have no outcome data. Therefore, the predictive probability of trial success with the current sample size is based on the predictive distribution of the final endpoints for the subjects enrolled but who do not yet have final 90-day mRS endpoints recorded. At the 200 and 250-patient interim analyses, if the probability of trial success based on just the enrolled patients is at least 95%, the trial will stop for expected success. The threshold for expected success stopping decreases as the trial continues: after 300 or 350 subjects, it decreases to 90%, then to 85% after 400 subjects and 80% after 450 subjects.

In order that the trial stop for expected success, the predictive probability of trial success must exceed the threshold described above, and the analogous predictive probability of trial success for a dichotomized version of the utility function must also exceed the same threshold. The dichotomized utility function is also discussed in Appendix B.

4.4 Longitudinal Modeling during Prediction Step

Patients will also be evaluated for modified Rankin score at 30 days. These scores are used to assist in making decisions during the course of the trial, and these include decisions to stop early and to cease

enrollment for a subset of the initially eligible population, but do not play a direct role in the final analysis.

Using patients who have completed the trial and provide both 30 and 90-day mRS scores, we estimate the probability distribution of 90 day mRS conditionally on 30 day mRS. Then we use this estimated distribution to construct a longitudinal model for imputing final mRS outcomes for subjects with known 30-day mRS but unknown 90-day mRS. The parameters of the longitudinal model will be updated during every interim analysis, and the longitudinal model is initialized to be non-informative (i.e. it does not use data from other studies to inform the distribution of 90-day mRS given 30-day mRS). The same longitudinal model parameters are used for both of the treatment arms. Appendix B includes full details of the longitudinal model.

5.0 Example Trial

In this section, we present graphical descriptions of the data collected, analyses performed, and decisions made during the course of the trial. Our example is just one of the many simulated data sets and trials.

The example trial presented in this section was created using a slightly different, older version of the design, in which the dichotomized version of the utility function is not used in the expected success stopping decisions.

In this scenario, we assume an accrual rate of 16.7 patients per month. Therefore, the first interim analysis occurs approximately 9 months after the first patient was enrolled and randomized. At this time, approximately 100 patients have reached their 90-day mRS outcome.

Figure 1 illustrates the first interim analysis, performed after 150 subjects have been enrolled. The raw data are shown in the leftmost plot. The x-coordinates are subject infarct sizes, while the y-coordinates display their mRS scores and weight scores, with good scores (mRS of 0 receiving a weight of 10) at the top of the plot. Red plot symbols denote control subjects, while blue plot symbols represent device subjects. Smaller plot symbols are used for subjects who have only reported 30 day endpoints, while subjects before 30 days are not shown. The same data are also shown in the second plot, and we have added estimates (solid lines) and confidence intervals (dotted lines) for the expected weights as a function of infarct size.

At this point, the device is estimated to be beneficial for subjects with infarct sizes smaller than roughly 30cc and more likely to be harmful than beneficial for larger infarcts, but there is still considerable uncertainty. The assumed discontinuity at 30cc is clearly visible. The rightmost plot shows the information used to make decisions. The green line and green triangles represent information used in deciding whether to stop for expected success: the triangles depict the predictive probability that the trial would be a success if we stopped enrollment altogether at this point, and the triangle for the current population would need to be above the green line at 95% in order to stop for expected success. The first enrichment mechanism is depicted using black diamonds and the black horizontal line. The black diamonds display the posterior probability that the average treatment effect is positive for four different subsets of infarct sizes (from left, 31 to 35, 36 to 40, 41 to 45, and 46 to <51). While the current population is 0 to 50, if the black diamond for the 46 to 50 range is below the black line at 40%, the trial enriches down to a subpopulation of 0 to 45. This requirement is satisfied, so we decide to enroll no more subjects with infarcts of 46 to <51 cc. The second enrichment mechanism is depicted using the bars. The bars represent the predictive probabilities of winning the trial if we enrich the population to each of the allowable subsets and then continue enrollment to the maximum of 500 subjects. The shortest bar corresponds to the full (0- <51) population, which has a predictive probability

of about 66% of winning. If all five bars were below the red line at 10%, the trial would stop for futility; this does not happen here. If one of the subpopulations of 0-30 through 0-40 has a 10% higher probability of a successful trial than the current population of 0-45, we enrich to one of those populations at this point. Since the predictive probabilities indicate that we can increase the probability of winning the trial by more than 10% by enriching, we will make an enrichment decision in this interim analysis by dropping to 0-40. The 0-35 subpopulation has a slightly higher predictive probability than 0-40, but the rule dictates that we make the least severe enrichment that gives us the 10% improvement. In summary, the trial will no longer enroll subjects with infarcts of 41cc or larger. Also, the trial must now enroll at least 250 patients before stopping for predicted success – it must enroll at least 100 patients after this enrichment decision is made.

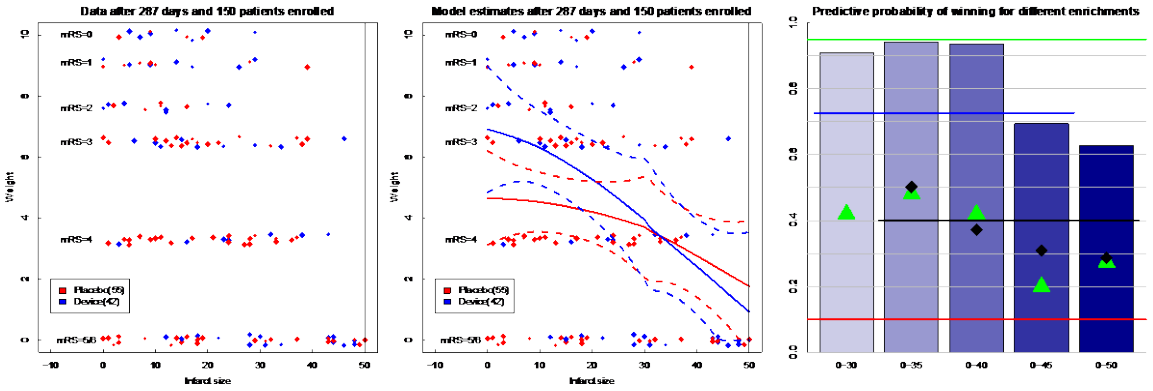


Figure 1: First interim analysis of example trial.

The second interim analysis is conducted after a total of 200 subjects have been enrolled. In all three plots, we indicate that subjects with large infarcts above 41 cc are no longer being enrolled: the current population is depicted by a vertical line in the first two plots and blacked out predictive probability bars in the rightmost plot. The solid curves in the middle picture show that the data acquired since the last interim analysis have been consistent with better outcomes for both placebo and device subjects with small infarcts, and the device seems to be lagging more for larger infarcts. The rightmost plot shows

that the 0-40 population has a large enough probability of winning the trial that we do not stop for futility or enrich with the idea of increasing the probability of winning. The black diamonds show that we believe that the probability of a positive device effect for 36-40 subjects is below the black horizontal line, which triggers a further enrichment to a population of 0-35. For two reasons, we cannot stop for expected success here: first, the third green triangle shows that the probability of a successful trial given that we stop now is well below the threshold given by the green line. Second, we have not enrolled 100 subjects since our last enrichment decision.

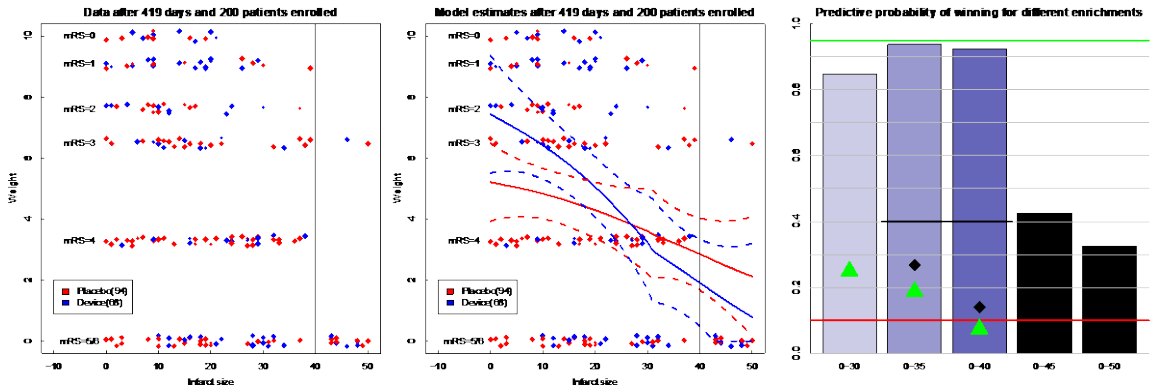


Figure 2: Interim analysis of example trial after 200 subjects enrolled.

The third interim analysis occurs when 250 patients are enrolled and is shown in Figure 3. No trial adaptations occur at this interim analysis. The probability of trial success exceeds 10% so the trial does not stop for futility. The trial does not stop accrual for predicted success: although the predictive probability of winning the trial given stopping now (green triangle) is >90% the trial has not yet enrolled 100 patients since the last enrichment decision then accrual continues.

Also, there is no enrichment adaptation: the treatment effect for 31-35 subjects is likely enough to be positive that we don't enrich for that reason, and the predictive probability of success given full enrollment is high enough so that it can't be improved with a further enrichment to 0-30 either.

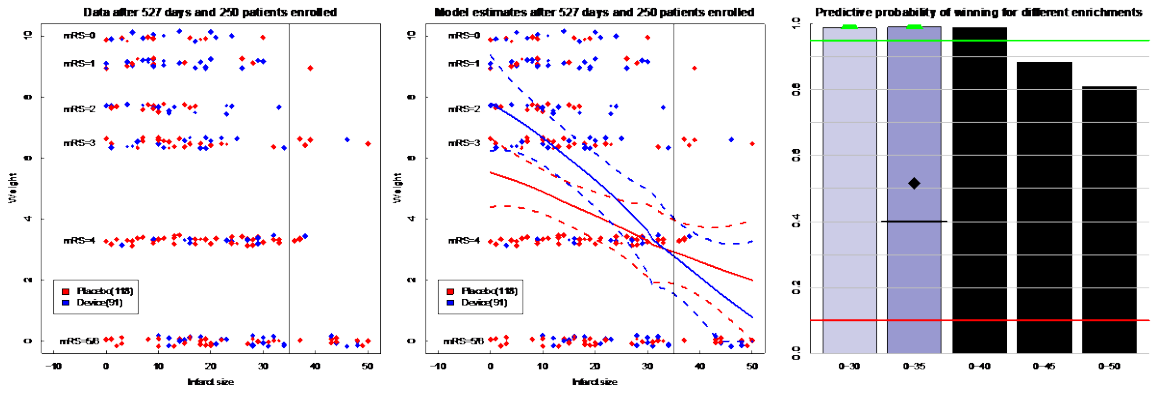


Figure 3: Third interim analysis for example trial.

The fourth interim analysis occurs after 300 patients are enrolled. This interim analysis, shown in Figure 4, results in the decision to stop for expected success, since the predictive probability is above 95% for the current population of 0-35 and since we have now enrolled 100 subjects since the final enrichment. We now follow the remaining 27 subjects to get their complete data and then perform the final analysis.

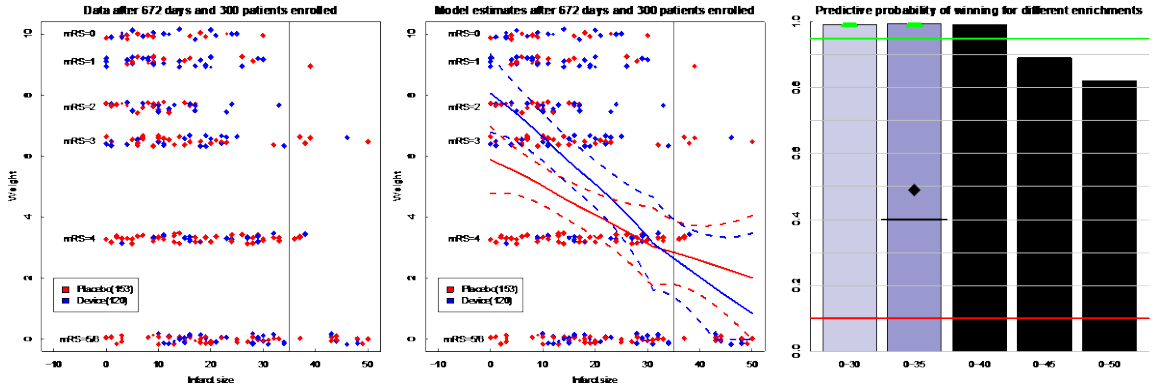


Figure 4: Fourth interim analysis for example trial.

Figure 5 displays the simulation truth for this trial and the results of the final analysis. The distributions of mRS values for each infarct size, both for control subjects (left plot) and for device subjects (center plot) are shown; the darkest blue represents mRS of 0, while the darkest red represents mortality. The device actually provides a benefit of 1 unit of expected weight for subjects with infarct sizes of 40 or less and no benefit to those with larger infarcts. The expected weights for device (blue) and control (red) are

plotted on the rightmost plot. As printed on the rightmost plot, the final posterior probability that the overall treatment effect for infarct sizes 0-35 is positive is 0.9996, which is larger than the critical value of 0.9895 that applies to this trial, in which the last enrichment happens when 200 subjects are enrolled and it excludes the 28 subjects with infarct sizes 36 or more from the primary analysis.

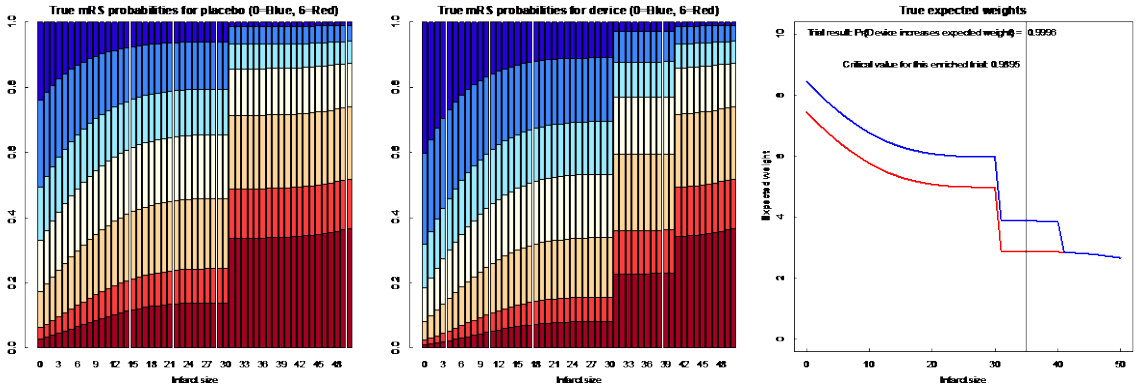


Figure 5: simulation truth and final analysis for example trial.

6.0 Operating Characteristics

The operating characteristics of this trial, including Type I error rate, power, sample size distribution, and likelihood of enrichment occurring, cannot be evaluated analytically. Therefore, we calculate the design's operating characteristics through simulation. By choosing a 'true' scenario for the distribution of mRS outcome given infarct size in the control and treatment groups, we can then simulate the trial. We sample patient's infarct size, their randomization assignment, and their outcome, and then perform the interim analyses as described, any adaptation defined by the design's rules, and continue to accrue simulated patients. We repeat this process thousands of times per scenario. Then for each 'true' scenario we summarize the proportion of trials that enrich, the sample size distribution, and the proportion that lead to successful efficacy claims. The proportion of trials that meet the standard to declare efficacy reflects the Type I error rate in scenarios where the device truly offers no efficacy and

reflects the power in scenarios in which the device is efficacious across at least some subset of infarct sizes.

A scenario consists of the following elements:

1. An accrual rate;
2. A distribution $\{\Pr\{I = i\} = \iota_i : 0 \leq i \leq 50\}$ of infarct sizes;
3. For each infarct size i, a distribution of 90-day mRS scores for subjects given standard treatment $\Pr\{Y_j = k | I_j = i, d_j = 0\}$, for $k = 0,1, \ldots ,6$.
4. For each infarct size i, a distribution of 90-day mRS scores for subjects randomized to the device $\Pr\{Y_j = k | I_j = i, d_j = 1\}$;
5. Conditional distributions of 30-day mRS scores given 90-day mRS scores, which in principle could also depend on infarct size and treatment.

The number of such scenarios is enormous. A realistic restriction is to assume that larger infarct sizes are unambiguously worse than smaller infarct sizes with respect to the distribution of outcomes. In practice we use probit models to simulate data, in which we define cut points $-\infty = K_7 < K_6 < \cdots < K_0 = \infty$, and a subject's outcome is determined by simulating a Gaussian random variable Z_j with mean $\mu(i, d)$and variance $v(d)$, after which the mRS score is equal to k if and only if $Z_j \in (K_{k+1}, K_k]$. In this case, if $\mu(i, d)$ is non-increasing in i for both d, then large infarct sizes are unambiguously worse.

We explore three different accrual rates. The default rate is 200 patients per year. We also consider a slower rate of 100 patients per year, and focus on this rate in the Type I error simulations. The slow rate would require 5 years of recruitment to achieve full enrollment. Finally, we consider a faster rate of 300 patients per year.

In addition, we assume an infarct size distribution. Our default assumption is as follows. We assume 30% of subjects have infarct sizes uniformly distributed between 0 and 10, 25% of subjects have infarct sizes uniformly distributed between 11 and 20, 20% of subjects have infarct sizes uniformly distributed between 21 and 30, and 25% of subjects have infarct sizes uniformly distributed between 31 and 50. We also consider a "Uniform" profile in which each infarct size is equally likely, and a "Lower" profile in which smaller infarcts are more common than in the default profile. These distributions are summarized in Table 2.

Infarct range	0-10	11-20	21-30	31-50
Default	30%	25%	20%	25%
Uniform	11/51	10/51	10/51	20/51
Lower	50%	30%	10%	10%

Table 2: Three distributions of infarct sizes used in the simulations. For these distributions, the probability of each infarct size is equal across infarct sizes in the same interval defined by the columns of the table.

The relationship between infarct size and outcome must also be established. While these could be extremely complex, we focus on five primary cases.

- Standard. This scenario is meant to be representative of a realistic relationship between infarct size and outcome, decreasing but with some curvature and a large discontinuity at 30 cc. This set of distributions is shown in left plot on Figure 6.
- Flat. This scenario assumes that outcome distributions are independent of infarct size between 0 and 30, and independent between 31 and 50. This set of distributions is shown in the middle plot on Figure 6.

- Jagged. The flexible models for expected weight as a function of infarct size punish deviations from linearity, so a scenario that maximizes curvature might be expected to be extreme with respect to Type I error. Such a scenario is piecewise constant with large changes at 15 and 40 cc. (Since the model assumes a discontinuity at 30 cc, roughness at that point should challenge the model less than above or below it). This set of distributions is shown in the right plot on Figure 6.
- Standard Plus. This scenario is similar to the standard scenario, but with higher probabilities of good outcomes for all infarct sizes.
- Standard Minus. This scenario is similar to the standard scenario, but with lower probabilities of good outcomes for all infarct sizes.

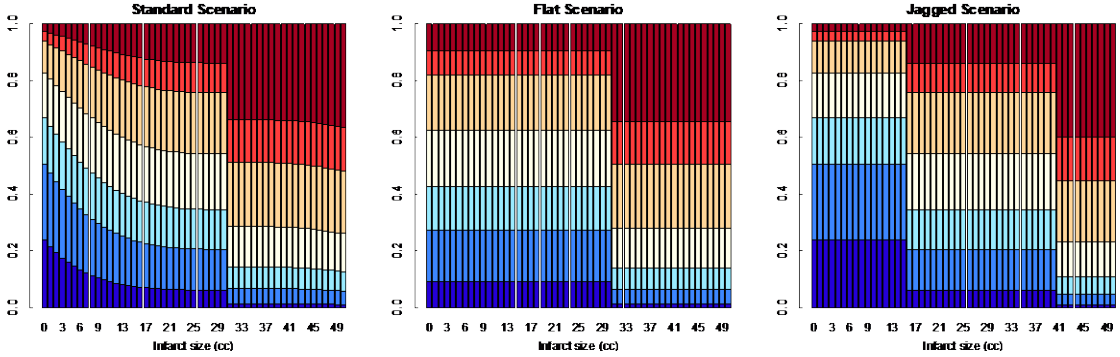


Figure 6: distributions of outcome as a function of infarct size, for the standard, flat, and jagged scenarios. Blue represents mRS of 0, red represents mRS of 6.

Finally, we consider three different relationships between 30-day outcomes and 90-day outcomes. The default relationship is specified in Table 3, where it is written in terms of the distribution of 30-day mRS conditional on 90-day mRS. This table is interpreted as follows: for example, if a subject is dead (mRS = 6) at 90 days, there was a 60% chance he was dead at 30 days, a 25% chance his 30 day mRS was 5, etc. Note that the conditional distributions in Table 3 are used only in the simulation experiments to

generate random patients; they will not be used in the conduct of the trial to inform the longitudinal model.

90 \ 30	0	1	2	3	4	5	6
0	0.50	0.30	0.10	0.05	0.05	0.00	0.00
1	0.10	0.50	0.20	0.15	0.05	0.00	0.00
2	0.05	0.15	0.40	0.30	0.09	0.01	0.00
3	0.02	0.08	0.25	0.45	0.18	0.02	0.00
4	0.02	0.03	0.15	0.20	0.55	0.05	0.00
5	0.01	0.02	0.03	0.09	0.25	0.60	0.00
6	0	0	0.01	0.04	0.10	0.25	0.60

Table 3: conditional distribution of 30 day mRS given 90 day mRS used as the default scenario in the power simulations.

We consider two other more extreme relationships as well. Under the "Perfect" longitudinal model, 30 day outcomes are identical to 90 day outcomes with probability 1. (The design does not utilize this information, however, except to the extent that it learns it as the trial progresses.) Under the "Uniform" longitudinal model, 30-day outcomes are almost useless for predicting 90-day outcomes: for any patient with a 90-day mRS between 0 and 5, the six possible 30-day mRS values of 0 to 5 are equally likely. Patients with 90-day mRS of 6 have probability 1/7 of each 30-day mRS between 0 and 6.

6.1 Type I error

First, we conservatively calculate Type I error by simulating the trial without allowing for the possibility of futility stopping. Allowing for futility stopping can only decrease the Type I error rate – it would allow for stopping some trials before they otherwise might result in an efficacy claim.

First, we define a "null scenario."

A null scenario is one in which $\Pr\{Y_j = k | I_j = i, d_j = 1\} = \Pr\{Y_j = k | I_j = i, d_j = 0\}$ for all i. In other words, there is no infarct size, *i*, in which the device offers a benefit vs. the standard of care on 90-day mRS scores.

In our simulations, we have emphasized scenarios that are likely to maximize the Type I error probability. In particular, Type I error probability is maximized if the scenario is such that one makes success stopping decisions based on complete information to the largest extent possible. Type I error is highest in these situations when all data is known, because stopping for predicted success will almost always produce a trial meeting the success criteria – a Type I error. When the trial is stopped for predicted success based upon more incomplete information, e.g. patients enrolled but lacking 90-day mRS scores, the outstanding data, once known, could result in a trial not quite meeting the success criteria – hence this regression to the mean due to the additional outstanding data will result in a dataset that does not lead to a Type I error.

In this trial, complete information is maximized two possible ways: (1) a very slow accrual rate leading to few patients lacking 90-day mRS scores at any particular interim analysis and (2) a strong correlation between patients' 30-day and 90-day mRS scores.

Therefore, to simulate scenarios that lead to higher Type I error probabilities, we simulate scenarios with very slow accrual rates such that it takes 5 years to enroll 500 subjects of infarct sizes 0 to 50.

Enrolling 500 patients in 5 years (60 months) results in approximately 8.3 patients per month, or just 25 enrolled patients for whom 90-day mRS scores are unknown at each interim analysis. In our default accrual rate assumption, 500 patients are accrued in 2.5 years, a rate of 16.7 patients per month. This rate would result in 50 patients who are enrolled but lack 90-day mRS data at each interim analysis.

Therefore, our conservative Type I error calculation has 2 times less outstanding data at each interim analysis.

Similarly, if 30 day mRS scores are maximally informative, i.e. if the day 30 endpoint is equal to the 90-day endpoint with probability one, Type I error is maximized.

Table 4 below defines the scenarios for which we have evaluated Type I error probability. The infarct size, mRS, longitudinal model, and accrual profiles were defined in the previous section.

Simulation	Infarct Size	*f*(mRs)	Effect Size	Longitudinal Model	Accrual rates
1	Default	Standard	0	Default	Default
2	Default	Standard	0	Default	Fast
3	Default	Standard	0	Default	Slow
4	Default	Standard	0	Uniform	Slow
5	Default	Standard	0	Perfect	Slow
6	Default	Standard+	0	Default	Slow
7	Default	Standard–	0	Default	Slow
8	Default	Flat	0	Default	Slow
9	Default	Jagged	0	Default	Slow
10	Uniform	Standard	0	Default	Slow
11	Lower	Standard	0	Default	Slow

Table 4: Definition of null scenarios in Type I error simulations.

Table 5 contains the simulation estimates of the Type I error probability for these scenarios, together with probability of futility (zero for all but the last scenario), average number of patients enrolled, and the probability that each possible final subpopulation is chosen in a successful trial. These estimates are based on 20000 simulations apiece. Type I error probabilities are no larger than 0.025 for all null

scenarios, even with no futility stopping and with slow accrual and a perfectly informative longitudinal model. Since no futility stopping is allowed, the trial enrolls the maximum number of subjects with high probability, so the expected sample sizes are nearly 500. Successful trials are rare, but when they do occur, the smallest or largest possible subpopulations are selected more often than the intermediate populations.

Simulation	P(Win)	Pr(Fut)	Mean SS	Final Infarct Size and Winning Trial				
				0-30	0-35	0-40	0-45	0-50
1	0.0136	0	499.4	0.004	0.002	0.002	0.001	0.005
2	0.0173	0	499.5	0.005	0.002	0.001	0.003	0.006
3	0.0194	0	498.4	0.005	0.004	0.001	0.002	0.007
4	0.0159	0	499.3	0.005	0.002	0.003	0.002	0.005
5	0.0171	0	498.4	0.005	0.004	0.001	0.002	0.005
6	0.0157	0	499.0	0.004	0.002	0.001	0.002	0.006
7	0.0151	0	499.0	0.005	0.002	0.002	0.002	0.004
8	0.0121	0	499.0	0.003	0.001	0.002	0.001	0.005
9	0.0162	0	498.4	0.005	0.003	0.002	0.002	0.004
10	0.0173	0	499.2	0.005	0.004	0.002	0.002	0.005
11	0.0160	0	499.3	0.004	0.002	0.003	0.003	0.004
12 W/ futility	0.0071	0.977	239.9	0.002	0.001	0.001	0.001	0.002

Table 5: estimated Type I error probability for several scenarios.

In fact, the Type I error probabilities for this design are conservative and well below the target value of 0.025. This is because the critical value (98.6%) for the design was selected using a previous version of the design, where the polychotomous utility function alone determined early stopping. Designs where early stopping decisions are based on criteria close to the primary analysis tend to inflate Type I error

above the nominal value more than designs that use other criteria. Including the dichotomous utility function in early stopping decisions therefore reduces the Type I error probability.

The primary analysis in this trial uses the utility function defined in Table 1. However, it may still be of interest to evaluate the probability that, at the end of the trial, a version of the primary analysis using the dichotomized utility function would yield a significant result (98.6% or higher posterior probability that the device is beneficial). Estimated probabilities for the same null scenarios are shown in Table 5D below. These probabilities are, in all cases, smaller than the analogous probabilities for the polychotomous utility function.

Simulation	Probability of superiority
1	0.0122
2	0.0154
3	0.0187
4	0.0129
5	0.0150
6	0.0129
7	0.0136
8	0.0099
9	0.0135
10	0.0147
11	0.0139
12 (W/ futility)	0.0043

Table 5D: estimated probability that the trial will demonstrate superiority with respect to the dichotomized utility function, using the form of the primary analysis with the utilities changed.

6.2 Power

In this section, we report simulation-based power estimates over a range of scenarios in which the Trevo device is efficacious for at least some subset of infarct sizes. Unlike in the Type I error simulations, the power simulations include the design's rules for futility stopping.

These scenarios are no longer those chosen to maximize Type I error probability; instead they are more representative of realistic scenarios. First, we consider a variety of scenarios in which the device's effect is constant for all infarct sizes. Specifically, we consider device effects of 0.5, 0.75, 1, 1.25, and 1.5 unit increases in expected utility, in addition to a case where the device has no benefit. The assumptions on accrual rate, infarct size distribution, mRS distribution, and longitudinal models are those that were defined in Section 6. For several combinations of these assumptions, we evaluate operating characteristics for each effect size. The scenarios are defined in Table 6 below.

Simulation Sets	**Infarct Size**	***f*(mRs)**	**Effect Size**	**Longitudinal Model**	**Accrual rates**
A	Default	Standard	0,0.5,0.75, 1,1.25,1.5	Default	Default
B	Default	Standard	0,0.5,0.75, 1,1.25,1.5	Default	Fast
C	Default	Standard	0,0.5,0.75, 1,1.25,1.5	Default	Slow
D	Default	Standard+	0,0.5,0.75, 1,1.25,1.5	Default	Default
E	Default	Standard–	0,0.5,0.75, 1,1.25,1.5	Default	Default
F	Default	Standard	0,0.5,0.75, 1,1.25,1.5	Perfect	Default

G	Default	Standard	0,0.5,0.75, 1,1.25,1.5	Random	Default
H	Uniform	Standard	0,0.5,0.75, 1,1.25,1.5	Default	Default
I	Lower	Standard	0,0.5,0.75, 1,1.25,1.5	Default	Default

Table 6: simulation sets for power calculations.

We investigate treatment effects that increase the expected weight by 0, 0.5, 0.75, 1, 1.25, and 1.5 units for all infarct sizes. Figure 7 below illustrates the conditional distributions of mRS given infarct size in the standard profile, for placebo and for five different treatment effect sizes, including a zero-treatment effect, which is identical to the placebo profile.

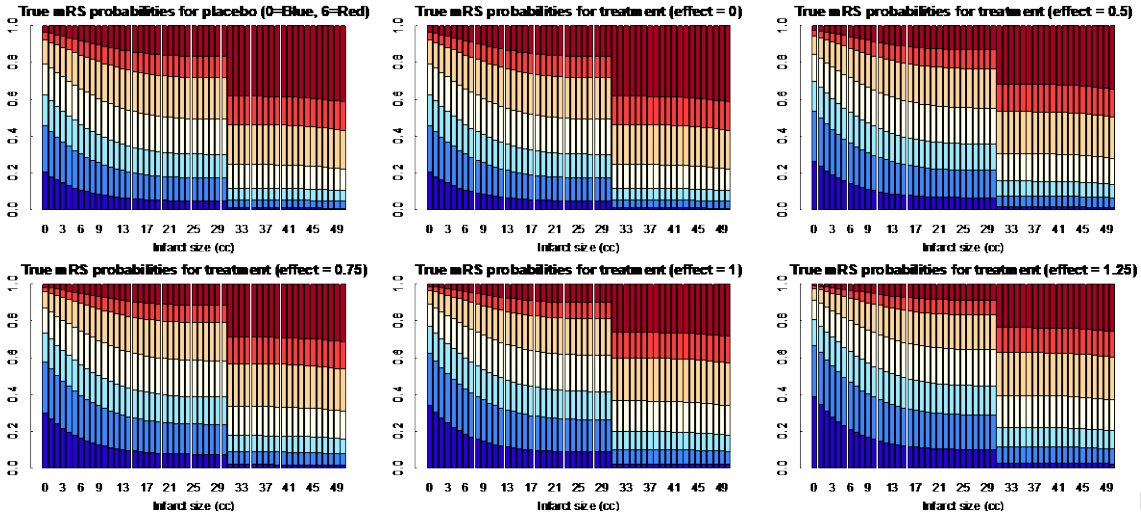


Figure 7: mRS distributions for all infarct sizes, for each of several treatment effects, for standard scenarios.

Table 7 shows the marginal distributions of mRS scores for each of these treatment effects, assuming the infarct size distribution described in Section 4. Due to rounding error, the numbers in a row may not sum to exactly one.

	mRS=0	mRS=1	mRS=2	mRS=3	mRS=4	mRS=5	mRS=6
Effect=0	0.07	0.13	0.12	0.17	0.20	0.11	0.19
Effect=0.5	0.09	0.16	0.13	0.18	0.19	0.10	0.15

Effect=0.75	0.10	0.17	0.14	0.18	0.19	0.09	0.13
Effect=1	0.12	0.18	0.14	0.18	0.18	0.08	0.12
Effect=1.25	0.13	0.19	0.15	0.18	0.17	0.08	0.10

Table 7: Marginal distributions of 90 day mRS assumed in the power simulations, based on averaging with respect to the infarct size distribution.

Operating characteristics for the trial are shown in Tables 8 through 16. Due to faster accrual rates and/or futility stopping, Type I error probabilities are in general lower than in the cases evaluated in Section 6.1. The effect size of 0.5 units of weight is small, and consequently a trial of this size is unlikely to detect it, the trial offers roughly 30% power in scenarios with this effect size. The effect sizes of 1.25 and 1.5, on the other hand, are very large and the trial offers better than 95% power to detect such improvements.

Expected sample sizes are smaller in the null case (when early stopping for futility is likely) and in the case of very large effect size (when early stopping for expected success is likely). The trial enrolls more subjects precisely when the data are inconclusive about whether the device has a substantial positive effect. The table also shows the fraction of trials that are successful and ultimately select a subpopulation smaller than the full population. As one would hope, aggressive enrichment is most common in the case of small to moderate treatment effects, when the design is trying to find a population for which the device works.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.017	0.971	243.2	0.01	0.00	0.00	0.00	0.00
0.50	0.289	0.664	359.2	0.05	0.03	0.04	0.04	0.13
0.75	0.617	0.337	401.3	0.11	0.07	0.05	0.07	0.32

1.0	0.871	0.116	390.1	0.11	0.07	0.07	0.11	0.51
1.25	0.967	0.029	333.8	0.09	0.06	0.06	0.11	0.65
1.50	0.997	0.003	292.4	0.06	0.04	0.04	0.10	0.76

Table 8: Operating characteristics for Simulation Set A, which features default profiles for accrual, infarct size distribution, mRS distributions, and longitudinal model.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.016	0.978	253.6	0.00	0.00	0.00	0.00	0.01
0.50	0.269	0.685	366.0	0.05	0.03	0.03	0.04	0.11
0.75	0.592	0.361	401.6	0.11	0.06	0.04	0.09	0.29
1.0	0.869	0.116	398.4	0.12	0.08	0.06	0.13	0.56
1.25	0.978	0.020	362.1	0.10	0.08	0.05	0.12	0.63
1.50	0.994	0.006	306.6	0.07	0.04	0.05	0.10	0.74

Table 9: Operating characteristics for Simulation Set B, which features a faster accrual rate. Expected sample sizes are larger than in the default scenario.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.018	0.974	243.6	0.00	0.00	0.00	0.00	0.01
0.50	0.292	0.661	359.0	0.05	0.04	0.04	0.02	0.14
0.75	0.616	0.352	382.6	0.11	0.06	0.05	0.06	0.33
1.0	0.880	0.102	368.0	0.10	0.08	0.07	0.11	0.53
1.25	0.978	0.021	321.6	0.07	0.06	0.06	0.11	0.67
1.50	0.994	0.006	264.5	0.05	0.04	0.03	0.09	0.79

Table 10: Operating characteristics for Simulation Set C, which features a relatively slow accrual rate. Successful trials are slightly more likely than in the default case, since early stopping for expected success decisions are made based on relatively more complete data. Sample sizes, however, tend to be smaller.

Effect	P(Win)	Pr(Fut)	Mean SS	Final Infarct Size and Winning Trial				
				0-30	0-35	0-40	0-45	0-50
0	0.032	0.962	255.6	0.01	0.00	0.00	0.01	0.01
0.50	0.372	0.580	382.8	0.09	0.05	0.03	0.03	0.17
0.75	0.729	0.236	401.8	0.16	0.10	0.07	0.08	0.32
1.0	0.942	0.047	365.4	0.13	0.10	0.07	0.10	0.55
1.25	0.994	0.003	301.4	0.09	0.08	0.06	0.10	0.68
1.50	0.998	0.001	246.8	0.04	0.04	0.04	0.09	0.80

Table 11: Operating characteristics for Simulation Set D, characterized by better mRS outcomes for given infarct sizes than the default profile. Successful trials are more common than in the default case, and large effects can be detected with fewer patients on the average. These findings likely depend on the specific choices for the infarct size and mRS distributions. The 0.032 probability for the null scenario is a random high in simulation results; there is almost a 10% chance of 32 or more successes out of 1000 trials with success probability 0.025.

Effect	P(Win)	Pr(Fut)	Mean SS	Final Infarct Size and Winning Trial				
				0-30	0-35	0-40	0-45	0-50
0	0.019	0.976	239.1	0.00	0.01	0.00	0.01	0.00
0.50	0.240	0.714	354.4	0.04	0.03	0.02	0.03	0.11

0.75	0.574	0.388	395.3	0.10	0.05	0.05	0.08	0.29
1.0	0.825	0.143	400.0	0.10	0.06	0.06	0.12	0.49
1.25	0.945	0.050	363.3	0.07	0.04	0.06	0.15	0.64
1.50	0.989	0.009	317.6	0.03	0.03	0.04	0.12	0.76

Table 12: Operating characteristics for Simulation Set E, which is characterized by worse mRS outcomes for given infarct sizes than in the default profile. In general, this scenario is slightly unfavorable to the default scenario in operating characteristics.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.015	0.979	245.2	0.00	0.00	0.00	0.00	0.01
0.50	0.315	0.630	362.6	0.06	0.03	0.04	0.04	0.14
0.75	0.640	0.325	388.4	0.09	0.07	0.05	0.08	0.35
1.0	0.876	0.103	365.2	0.12	0.07	0.08	0.11	0.50
1.25	0.970	0.027	309.5	0.08	0.05	0.05	0.10	0.68
1.50	0.994	0.004	258.5	0.06	0.03	0.05	0.08	0.78

Table 13: Operating characteristics for Simulation Set F, in which 30-day mRS scores are always identical to 90-day mRS scores. Since more information is available earlier, decisions to stop early for expected success are more successful, so that win probabilities are slightly larger and expected sample sizes are smaller.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.012	0.985	241.8	0.00	0.00	0.00	0.00	0.01
0.50	0.255	0.698	362.4	0.06	0.03	0.03	0.03	0.10

0.75	0.592	0.362	405.3	0.09	0.08	0.04	0.10	0.28
1.0	0.848	0.131	400.0	0.12	0.07	0.06	0.11	0.49
1.25	0.959	0.035	372.4	0.13	0.08	0.07	0.12	0.56
1.50	0.992	0.008	321.0	0.08	0.06	0.05	0.10	0.71

Table 14: Operating characteristics for Simulation Set G, in which 30-day mRS scores are nearly unrelated to 90-day mRS scores. Win probabilities are generally decreased and expected sample sizes increased.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.014	0.981	256.6	0.00	0.00	0.00	0.00	0.01
0.50	0.260	0.691	368.6	0.05	0.05	0.02	0.04	0.11
0.75	0.563	0.365	405.6	0.09	0.08	0.06	0.05	0.30
1.0	0.860	0.108	397.8	0.12	0.10	0.07	0.09	0.49
1.25	0.967	0.026	347.2	0.08	0.08	0.05	0.08	0.67
1.50	0.990	0.008	295.6	0.05	0.05	0.04	0.06	0.79

Table 15: Operating characteristics for Simulation Set H, in which each infarct size between 0 and 50 is equally likely. This change in assumptions does not have a dramatic effect on power, but enrichment is less likely and the full subpopulation is more likely to be chosen by the design.

Effect	**P(Win)**	**Pr(Fut)**	**Mean SS**	**Final Infarct Size and Winning Trial**				
				0-30	**0-35**	**0-40**	**0-45**	**0-50**
0	0.014	0.979	239.6	0.01	0.00	0.00	0.00	0.00
0.50	0.291	0.668	351.8	0.05	0.02	0.04	0.04	0.14

0.75	0.624	0.354	384.7	0.09	0.05	0.09	0.10	0.29
1.0	0.874	0.112	374.6	0.09	0.06	0.10	0.13	0.49
1.25	0.981	0.021	335.3	0.09	0.06	0.10	0.16	0.57
1.50	0.996	0.004	282.4	0.05	0.03	0.08	0.16	0.68

Table 16: Operating characteristics for Simulation Set I, in which infarcts tend to be smaller than in the default profile. This change in assumptions makes it relatively easier to detect a device effect, and smaller subpopulations are more likely to be chosen by the design.

6.3 Population Enrichment

In this section, we investigate the enrichment decisions made by the design. In particular, we investigate how likely the design is to select the right subpopulation when the device benefits only a fraction of the population. For the purpose of this study, we define scenarios where the device has a constant effect on the expected weight scale for infarct sizes less than a specified value, and it has no effect for larger infarct sizes. We also investigate scenarios in which the treatment effect decreases more slowly with effect size. Table 17 summarizes these scenarios. For reference, scenario E1 offers a one unit expected weight benefit for all patients. In scenarios E2 through E5, the device offers a benefit of 1 unit expected weight to subjects with specified infarct sizes or smaller, and no benefit to larger infarct sizes. Scenarios E6 and E7 benefit large infarct sizes substantially, and the benefit decreases steadily for larger infarcts.

Simulation Sets	**0-30**	**31-35**	**36-40**	**41-45**	**46-50**
E1	1	1	1	1	1
E2	1	1	1	1	0

E3	1	1	1	0	0
E4	1	1	0	0	0
E5	1	0	0	0	0
E6	1.5	1.25	1	0.75	0.5
E7	1.25	1	0.75	0.5	0

Table 17: effect sizes as functions of infarct size in the enrichment scenarios.

Table 18 shows the results. We show the power, the probability that the design selects the various possible subpopulations, and the probability that the trial is successful with each possible subpopulation chosen. Power increases with the size of the population that benefits: with a large population that benefits, the final analysis is less likely to involve patients that do not benefit, and the sample size for the final selected population is likely to be larger. Also, the design is likelier to select a small population if the truth is that only a small population benefits. These numbers are based on 1000 simulated trials per scenario.

Simulation Effect	**P(Win)**	**Final Infarct Size**					**Final Infarct Size and Winning Trial**				
		0-30	**0-35**	**0-40**	**0-45**	**0-50**	**0-30**	**0-35**	**0-40**	**0-45**	**0-50**
E1	0.871	0.14	0.08	0.07	0.13	0.58	0.11	0.07	0.07	0.11	0.51
E2	0.844	0.16	0.11	0.12	0.20	0.41	0.12	0.09	0.11	0.19	0.35
E3	0.827	0.20	0.15	0.19	0.16	0.30	0.16	0.13	0.17	0.15	0.22
E4	0.807	0.25	0.27	0.13	0.11	0.24	0.20	0.24	0.12	0.09	0.16
E5	0.793	0.38	0.20	0.11	0.09	0.22	0.32	0.17	0.09	0.07	0.14
E6	0.992	0.11	0.09	0.12	0.19	0.49	0.11	0.09	0.12	0.19	0.49
E7	0.955	0.16	0.14	0.17	0.18	0.33	0.15	0.15	0.17	0.18	0.31

Table 18: power and enrichment decisions when the truth is that the benefit of the device depends on infarct size. For each scenario, we present the total probability of a successful trial in the P(Win)

column. We then present the probability that the design selects each of the five possible subpopulations regardless of whether the trial is successful, and finally the probability that a trial is successful and chooses each of the five possible subpopulations.

6.4 Probability of successful final analysis with dichotomized utility

We can also evaluate the probability of a significant result when the final data are analyzed using the dichotomized utility function. For brevity, we focus on the scenarios in Simulation Set A of Section 6.2, so the numbers in Table 19 can be compared to those in Table 8.

Effect	P(Win)*	Pr(Fut)	Mean SS	Final Infarct Size and Winning Trial				
				0-30	0-35	0-40	0-45	0-50
0	0.009	0.971	243.2	0.01	0.00	0.00	0.00	0.00
0.50	0.174	0.664	359.2	0.05	0.03	0.04	0.04	0.13
0.75	0.447	0.337	401.3	0.11	0.07	0.05	0.07	0.32
1.0	0.681	0.116	390.1	0.11	0.07	0.07	0.11	0.51
1.25	0.902	0.029	333.8	0.09	0.06	0.06	0.11	0.65
1.50	0.980	0.003	292.4	0.06	0.04	0.04	0.10	0.76

Table 19: Operating characteristics for Simulation Set A, which features default profiles for accrual, infarct size distribution, mRS distributions, and longitudinal model. The numbers in the P(Win)* column refer to analyses performed on the final data using the dichotomized utility function. The other numbers are reproduced from Table 8.

Probabilities of significant results for the dichotomous analysis are considerably smaller than for the primary analysis (which uses the utility function in Table 1). The utility analysis is a more sensitive method of detecting a benefit. Dichotomizing the endpoint has the effect of maximizing the variance of

the endpoint as a function of its mean. In the case of Simulation Set A, the estimated standard deviations of outcomes are about 50% larger for the dichotomized analysis than they are for the utility analysis, so loss of power is inevitable.

7.0 Summary

The randomized controlled Bayesian adaptive design described here offers high power to detect an improvement offered by the Trevo device vs. standard of care in patients with an acute ischemic stroke in the combined late and wake up stroke population.

Furthermore, the design permits the trial to stop early for futility or stop accrual early for predicted success. In any success scenario, all enrolled patients are followed to their 90-day outcomes before any final analysis is performed.

Furthermore, the innovative design adapts to the responding subpopulation of infarct sizes.

As described, however, the trial design maintains control of Type I error even in conservative circumstances. Trial sample sizes tend to be lower when the device offers no efficacy or strong efficacy. When the benefit is more ambiguous the trial enrolls more patients in an attempt to identify which patients (patients with which size infarcts) are most likely to benefit by the Trevo therapy. In such a situation, at least 100 additional patients are enrolled in order to validate this finding.

Appendix A: Statistical Model For Final Analyses

Denote the j'th subject's 90-day mRS by S_j, and her resulting weight score by Y_j; $Y_j = W_k$ if $S_j = k$. Further denote the jth subject's infarct volume by I_j and the treatment to which she was randomized by d_j (d_j is either 0 for placebo or 1 for the TREVO device). Define, for $0 \leq i \leq 50$ and $d \in \{0,1\}$,

$$\Pr\{Y_j = k \mid I_j = i, d_j = d\} = p_k(i,d).$$

For the purposes of the final analysis, we model the $p_k(i,d)$ as Gaussian with expected values that depend on i, with a common treatment effect θ, and with variances σ_d^2 that depend on the treatment:

$$E\left(Y_j \middle| I_j = i, d_j = d\right) = \sum_{k=0}^{6} p_k(i,d) W_k = \phi_i + \theta\, I\{d = 1\},$$

for all infarct sizes i that are in the enriched population at the end of the trial. (The final analysis ignores the subjects outside the enriched population.) We model the ϕ_i flexibly, and assume that they come from a second order normal dynamic linear model (NDLM). Specifically, the prior distribution for the ϕ_i assumes that for $2 \leq i \leq 50$, we have

$$\phi_i \sim Normal(2\phi_{i-1} - \phi_{i-2}, \tau^2 \delta^{2I\{i=31\ or\ i=32\}}).$$

This form of the normal dynamic linear model encourages the ϕ_i to be linear, and allows a discontinuity between infarct sizes of 30 and 31, where the entry criteria change.

We use the following prior distributions:

$$\theta \sim Normal(0, 2.5^2),$$

$$\sigma_d^2 \sim Inverse\ Gamma\ (1, 10)\ (d = 0,1),$$

$$\tau^2 \sim Inverse\ Gamma\ (10, 0.005), \text{ and}$$

$$\delta^2 \sim Inverse\ Gamma\ (1, 1000).$$

We evaluate the posterior distribution of the parameters of this model using the Gibbs sampler. Conditionally on the other parameters, (ϕ, θ) have a multivariate normal distribution, and the remaining parameters have inverse Gamma conditional distributions.

The primary output of the final analysis is the posterior probability that $\theta > 0$. If this probability is large enough, the trial is considered to be a success. The threshold for determining significance is computed as follows. Let N_1 be the number of subjects in the enriched population that have been enrolled at the time of the last enrichment, let N_2 be the number of subjects enrolled but outside the enriched population at the time of the last enrichment, and let N_3 be the number of subjects still to be enrolled at the time of the last enrichment (all of them are by definition in the enriched population; this number may be larger than the number of subjects actually enrolled due to stopping early for expected success). Then the threshold is

$$\Phi\left(\sqrt{1 + \frac{N_2}{N_1 + N_3}}\,\Phi^{-1}(p_{crit})\right),$$

where Φ is the standard normal cumulative distribution function, and p_{crit} is a critical probability evaluated via simulation to control Type I error probability; we use $p_{crit} = 0.986$. As an example of how enrichment inflates the threshold, suppose that after 400 subjects have been enrolled, with 50 of them having infarct sizes larger than 40, we enrich the population to infarct sizes of 0 to 40, and then enroll the last 100 subjects. In this case $N_1 = 350, N_2 = 50$, and $N_3 = 100$, so that the adjusted threshold is 0.9906. On the other hand, suppose that after 150 subjects, 50 of which have infarct sizes larger than 30, we enrich to infarct sizes of 0 to 30 and then enroll a further 150 subjects, at which time we stop for expected success. In this case, $N_1 = 100, N_2 = 50$, and $N_3 = 150$, the threshold increases to 0.9920.

The accuracy of Monte Carlo estimates of $\Pr\{\theta > 0|Y\}$ can be improved substantially by using *Rao-Blackwellization:* instead of simply counting the number of positive θs in N Monte Carlo simulations, one instead takes the Monte Carlo average of

$$\Pr\{\theta > 0|Y, \sigma_a^2, \tau^2, \delta^2)$$

obtained during the Gibbs sampling steps in which one draws new values of (ϕ, θ).

Appendix B: Statistical Model for Futility and Expected Success Analyses

The statistical model used during the trial to make decisions about enrichment and early stopping is more detailed than the final analysis model. The largest change is that instead of a single treatment effect θ that applies for all infarct sizes in the enriched population, we estimate a different treatment effect for each infarct size, and the shape of the treatment effect function of infarct size is also modeled flexibly using a second-order normal dynamic linear model. We also use a longitudinal model to impute values of final endpoints for subjects for whom we have 30-day mRS scores but not 90-day scores; we estimate the probability distribution of final endpoint values given early endpoint values. The third major change is that we also estimate the distribution of infarct sizes for enrolled subjects; for predicting whether the trial will be successful it is critical to be able to forecast what kinds of subjects will appear in the future.

Specifically, we denote the treatment effect (i.e., increase in expected weight) for subjects with infarct sizes i by θ_i, and assume a second order normal dynamic linear model: for $2 \leq i \leq 50$, assume

$\theta_i \sim Normal(2\theta_{i-1} - \theta_{i-2}, \tau_\theta^2 \delta^{2I\{i=31 \text{ or } i=32\}})$.

The same discontinuity multiplier $\delta^2$is used in the prior distributions for both ϕ and θ, but the smoothness parameter τ_θ^2 is different from the analogous parameter used for the ϕ_i (the expected weight for placebo subjects). Additionally, whereas in the final analysis we use a non-informative prior with no information about the overall level of the θ_i or the overall slopes of the θ_i as a function of i, we now use the following prior distributions:

$$\theta_0 \sim N(0, 2.5^2), \theta_{50} \sim N(0, 2.5^2), \theta_1 \sim N(\theta_0, 0.25^2), and\ \theta_{49} \sim N(\theta_{50}, 0.25^2),$$

along with

$\phi_0 \sim N(5, 2.5^2), \phi_{50} \sim N(5, 2.5^2), \phi_1 \sim N(\phi_0, 0.25^2), and\ \phi_{49} \sim N(\phi_{50}, 0.25^2)$. The prior distribution for τ_θ^2 is Inverse Gamma (10, 0.005).

Writing Y_j^{30} for the 30-day mRS value for the jth subject, we estimate the probabilities $\lambda_{mk} = \Pr\{Y_j = k \mid Y_j^{30} = m\}$ using a multinomial model with prior distributions

$$(\lambda_{m0}, \lambda_{m1}, \lambda_{m2}, \lambda_{m3}, \lambda_{m4}, \lambda_{m5}, \lambda_{m6}) \sim Dirichlet(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$$

for $m = 0,1,2,3,4,5$. We have $\lambda_{66} = 1$ and $\lambda_{6k} = 0$ for $k < 6$. As mentioned earlier, the longitudinal model plays no role in the final analysis. The parameters of the longitudinal model are updated at each interim analysis and are based on all subjects with complete 30-day and 90-day data to that point. We note that we are not using data from other studies to inform the parameters of the longitudinal model. We use the same longitudinal model for both arms (i.e. we pool the data for all patients to estimate the probability distribution of 90-day outcome given 30-day outcome). The final piece of the statistical model for interim analyses is the model for the infarct size distribution $\Pr\{I_j = i\} = \iota_i$, which is also a Dirichlet-multinomial model with prior distribution

$$(\iota_0, \iota_1, \dots, \iota_{50}) \sim Dirichlet\ (\frac{1}{3}, \frac{1}{3}, \dots, \frac{1}{3}).$$

The prior distributions for σ^2 and δ^2 are as specified in the description of the final analysis.

During an interim analysis, we estimate the parameters $(\phi, \theta, \sigma^2, \tau_\phi^2, \tau_\theta^2, \delta^2, \lambda, \iota)$ of this model using Gibbs sampling. We then use these samples to estimate several predictive quantities. First, for each candidate enrichment that is allowed at the current interim analysis, we calculate the probability that the trial would end with a significant result if we enriched to that population and enrolled subjects up to the maximum sample size. (Enrichments that are not allowed include re-expanding the population to

reverse a previous enrichment, or any enrichment with fewer than 100 subjects before the maximum sample size.) This calculation consists of the following steps: for a given MCMC sample,

1. Using the λs, impute 90-day endpoint values for the subjects enrolled and with 30-day data.
2. Simulate random infarct sizes and treatment assignments for the subjects yet to be enrolled, using the ιs and assuming that subject accrual is restricted to subjects in the enriched population. Augment this list of subjects with the subjects included in the trial who have not yet provided 30-day data.
3. Calculate the probability, given that list of subjects, that final 90-day data will result in a significant trial. We use a closed form expression for this probability that assumes fixed values of the parameters $\left(\sigma^2, \tau_\phi^2, \tau_\theta^2, \delta^2\right)$ and the normal model for the Ys given the parameters.

One may choose to repeat step 2 multiple times for a given MCMC sample. Compute the average of the resulting probabilities. These probabilities will be used to make enrichment decisions and futility stopping decisions.

Second, we calculate the probability that if the trial stopped enrollment now, the subjects who have been enrolled but who have not provided final endpoint data will ultimately yield a significant result. This calculation has a similar form to the one previously discussed, but it doesn't require simulating subjects yet to be enrolled in the trial. This probability is used to make decisions to stop for expected success.

Also used in the decisions to stop for expected success is a slight modification of this predictive probability. We replace the primary utility function with a dichotomized version (as in Table 1A below) and then calculate the expected success predictive probability using the dichotomous utility function. (We first re-analyze the available data using the model in Appendix A and the dichotomized endpoint values, and then use this model to predict future data with dichotomized endpoint values.) For the trial

to stop for expected success, the predictive probability must exceed the threshold for both the primary utility function and the dichotomized utility function.

mRS	0	1	2	3	4	5	6
Rivero-Arias et al	10	8.7	7.3	6.0	2.8	-0.1	0
Hong & Saver	10	9.5	7.9	6.7	3.5	0.1	0
This Trial	10	9.1	7.6	6.5	3.3	0	0
Dichotomized	10	10	10	0	0	0	0

Table 1A: dichotomized utility function used in expected success predictive probability calculations.

Supplemental Text Section S5: Sensitivity Analysis to Account for the 4 Patients with who were Lost to Follow Up 90 days (all of whom had 30 days follow-up data)

A total of four patients were lost to follow up after day 30 as following.

Subject ID	Treatment Arm	Infarct Size	30-day mRS
13100194	Control	45.2	5
13100223	Treatment	10.9	4
19900098	Control	25	5
71300135	Control	6	3

The imputed data set most unfavorable to Trevo thrombectomy had all three control arm subjects with missing data improving to an mRS of 0, and the device arm subject scoring an mRS of 5 or 6. This data set has an imputation probability of 5×10^{-7} and results in a probability of benefit for the device of 0.99987. Given this data set results in a large probability of benefit, any imputation model will conclude likewise.

For this dichotomized version of the analysis, there were only 16 different imputed data sets. Two imputed data sets are much more likely than the others: the data set where all four subjects have mRS scores of 3 or worse has probability 0.42 and results in a probability of benefit of $1 - 2\times10^{-8}$. The data set where the fourth subject (with a 30-day score of 3) is the only subject with a score of 2 or better has probability 0.37 and a probability of benefit of $1 - 5\times10^{-8}$.

The probabilities of benefit for the 16 possible imputed data sets range from $1 - 8\times10^{-9}$ to $1 - 4\times10^{-7}$, and the overall probability is $1 - 4\times10^{-8}$.

Supplemental Text Section S6: Principles of Medical and Procedural Management

A. Medical Management

In order to avoid performance bias, patients in both arms were admitted to intensive care units or acute stroke units providing semi intensive care and continuous cardiovascular monitoring. All subjects enrolled into this study were medically managed according to the 2013 AHA guidelines[47], according to the 2008 European Stroke Organization ESO Guidelines[48], according to the 2007 Australian Clinical Guidelines for Acute Stroke Management [49]and according to the 2015 Canadian Acute Stroke Best Practice recommendations [50] depending on geographical location and hospital specific policies. Common elements representing most important aspects of the medical management included:

Antiplatelet therapy: Patients not already on an antiplatelet agent were given at least 81 mg of acetylsalicylic acid (ASA) immediately as a one-time dose after brain imaging has excluded intracranial hemorrhage. In patients treated with tissue plasminogen activator (tPA), ASA was delayed until after the 24-hour post-thrombolysis scan has excluded intracranial haemorrhage as per standard post iv t-PA care guidelines. ASA (81 to 325 mg daily), Clopidogrel (75 mg daily) or a combination thereof

was then continued indefinitely or until an alternative antithrombotic regime was started. In dysphagic patients, ASA could be given by enteral tube or by rectal suppository and Clopidogrel by enteral tube.

Systemic Thrombolysis: In eligible patients, intravenous tPA was administered as a dose of 0.9 mg/kg to a maximum of 90 mg total dose, with 10 percent (0.09 mg/kg) given as intravenous bolus over one minute and the remaining 90 percent (0.81mg/kg) given as an intravenous infusion over 60 minutes) within 4.5 hours of the onset of stroke symptoms in accordance with criteria adapted from National Institute of Neurological Disorders and Stroke (NINDS) tPA Stroke Study and the European Cooperative Acute Stroke Study (ECASSIII). [48,51]

Blood pressure management: In subjects who received IV tPA, blood pressure should be managed according to post IV tPA management guidelines within the first 24 hours. Very high BP

(>185/110mmHg) was treated to lower the BP below these values. In ischemic stroke patients not eligible for thrombolytic therapy treatment of hypertension was not recommended to be routinely undertaken unless extreme blood pressure elevation (e.g. systolic > 220 or diastolic > 120mmHg),

which could be treated to gently reduce the blood pressure below these values over the first 24h with gradual reduction thereafter. In control group patients or in thrombectomy group patients who did not reperfuse, augmentation of blood pressure with pharmacological agents according to site specific practice was permitted.

In subjects who reperfused after mechanical thrombectomy (defined as achieving TICI 2b or TICI 3) it was recommended to maintain systolic blood pressure below 140 mm Hg in the first 24 hours to minimize the risk of reperfusion related intracerebral hemorrhage. In subjects who did not achieve reperfusion after thrombectomy similar BP management orders were applied as for the control subjects within each center.

Stroke unit care: Upon admission or after discharge from the intensive care unit, patients were cared for in an inpatient stroke unit, which is a specialized, geographically defined hospital unit dedicated to the management of stroke patients and supported by a core multidisciplinary team. The team should consist of healthcare professionals with stroke expertise, that included vascular neurologists, nursing, occupational therapy, physiotherapy, speech-language pathology, social work and clinical nutrition (dietician).

Complication Prevention: Appropriate investigations and management strategies were implemented for all patients to optimize recovery, diagnose and avoid complications, and prevent stroke recurrence. This included deep vein thrombosis (DVT) prophylaxis with early mobilization, +/- low-molecular weight heparin (LMWH) or unfractionated heparin, sequential compression devices (SCD), aspiration prevention with dysphagia screening and enteral feeding support when required.

Hemicraniectomy: decompressive surgery was considered in younger patients

in the early stages of malignant middle cerebral artery territory ischemic stroke.

Stroke specific rehabilitation consisted of care formally coordinated and organized by a multidisciplinary rehabilitation team consisting at a minimum of a physician (physiatrist, neurologist, or other physician with expertise/core training in stroke rehabilitation), nurse, physical

therapist, occupational therapist, speech-language pathologist, social worker and dietician.

B. **Interventional management**

In subjects randomized to the Trevo thrombectomy plus medical management arm (thrombectomy arm) the procedure was performed using the FDA-approved Trevo Retriever (ProVue and XP ProVue). The use of other endovascular reperfusion methods (other stentrievers, primary aspiration, intracranial stenting/angioplasty, thrombolytic drugs) was not allowed and constituted a major protocol deviation.

The procedure had to be started (defined as the time of arterial access) no earlier than 6 hours, but before 24 hours, from TLSW and performed according to the most current Instructions for Use (IFU) of the device. It was recommended that the interventional procedure starts within 60 minutes of randomization and is completed within two (2) hours of arterial access. No more than six (6) retrieval attempts in the same vessel using any of the available Trevo devices and no more than three (3) passes per Trevo device were allowed. In cases of tandem extracranial high-grade stenosis or occlusion with accompanying intracranial occlusion, stenting of the extracranial lesion was not allowed. However, angioplasty could be performed if considered necessary in order to access the intracranial lesion.

Supplemental Text Section S7: Post hoc models to assess for the effect of imbalances in baseline variables.

Table 2 Univariate Logistic Regression modelled for mRS 0-2

Ratio for Binary Variables	Coefficient	Standard Error	Odds Ratio [95% CI]	p-value
Trevo vs MM	1.8331	0.3549	6.25 (3.12-12.5)	< 0.0001
Baseline clot location M1 vs ICA	-0.1941	0.3717	0.82 (0.40-1.71)	0.6016
Male vs Female	0.0595	0.3010	1.06 (0.59-1.91)	0.8433
Pre-NIHSS less than 20 Yes vs No	1.3199	0.3719	3.74 (1.81-7.76)	0.0004
Pre-mRS 1 vs 0	-0.2176	0.6942	0.80 (0.21-3.14)	0.7540
Blood Glucose (mg/dl)	-0.0143	0.00475	0.99 (0.98-1.00)	0.0026
Extracranial Carotid Artery Disease history Yes vs No	-0.0268	0.5200	0.97 (0.35-2.70)	0.9589
Peripheral Vascular Disease history Yes vs No	-0.1133	0.5563	0.89 (0.30-2.66)	0.8386
Atrial Fibrillation history Yes vs No	-0.2223	0.3278	0.80 (0.42-1.52)	0.4976
Diabetes Mellitus history Yes vs No	-1.0367	0.3885	0.35 (0.17-0.76)	0.0076
Heart Failure Yes history vs No	-0.3648	0.4222	0.69 (0.30-1.59)	0.3875
Dyslipidemia history Yes vs No	0.1515	0.3096	1.16 (0.63-2.13)	0.6245
TIA history Yes vs No	0.6438	0.5417	1.90 (0.66-5.50)	0.2347
Previous Ischemic Stroke Yes vs No	-0.6049	0.5302	0.55 (0.19-1.54)	0.2539

Ratio for Binary Variables	Coefficient	Standard Error	Odds Ratio [95% CI]	p-value
IV lytic administered before randomization Yes vs No	-0.1978	0.5490	0.82 (0.28-2.41)	0.7186
Stenosis > 50%(Baseline Carotid Stenosis by CTA or MRA Yes vs No)	-0.4945	0.3613	0.61 (0.30-1.24)	0.1711
Occlusion location by CTA MRA Right vs Left	0.4185	0.3017	1.52 (0.84-2.75)	0.1653
Wake-up Stroke vs. other	0.1854	0.3031	1.20 (0.66-2.18)	0.5408
Core Infarct size by RAPID	-0.0323	0.0151	0.97 (0.94-1.00)	0.0325
Age (yrs)	-0.0336	0.0113	0.97 (0.95-0.99)	0.0028
Baseline Aspects score (Core lab)	0.2755	0.1110	1.32 (1.06-1.64)	0.0131

Functional Independence (mRS 0-2) was imputed by LOCF

Table 3 Multivariate Step wise model. Predictors of functional independence (mRS 0-2)

Ratio for Binary Variables	Coefficient	Standard Error	Odds Ratio [95% CI]	p-value
Age (Yrs)	-0.0487	0.0143	0.95 (0.93-0.98)	0.0007
Core Infarct size by RAPID	-0.00605	0.0227	0.99 (0.95-1.04)	0.7897
Previous Diabetes Mellitus No vs Yes	1.0152	0.4526	2.76 (1.14-6.70)	0.0249
Trevo vs MM	2.2315	0.4217	9.31 (4.08-21.3)	< 0.0001
Baseline NIHSS < 20 vs NIHSS >= 20	1.4161	0.4595	4.12 (1.67-10.1)	0.0021

Table 3. Multivariate Logistic Regression model. Functional Independence (mRS 0-2) by propensity score matching

Ratio for Binary Variables	Coefficient	Standard Error	Odds Ratio [95% CI]	p-value
Age (Yrs)	-0.0435	0.0155	0.96 (0.93-0.99)	0.0050
Trevo vs MM	2.3286	0.4617	10.3 (4.15-25.4)	< 0.0001
Previous Diabetes Mellitus No vs Yes	1.5278	0.5480	4.61 (1.57-13.5)	0.0053
Baseline NIHSS < 20 vs NIHSS >= 20	1.9621	0.5876	7.11 (2.25-22.5)	0.0008

References

1. Berger C, Fiorelli M, Steiner T, et al. Hemorrhagic transformation of ischemic brain tissue: asymptomatic or symptomatic? Stroke; a journal of cerebral circulation 2001;32:1330-5.
2. Bath PM, Lees KR, Schellinger PD, et al. Statistical analysis of the primary outcome in acute stroke trials. Stroke; a journal of cerebral circulation 2012;43:1171-8.
3. Saver JL. Optimal end points for acute stroke therapy trials: best ways to measure treatment effects of drugs and devices. Stroke; a journal of cerebral circulation 2011;42:2356-62.
4. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. Stroke; a journal of cerebral circulation 2007;38:3055-62.
5. Hong KS, Saver JL. Quantifying the value of stroke disability outcomes: WHO global burden of disease project disability weights for each level of the modified Rankin Scale. Stroke; a journal of cerebral circulation 2009;40:3828-33.
6. Rivero-Arias O, Ouellet M, Gray A, Wolstenholme J, Rothwell PM, Luengo-Fernandez R. Mapping the modified Rankin scale (mRS) measurement into the generic EuroQol (EQ-5D) health outcome. Med Decis Making 2010;30:341-54.
7. Chaisinanunkul N, Adeoye O, Lewis RJ, et al. Adopting a Patient-Centered Approach to Primary Outcome Analysis of Acute Stroke Trials Using a Utility-Weighted Modified Rankin Scale. Stroke; a journal of cerebral circulation 2015;46:2238-43.
8. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. The New England journal of medicine 1988;318:1728-33.
9. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995;310:452-4.
10. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014;312:1342-3.
11. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health Serv Res 1976;11:478-507.
12. Kaplan RM. The minimally clinically important difference in generic utility-based measures. COPD 2005;2:91-7.

13. Chaisinanunkul NN SS, Gornbein JA, et al. Abstract 138: Assessing Disability Outcome in Acute Stroke Clinical Trials: Increased Study Power with Linear (ALDS) Compared with Ordinal (mRS) Scales. . Stroke;45:A138.
14. Kelly A, Rush J, Shafonsky E, et al. Detecting short-term change and variation in health-related quality of life: within- and between-person factor structure of the SF-36 health survey. Health Qual Life Outcomes 2015;13:199.
15. Gershon RC, Lai JS, Bode R, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. Qual Life Res 2012;21:475-86.
16. Lansberg MG, Schrooten M, Bluhmki E, Thijs VN, Saver JL. Treatment time-specific number needed to treat estimates for tissue plasminogen activator therapy in acute stroke based on shifts over the entire range of the modified Rankin Scale. Stroke; a journal of cerebral circulation 2009;40:2079-84.
17. Saver JL, Gornbein J, Grotta J, et al. Number needed to treat to benefit and to harm for intravenous tissue plasminogen activator therapy in the 3- to 4.5-hour window: joint outcome table analysis of the ECASS 3 trial. Stroke; a journal of cerebral circulation 2009;40:2433-7.
18. Howard G, Waller JL, Voeks JH, et al. A simple, assumption-free, and clinically interpretable approach for analysis of modified Rankin outcomes. Stroke; a journal of cerebral circulation 2012;43:664-9.
19. Rawls J. A Theory of Justice. Belknap Press 1971.
20. Harsanyi JC. Cardinal Utility in Welfare Economics and in the Theory of Risk-taking. . Journal of Political Economy 1953;61:434-5.
21. Harsanyi JC. Games with Incomplete Information Played by “Bayesian” Players, I–III Part I. The Basic Model. Management Science 1967;14:159-82.
22. Saver JL, Goyal M, van der Lugt A, et al. Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-analysis. JAMA 2016;316:1279-88.
23. Fransen PS, Beumer D, Berkhemer OA, et al. MR CLEAN, a multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke in the Netherlands: study protocol for a randomized controlled trial. Trials 2014;15:343.
24. Goyal M, Demchuk AM, Menon BK, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. The New England journal of medicine 2015;372:1019-30.
25. Saver JL, Goyal M, Bonafe A, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. The New England journal of medicine 2015;372:2285-95.

26. Campbell BC, Mitchell PJ, Kleinig TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. N Engl J Med 2015;372:1009-18.
27. Jovin TG, Chamorro A, Cobo E, et al. Thrombectomy within 8 Hours after Symptom Onset in Ischemic Stroke. The New England journal of medicine 2015.
28. Bracard S, Ducrocq X, Mas JL, et al. Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (THRACE): a randomised controlled trial. Lancet Neurol 2016;15:1138-47.
29. Muir KW, Ford GA, Messow CM, et al. Endovascular therapy for acute ischaemic stroke: the Pragmatic Ischaemic Stroke Thrombectomy Evaluation (PISTE) randomised, controlled trial. J Neurol Neurosurg Psychiatry 2017;88:38-44.
30. Jovin T, Gupta R, Liebeskind D, et al. Imaging Based Intraarterial Treatment for Acute Anterior Circulation Ischemic Stroke Due to Intracranial Large Vessel Occlusion beyond 8 Hours from Symptoms Onset – Preliminary Multicenter Experience. 61st Annual Meeting of the American Academy of Neurology; 2009; Seattle, WA.
31. Lansberg MG, Cereda CW, Mlynash M, et al. Response to endovascular reperfusion is not time-dependent in patients with salvageable tissue. Neurology 2015;85:708-14.
32. Nogueira RG, Kemmling A, Souza LM, et al. Clinical diffusion mismatch better discriminates infarct growth than mean transit time-diffusion weighted imaging mismatch in patients with middle cerebral artery-M1 occlusion and limited infarct core. J Neurointerv Surg 2017;9:127-30.
33. Bouslama M, Bowen MT, Haussen DC, et al. Selection Paradigms for Large Vessel Occlusion Acute Ischemic Stroke Endovascular Therapy. Cerebrovasc Dis 2017;44:277-84.
34. Nogueira R.G., G.W. A, D.C. H, et al. Clinical-Imaging Mismatch versus Perfusion Imaging Mismatch Selection for Stroke Patients Undergoing Mechanical Thrombectomy: a Post-Hoc Analysis of the SWIFT PRIME Trial. . European Stroke Organization Congress; May 10-12, 2016; Barcelona, Spain.
35. Zaidi SF, Aghaebrahim A, Urra X, et al. Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy. Stroke; a journal of cerebral circulation 2012;43:3238-44.
36. Yoo AJ, Chaudhry ZA, Nogueira RG, et al. Infarct volume is a pivotal biomarker after intra-arterial stroke therapy. Stroke; a journal of cerebral circulation 2012;43:1323-30.

37. Ribo M, Tomasello A, Lemus M, et al. Maximal Admission Core Lesion Compatible With Favorable Outcome in Acute Stroke Patients Undergoing Endovascular Procedures. Stroke; a journal of cerebral circulation 2015;46:2849-52.
38. Lima A, Haussen DC, Rebello LC, et al. Endovascular Therapy for Large Vessel Stroke in the Elderly: Hope in the New Stroke Era. Cerebrovasc Dis 2016;42:421-7.
39. Rangaraju S, Liggins JT, Aghaebrahim A, et al. Pittsburgh outcomes after stroke thrombectomy score predicts outcomes after endovascular therapy for anterior circulation large vessel occlusions. Stroke; a journal of cerebral circulation 2014;45:2298-304.
40. Broderick JP, Palesch YY, Demchuk AM, et al. Endovascular therapy after intravenous t-PA versus t-PA alone for stroke. The New England journal of medicine 2013;368:893-903.
41. Farzin B, Fahed R, Guilbert F, et al. Early CT changes in patients admitted for thrombectomy: Intrarater and interrater agreement. Neurology 2016;87:249-56.
42. Cereda CW, Christensen S, Campbell BC, et al. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. J Cereb Blood Flow Metab 2016;36:1780-9.
43. Albers GW, Goyal M, Jahan R, et al. Ischemic core and hypoperfusion volumes predict infarct size in SWIFT PRIME. Ann Neurol 2016;79:76-89.
44. Menjot de Champfleur N, Saver JL, Goyal M, et al. Efficacy of Stent-Retriever Thrombectomy in Magnetic Resonance Imaging Versus Computed Tomographic Perfusion-Selected Patients in SWIFT PRIME Trial (Solitaire FR With the Intention for Thrombectomy as Primary Endovascular Treatment for Acute Ischemic Stroke). Stroke; a journal of cerebral circulation 2017;48:1560-6.
45. Hacke W, Kaste M, Bluhmki E, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. The New England journal of medicine 2008;359:1317-29.
46. Marks MP, Olivot JM, Kemp S, et al. Patients with acute stroke treated with intravenous tPA 3-6 hours after stroke onset: correlations between MR angiography findings and perfusion- and diffusion-weighted imaging in the DEFUSE study. Radiology 2008;249:614-23.
47. Jauch EC, Saver JL, Adams HP, Jr., et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke; a journal of cerebral circulation 2013;44:870-947.

48. European Stroke Organisation Executive C, Committee ESOW. Guidelines for management of ischaemic stroke and transient ischaemic attack 2008. Cerebrovasc Dis 2008;25:457-507.
49. Hill K, Acute Stroke Management Expert Working G. Australian Clinical Guidelines for Acute Stroke Management 2007. Int J Stroke 2008;3:120-9.
50. Casaubon LK, Boulanger JM, Blacquiere D, et al. Canadian Stroke Best Practice Recommendations: Hyperacute Stroke Care Guidelines, Update 2015. Int J Stroke 2015;10:924-40.
51. National Institute of Neurological D, Stroke rt PASSG. Tissue plasminogen activator for acute ischemic stroke. The New England journal of medicine 1995;333:1581-7.