



A statistical approach to identify optimal inclusion criteria: An application to acute stroke clinical trials

Robyn L. Ball^a, Bin Jiang^b, Manisha Desai^a, Patrik Michel^d, Ashraf Eskandari^d, Tudor Jovin^c, Max Wintermark^{b,*}

^a Department of Medicine, Quantitative Sciences Unit, Stanford University, Stanford, CA, 1070 Arastradero Rd, Palo Alto, CA, 94305, USA

^b Department of Radiology, Neuroradiology Section, Stanford University School of Medicine, Stanford, CA, 300 Pasteur Dr., Room S047, Stanford, CA, 94305, USA

^c Department of Neurology, University of Pittsburgh, Pittsburgh, PA, 811 Kaufmann Medical Building, 3471 Fifth Ave, Pittsburgh, PA, 15213, USA

^d Department of Neurology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland, BH13.713, Rue Du Bugnon 46, 1011, Lausanne, Switzerland

ARTICLE INFO

Keywords:

Inclusion criteria
Patient selection
Clinical trial
Acute stroke

ABSTRACT

Purpose: To develop a statistical approach that compares patient selection strategies across clinical trials and apply this approach to acute ischemic stroke clinical trials to identify the optimal inclusion criteria.

Methods: We developed a statistical approach that compares the number needed to treat to achieve one success (NNT) along with the number needed to screen to achieve one success (NNS) and assesses if there are significant differences in inclusion criteria, treatment course, and clinical outcome among patients that may have been included/excluded in the trials. We applied this approach to the study population from four recent positive acute stroke clinical trials: MR CLEAN, EXTEND-IA, ESCAPE, and SWIFT PRIME, applying published trial criteria to an independent registry of 612 acute stroke patients, since we did not have access to the complete trial data.

Results: Although reported NNT were similar for EXTEND-IA, SWIFT PRIME and ESCAPE, and somewhat higher for MR CLEAN, NNS varied across the trials from 21 for EXTEND-IA, 27 for MR CLEAN, to 46 for ESCAPE and 64 for SWIFT PRIME, reflecting less and more stringent inclusion criteria, respectively. Although there were significant differences in imaging biomarkers and other clinical characteristics among patients that may have been included/excluded in the trials, these differences did not translate to significant differences in treatment course or clinical outcomes.

Conclusions: Our study proposes a robust statistical approach that can be applied to a larger pooled trial dataset, if made available, to objectively compare across clinical trials and inform inclusion criteria of future trials. Pooled analysis of the acute stroke trial data is needed to determine which imaging biomarker inclusion criteria are critical and which may be relaxed. If this procedure were applied across the pooled trial data, it could decrease costs and refine the design of future trials to be the most efficacious for the greatest number of patients.

1. Introduction

Utilization of study population data is essential to clinical trial design. Optimizing inclusion criteria can not only decrease costs associated with expensive screening strategies but it can also help ensure trials are both efficient and include patients who may benefit from the trial. With an increasing amount of trial data becoming available, straightforward statistical approaches and methodologies are poised to help clinical researchers best utilize existing trial data to inform the design of upcoming trials. For example, there has been debate regarding the use of advanced imaging biomarkers in recent positive acute stroke clinical trials [1]. Screening for advanced imaging biomarkers is

expensive and it is critical that trials are designed to benefit the greatest number of patients.

Four recent clinical trials have demonstrated the value of endovascular revascularization therapy in patients with acute ischemic stroke up to 24 h since symptom onset, including MR CLEAN [2], ESCAPE [3], SWIFT PRIME [4], and EXTEND-IA [5]. In comparison with the earlier trials which showed negative results [6], more patients benefitted from endovascular therapy with mechanical devices and the average modified Rankin score improved, in varying degrees. The major difference of these recent positive trials compared to the older negative trials was the use of imaging biomarkers to select patients for acute revascularization therapy. However, the positive trials differ from

* Corresponding author. Stanford University, Department of Radiology, Neuroradiology Division, 300 Pasteur Drive, Room S047, Stanford, CA, 94305-5105, USA.
E-mail address: Max.Wintermark@gmail.com (M. Wintermark).

each other in terms of the exact imaging biomarkers used for inclusion criteria.

Noncontrast brain CT and CT angiography (CTA) were the primary imaging biomarkers used in the aforementioned four clinical trials, while perfusion-CT (PCT) and/or collateral imaging was employed in EXTEND-IA and SWIFT PRIME to determine the infarct core and penumbra. In these trials, only those patients with relative large mismatch (> 1.2 penumbra mismatch or > 1.8 target mismatch, respectively) were considered for enrollment. This strategy has resulted in concern that these trials succeeded because of selection bias, “cherry-picking” the patients that may have the best outcome while excluding others who may also have benefitted. It is critical to refine the design of clinical trials to be the most efficacious for the greatest number of patients.

This study aims to develop a straightforward statistical approach that can be used to identify optimal inclusion criteria and inform future clinical trial design. We present an approach that estimates the number needed to screen to achieve one success (NNS), using trial results and inclusion criteria, which could be used to evaluate the feasibility of a trial under consideration and assess inclusion criteria across trials. We illustrate the statistical approach by comparing imaging biomarker-based selection strategies across four acute stroke clinical trials.

2. Methods

2.1. Data

Ethics approval was obtained from our local Institutional Review Board (IRB) and since this was a retrospective study, the IRB waived the need for patient consent. Ideally, this study would utilize the pooled trial data but since the complete trial data are unavailable, we compared the inclusion criteria of imaging biomarkers across trials through utilizing the published results and criteria from the MR CLEAN, ESCAPE, SWIFT PRIME, and EXTEND-IA trials (Table 1) in combination with an independent registry of stroke imaging data [1]. Published trial results include the demographic and imaging biomarker inclusion criteria, number of patients in the treated and control arms, proportion of patients in each arm with a positive outcome, as well as number needed to treat (NNT). The previously characterized independent registry [1] includes baseline parenchymal, perfusion, and vascular imaging data of 612 acute ischemic stroke patients, as well as the time since symptom onset, demographic information, treatment course, and clinical outcomes of these patients. An interactive web application, available at

Table 1
Published inclusion criteria and trial results.

Inclusion criteria	MR CLEAN	ESCAPE	SWIFT PRIME	EXTEND-IA
Age (years)	18+	18+	18–80	18+
Time since symptom onset (hours)	< 6	< 12	< 6	< 6
Location	ICA or M1 or M2 (A1& A2)	ICA or M1	ICA or M1	ICA or M1 or M2
Infarct volume	any	any	< 50 ml	< 70 ml
ASPECT	any	6–10	6–10	any
Target mismatch	any	any	> 1.8	any
Penumbra mismatch	any	any	any	> 1.2
Penumbra volume	any	any	> 15 ml	> 10 ml
Collateral	any	2 or 3	any	any
Trial results	MR CLEAN	ESCAPE	SWIFT PRIME	EXTEND-IA
Control (N)	267	150	98	35
Treated (N)	233	165	98	35
Absolute Risk Reduction (ARR)	0.14	0.24	0.24	0.31

<https://neuroradiology.shinyapps.io/strokeimbio/>, can be used to explore this registry.

2.2. Statistical methods

From the published trial results, we extracted the absolute risk reduction (ARR) as the proportion of patients in the treatment arm with a good outcome ($mRS \leq 2$) minus the proportion of patients in the control arm with a good outcome; the number needed to treat (NNT) is the inverse of ARR.

For each trial, we used the published inclusion criteria to determine which patients from the independent registry met the initial screening criteria for possible enrollment based on age and time since symptom onset (N). Of those patients who met the screening criteria, we identified those that could have been included in each trial based on baseline parenchymal, perfusion, and vascular imaging biomarkers (n). Next, for each trial, we calculated the proportion ($p = n/N$) of those who met the screening criteria that could have been included in the trial based on the baseline imaging biomarkers. Using this proportion and the ARR, the number needed to screen to achieve one success (NNS) is the inverse of ARR times p , $1/(ARR \cdot p)$.

To assess the uncertainty around these estimates, we used the bootstrap method and sampled with replacement from the independent registry 10,000 times. Complete statistical details and the algorithm can be found in Supplemental Methods in the Online Supplement.

We compared the demographics, imaging biomarkers, treatment course, and clinical outcome of patients across trials by considering subgroups of patients that met the inclusion criteria of a trial but who would have been excluded from another trial. For categorical outcomes, we used Fisher's exact test and for continuous outcomes, we used the nonparametric Wilcoxon rank sum test. Significance was assessed at the 0.05 level. All statistical analyses were carried out in the R statistical computing interface [7].

3. Results

From the published trial results, we tabulated the inclusion criteria and calculated the absolute risk reduction (Table 1). Applying the published inclusion criteria to the independent registry of 612 acute ischemic stroke patients with < 24 h since symptom onset, the following met the screening criteria for possible enrollment, based on age and time since symptom onset: 445 (72.7%) for MR CLEAN and EXTEND-IA, 541 (88.4%) for ESCAPE, and 372 (60.2%) for SWIFT PRIME. Of those who met the screening criteria, the following could have been included in the trial based on imaging biomarkers: 120 for MR CLEAN (27.0%), 72 (16.2%) for EXTEND-IA, 50 (9.2%) for ESCAPE, and 25 (6.7%) for SWIFT PRIME. We found that those who could have been included in SWIFT PRIME were a subset of those who could have been included in EXTEND-IA, and these were a subset of those patients who could have been included in MR CLEAN (Fig. 1), illustrating the progression of stricter imaging inclusion criteria. While there was overlap among those patients who met the inclusion criteria of ESCAPE and the other three trials (Fig. 1), ESCAPE included some patients with longer time since symptom onset that would have been excluded from all three of the other trials.

Per the reported results in the literature, NNT was similar for EXTEND-IA [5], SWIFT PRIME [4] and ESCAPE [3], and somewhat higher for MR CLEAN [2] (Fig. 2A). NNS was more variable and reflected the relative stringency of the inclusion criteria. NNS was 21 for EXTEND-IA, 27 for MR-CLEAN, 46 for ESCAPE and 64 for SWIFT PRIME (Fig. 2B).

As seen in Table 2, patients that could have been included in MR CLEAN but excluded from EXTEND-IA ($N = 48$), had lower median ASPECT scores (7 compared to 8; $p = 0.006$), larger median infarct volumes (83.3 ml compared to 17.9 ml; $p < 0.001$), smaller median penumbra volumes (43.8 ml compared to 101.2 ml; $p < 0.001$) and

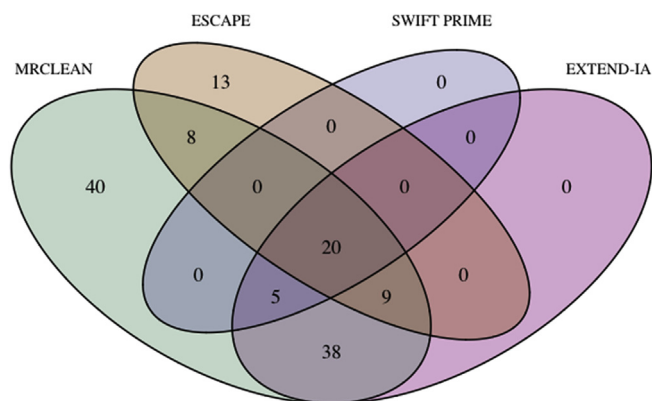


Fig. 1. Number of patients in the independent registry that could have been included in the trials based on published inclusion criteria.

thus, smaller median target mismatch (1.8 compared to 6.2; $p < 0.001$).

SWIFT PRIME had the strictest inclusion criteria of the four trials, resulting in only 25 of 445 (6.7%) registry patients eligible to be included in the trial who met the screening criteria. Of the patients that could have been included in MR CLEAN but excluded from SWIFT PRIME ($N = 95$), 61 of 95 (64.2%; $p < 0.001$) would have been excluded because of M2 involvement. Additionally, patients that would have been excluded from SWIFT PRIME but who could have been included in MR CLEAN had significantly higher median age (73 compared to 66; $p = 0.034$), larger median infarct volumes (41.4 ml compared to 15.8 ml; $p = 0.001$), smaller median penumbra volumes (62.6 ml compared to 118.5 ml; $p < 0.001$), resulting in lower median target mismatch (2.4 compared to 7.1; $p < 0.001$).

ESCAPE was not quite as strict as SWIFT PRIME in its inclusion criteria but differed from other trials by including patients up to 12 h since symptom onset while excluding patients with less than two collaterals (Table 1). Not surprisingly, of patients that could have been included in MR CLEAN but excluded from ESCAPE, 47 of 83 (56.6%; $p < 0.001$) were excluded due to having less than two collaterals and 61 of 83 (73.5%; $p < 0.001$) were excluded due to M2 involvement. Additionally, patients that would have been excluded from ESCAPE but

who could have been included in MR CLEAN had larger median infarct volume (38.0 ml compared to 18.4 ml; $p = 0.002$), smaller median penumbra volume (57.4 ml compared to 102.3 ml; $p = 0.001$), resulting in smaller median target mismatch (2.4 compared to 6.1; $p < 0.001$).

These differences among patients from the independent registry that could have been included/excluded from the trials did not translate to statistically significant differences in treatment course or clinical outcome in these patient subgroups (Table 2). For instance, the treatment effect of endovascular therapy and clinical outcomes in stroke patients who were MR CLEAN candidates but not EXTEND-IA candidates was not significantly different from the treatment effect of endovascular therapy and clinical outcomes in stroke patients who were MR CLEAN candidates and EXTEND-IA candidates.

4. Discussion

Despite having a smaller number needed to treat to achieve one success (NNT), acute stroke trials involving advanced imaging (for instance, EXTEND-IA) resulted in a similar number needed to screen to achieve one success (NNS) as those for acute stroke trials involving conventional imaging (for instance, MR CLEAN). This finding indicates that advanced imaging inclusion criteria result in a more pronounced, and stricter, selection of the screened population. Patients meeting imaging inclusion criteria for acute stroke trials involving advanced imaging were subsets of patients meeting imaging inclusion criteria for acute stroke trials involving conventional imaging.

Ideally, this approach would be applied to the pooled complete trial data, providing the opportunity for a robust assessment and identification of the optimal inclusion criteria. Since these data were unavailable, we illustrated the approach using an independent registry. We tried to determine whether different outcomes following revascularization therapy were observed in the patients meeting imaging inclusion criteria for acute stroke trials involving conventional imaging but not for acute stroke trials involving advanced imaging, but we did not find any statistically significant differences in treatment course or clinical outcome. However, the sample size of the independent registry is limited, and we lacked sufficient power to detect small differences. Further and most importantly, the patients included in the independent registry were not treated based on the inclusion criteria of the different trials, which limits the interpretation that can be made of the outcomes

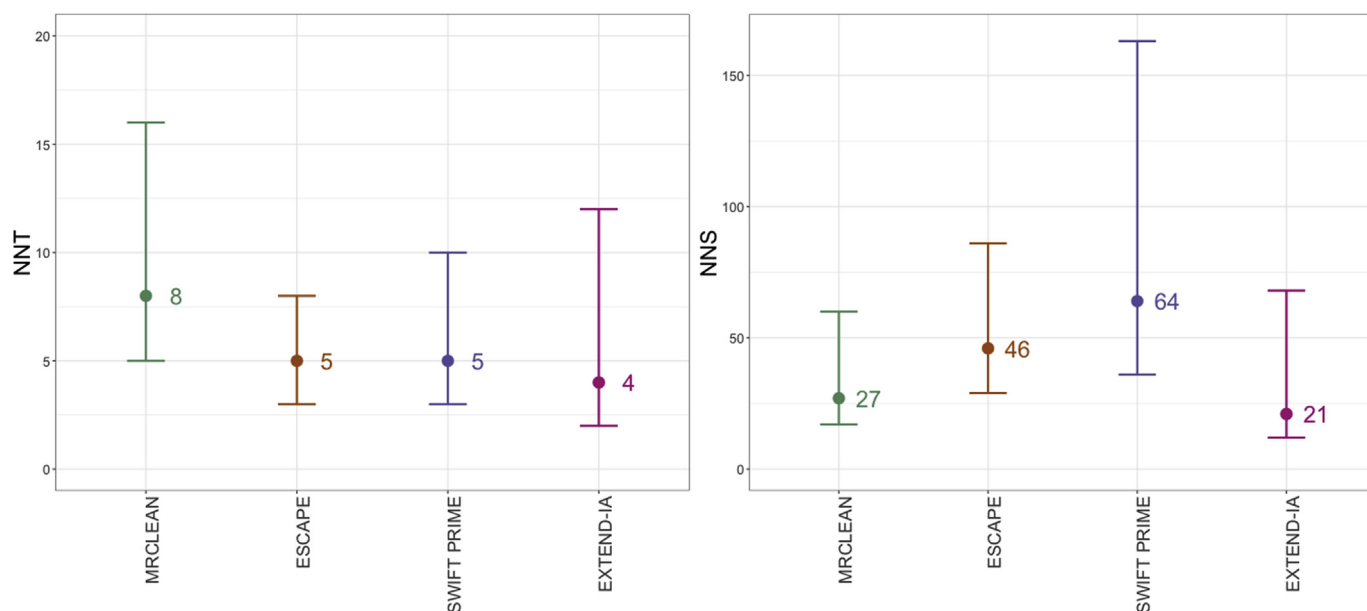


Fig. 2. (A) Number needed to treat to achieve one success (NNT) and (B) Number needed to screen to achieve one success. Bars represent 95% bootstrap confidence intervals around the estimates.

Table 2

Comparison of patients in the registry that could have been included in MRCLEAN but excluded from another trial versus those that could have been included in either MRCLEAN or another trial. For example, columns 2 and 3 refer to patients in the registry that could have been included in MRCLEAN but excluded from EXTEND-IA (n = 48) versus those that could have been included in either MRCLEAN or EXTEND-IA (n = 72); see Fig. 1. Continuous characteristics summarized with median [IQR] and categorical Characteristics summarized with N (%). For categorical outcomes, we used Fisher's exact test and for continuous outcomes, we used the nonparametric Wilcoxon rank sum test. Significant p-values in bold.

Imaging Biomarkers	EXTEND-IA	MRCLEAN	p-value	SWIFT PRIME	MRCLEAN	p-value	ESCAPE	MRCLEAN	p-value
Number compared	72	48		25	95		50	83	
Time since symptom onset (in hours)	2.5 [1.8, 3.0]	2.5 [1.5, 3.0]	0.244	2.5 [2.0, 3.0]	2.5 [1.5, 3.0]	0.309	3.0 [2.0, 5.8]	2.5 [1.6, 3.0]	0.003
ASPECT	8.0 [7.0, 10.0]	7.0 [6.0, 9.0]	0.006	8.0 [8.0, 9.0]	8.0 [6.5, 10.0]	0.318	8.0 [8.0, 9.0]	8.0 [6.5, 10.0]	0.096
Infarct volume (in ml)	17.9 [7.5, 32.5]	83.3 [49.2, 107.0]	< 0.001	15.8 [7.3, 30.7]	41.4 [15.1, 83.3]	0.001	18.4 [6.5, 38.1]	38.0 [16.8, 80.3]	0.002
Penumbra volumes (in ml)	101.2 [57.1, 123.6]	43.8 [23.5, 81.4]	< 0.001	118.5 [100.7, 152.8]	62.6 [34.0, 102.9]	< 0.001	102.3 [69.6, 134.0]	57.4 [34.0, 105.2]	0.001
Target mismatch (P+I)/I	6.2 [3.7, 10.6]	1.8 [1.4, 2.0]	< 0.001	7.1 [4.2, 18.4]	2.4 [1.8, 5.0]	< 0.001	6.1 [3.0, 17.9]	2.4 [1.8, 5.0]	< 0.001
Age (in years)	68.5 [59.0, 80.0]	73.0 [59.8, 78.3]	0.841	66.0 [57.0, 73.0]	73.0 [60.0, 80.5]	0.034	72.5 [56.0, 79.0]	67.0 [59.0, 78.5]	0.726
Collateral (%)			0.086			0.133			< 0.001
0	4 (5.6)	2 (4.2)		0 (0.0)	6 (6.3)		0 (0.0)	6 (7.2)	
1	17 (23.6)	22 (45.8)		5 (20.0)	34 (35.8)		0 (0.0)	39 (47.0)	
2	32 (44.4)	16 (33.3)		11 (44.0)	37 (38.9)		25 (50.0)	27 (32.5)	
3	19 (26.4)	8 (16.7)		9 (36.0)	18 (18.9)		25 (50.0)	11 (13.3)	
ICA (%)	11 (15.3)	2 (4.2)	0.073	7 (28.0)	6 (6.3)	0.006	15 (30.0)	3 (3.6)	< 0.001
M1 (%)	28 (38.9)	18 (37.5)	1	18 (72.0)	28 (29.5)	< 0.001	35 (70.0)	19 (22.9)	< 0.001
M2 (%)	33 (45.8)	28 (58.3)	0.197	0 (0.0)	61 (64.2)	< 0.001	0 (0.0)	61 (73.5)	0.001
Treatment course and clinical outcome	EXTEND-IA	MRCLEAN	p-value	SWIFT PRIME	MRCLEAN	p-value	ESCAPE	MRCLEAN	p-value
Number treated (%)	49 (68.1)	38 (79.2)	0.214	19 (76.0)	68 (71.6)	0.803	30 (60.0)	59 (71.1)	0.254
Treatment ^a for those treated (%)			0.764			1			0.514
IV	47 (95.9)	36 (94.7)		19 (100.0)	64 (94.1)		28 (93.3)	57 (96.6)	
IA	1 (2.0)	2 (5.3)		0 (0.0)	3 (4.4)		2 (6.7)	1 (1.7)	
IV and IA	1 (2.0)	0 (0.0)		0 (0.0)	1 (1.5)		0 (0.0)	1 (1.7)	
Recanalization (%)	41 (83.7)	26 (68.4)	0.125	16 (84.2)	51 (75.0)	0.543	27 (90.0)	42 (71.2)	0.06
mRS ≤ 2 (%)	22 (44.9)	21 (55.3)	0.391	11 (57.9)	32 (47.1)	0.446	15 (50.0)	30 (50.8)	1
mRS (%)			0.774			0.75			0.631
0	7 (14.3)	3 (7.9)		4 (21.1)	6 (8.8)		3 (10.0)	7 (11.9)	
1	8 (16.3)	11 (28.9)		3 (15.8)	16 (23.5)		5 (16.7)	14 (23.7)	
2	7 (14.3)	7 (18.4)		4 (21.1)	10 (14.7)		7 (23.3)	9 (15.3)	
3	10 (20.4)	4 (10.5)		3 (15.8)	11 (16.2)		6 (20.0)	8 (13.6)	
4	7 (14.3)	5 (13.2)		3 (15.8)	9 (13.2)		2 (6.7)	10 (16.9)	
5	1 (2.0)	1 (2.6)		0 (0.0)	2 (2.9)		0 (0.0)	2 (3.4)	
6	6 (12.2)	4 (10.5)		1 (5.3)	9 (13.2)		4 (13.3)	6 (10.2)	
missing	3 (6.1)	3 (7.9)		1 (5.3)	5 (7.4)		3 (10.0)	3 (5.1)	

^a IV = intravenous drug, IA = intraarterial, endovascular drug/technique.

in the different subgroups of this registry. However, if this methodology were applied to the pooled complete trial data and significant differences were found among imaging criteria and their relation to clinical outcomes, inclusion criteria for future trials could be optimized, allowing for more prescriptive criteria for imaging biomarkers that are significantly associated with clinical outcomes and less strict, more inclusive criteria, for those imaging biomarkers that are not significantly associated with clinical outcomes.

Despite these limitations, our study should be seen as a robust statistical approach that can be applied to a larger pooled trial dataset, if made available, to objectively compare across clinical trials and inform inclusion criteria of future trials and decrease the high costs associated with these trials. Pooled analysis of the acute stroke trial data is needed to compare clinical outcomes between subgroups and to determine which imaging biomarker inclusion criteria are critical and which may be relaxed. If this procedure were applied across the pooled trial data, it could refine the design of future trials to be the most efficacious for the greatest number of patients.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.conctc.2019.100355>.

doi.org/10.1016/j.conctc.2019.100355.

Disclosures

None.

References

- [1] B. Jiang, R.L. Ball, P. Michel, T. Jovin, M. Desai, A. Eskandari, et al., Prevalence of imaging biomarkers to guide the planning of acute stroke reperfusion trials, *Stroke* 48 (2017) 1675–1677.
- [2] O.A. Berkhemer, P.S. Fransen, D. Beumer, L.A. van den Berg, H.F. Lingsma, A.J. Yoo, et al., A randomized trial of intraarterial treatment for acute ischemic stroke, *N. Engl. J. Med.* 372 (2015) 11–20.
- [3] M. Goyal, A.M. Demchuk, B.K. Menon, M. Eesa, J.L. Rempel, J. Thornton, et al., Randomized assessment of rapid endovascular treatment of ischemic stroke, *N. Engl. J. Med.* 372 (2015) 1019–1030.
- [4] J.L. Saver, M. Goyal, A. Bonafe, H.C. Diener, E.I. Levy, V.M. Pereira, et al., Stent-retriever thrombectomy after intravenous t-pa vs. t-pa alone in stroke, *N. Engl. J. Med.* 372 (2015) 2285–2295.
- [5] B.C. Campbell, P.J. Mitchell, T.J. Kleinig, H.M. Dewey, L. Churilov, N. Yassi, et al., Endovascular therapy for ischemic stroke with perfusion-imaging selection, *N. Engl. J. Med.* 372 (2015) 1009–1018.
- [6] J.P. Broderick, Y.Y. Palesch, A.M. Demchuk, S.D. Yeatts, P. Khatri, M.D. Hill, Endovascular therapy after intravenous t-pa versus t-pa alone for stroke, *N. Engl. J. Med.* 368 (2013) 893–903.
- [7] R Core Team, A Language and Environment for Statistical Computing, The R Project for Statistical Computing, 2015, <https://www.R-project.org/>.