

RESEARCH ARTICLE

Open Access



# Improving quality of stroke care through benchmarking center performance: why focusing on outcomes is not enough

Marzyeh Amini<sup>1\*</sup> , Nikki van Leeuwen<sup>1</sup>, Frank Eijkenaar<sup>2</sup>, Maxim J. H. L. Mulder<sup>3</sup>, Wouter Schonewille<sup>4</sup>, Geert Lycklama à Nijeholt<sup>5</sup>, Wouter H. Hinsenveld<sup>6</sup>, Robert-Jan B. Goldhoorn<sup>6</sup>, Pieter Jan van Doormaal<sup>7</sup>, Sjoerd Jenniskens<sup>8</sup>, Jan Hazelzet<sup>1</sup>, Diederik W. J. Dippel<sup>3</sup>, Bob Roozenbeek<sup>3</sup>, Hester F. Lingsma<sup>1</sup> and on behalf of the MR CLEAN Registry Investigators

## Abstract

**Background:** Between-center variation in outcome may offer opportunities to identify variation in quality of care. By intervening on these quality differences, patient outcomes may be improved. However, whether observed differences in outcome reflect the true quality improvement potential is not known for many diseases. Therefore, we aimed to analyze the effect of differences in performance on structure and processes of care, and case-mix on between-center differences in outcome after endovascular treatment (EVT) for ischemic stroke.

**Methods:** In this observational cohort study, ischemic stroke patients who received EVT between 2014 and 2017 in all 17 Dutch EVT-centers were included. Primary outcome was the modified Rankin Scale, ranging from 0 (no symptoms) to 6 (death), at 90 days. We used random effect proportional odds regression modelling, to analyze the effect of differences in structure indicators (center volume and year of admission), process indicators (time to treatment and use of general anesthesia) and case-mix, by tracking changes in  $\tau^2$ , which represents the amount of between-center variation in outcome.

**Results:** Three thousand two hundred seventy-nine patients were included. Performance on structure and process indicators varied significantly between EVT-centers ( $P < 0.001$ ). Predicted probability of good functional outcome (modified Rankin Scale 0–2 at 90 days), which can be interpreted as an overall measure of a center's case-mix, varied significantly between 17 and 50% across centers. The amount of between-center variation ( $\tau^2$ ) was estimated at 0.040 in a model only accounting for random variation. This estimate more than doubled after adding case-mix variables ( $\tau^2$ : 0.086) to the model, while a small amount of between-center variation was explained by variation in performance on structure and process indicators ( $\tau^2$ : 0.081 and 0.089, respectively). This indicates that variation in case-mix affects the differences in outcome to a much larger extent.

**Conclusions:** Between-center variation in outcome of ischemic stroke patients mostly reflects differences in case-mix, rather than differences in structure or process of care. Since the latter two capture the real quality improvement potential, these should be used as indicators for comparing center performance. Especially when a strong association exists between those indicators and outcome, as is the case for time to treatment in ischemic stroke.

**Keywords:** Stroke, Endovascular treatment, Benchmarking, Quality of care, Outcome differences, Case-mix, Process of care

\* Correspondence: [m.amini@erasmusmc.nl](mailto:m.amini@erasmusmc.nl)

<sup>1</sup>Department of Public Health, Erasmus MC University Medical Center, P.O.

Box 2040, 3000, CA, Rotterdam, The Netherlands

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Endovascular treatment (EVT) has been shown to be a highly effective treatment for patients with ischemic stroke due to a proximal intracranial occlusion in the anterior circulation [1–3]. EVT is defined as arterial catheterization with a micro-catheter to the level of the occlusion, followed by mechanical thrombectomy or thrombus aspiration, or both, with or without delivery of a thrombolytic agent. Currently EVT is widely implemented in routine clinical practice, and the challenge is how to continuously improve the quality of this service.

Worldwide, healthcare systems and practices are being reorganized with a strong focus on measuring and improving outcomes of care. The quality of care for patients with a specific medical condition is judged by the achieved outcomes that are relevant for those patients. A central aspect of this development is benchmarking, comparing quality of care and specifically outcomes between healthcare providers, in this case EVT centers. If a specific center A has better outcomes for a certain condition than center B, this suggests that center B should copy the medical management strategy of center A in order to improve the quality of care in that center. An important problem of this approach, however, is the variability in baseline characteristics of patients ('case-mix') that often exists between centers. If center B treats more severely affected patients than center A, this might (partly) explain the better outcomes of center A. Moreover, especially if centers are relatively small, between-center differences in outcome may be caused by chance ('random variation'). Therefore, between-center comparisons of outcome should be adjusted for case-mix and random variation. If not done properly, such comparisons are likely to miss their purpose and could even be counterproductive as clinicians may base their decisions on flawed information. Furthermore, following Donabedian's framework for evaluating healthcare quality, the quality improvement potential is especially captured by variation in structures ('How is care organized?') and processes ('What is done?') of care [4]. For many diseases, however, it is unknown whether between-center variation in outcome reflects true differences in quality of care, captured by this framework.

Using data from a large nation-wide registry, the aim of this study was to assess the effect of structure and process indicators on between-center variation in outcome for ischemic stroke patient treated with EVT, while adjusting for case-mix and random variation.

## Methods

### Study design and patients

For this study we used data collected between March 2014 and November 2017 from the MR CLEAN Registry, a prospective, observational study in all 17 centers that perform EVT in the Netherlands (Supplementary Figure 1) [5]. This registry is unique since it includes clinical and neuro-

imaging data of all patients treated with EVT in one country during a multi-year period and thereby reflects clinical practice. All patients undergoing EVT for acute ischemic stroke have been registered. Inclusion criteria were: age 18 years and older, treatment in a center that participated in the MR CLEAN trial, and proximal intracranial vessel occlusion in the anterior circulation (internal carotid artery (ICA), internal carotid artery terminus (ICA-T), middle (M1/M2) cerebral artery, or anterior (A1/A2) cerebral artery), as shown by computed tomography angiography (CTA). Details on the study design and objectives of the MR CLEAN Registry have been described elsewhere [5]. Overall, data from 3279 patients were included for the current analysis (Supplementary Figure 2).

The MR CLEAN Registry was approved by the ethics committee of the Erasmus University MC, Rotterdam, The Netherlands (MEC-2014-235). With this approval it was approved by the research board of each participating center. At UMC Utrecht, approval to participate in the study has been obtained from their own research board and ethics committee.

### Case-mix indicators

For case-mix adjustment we used the following patient and neuro-imaging characteristics: age, sex, relevant medical history (i.e. previous stroke, atrial fibrillation, myocardial infarction, peripheral arterial disease, hypertension, diabetes mellitus, hypercholesterolemia), pre-stroke score on the modified Rankin Scale (mRS), the baseline score on the National Institute of Health Stroke Scale (NIHSS) as a stroke-related neurologic deficit score, and occlusion location and collateral grade on CT angiography. These characteristics were selected based on clinical knowledge and previous studies [6, 7]. Two additional characteristics were considered as case-mix indicators, because these are strongly associated with outcome but not influenceable by the centers: time between stroke onset and arrival at the emergency department (ED) of the intervention center, and whether or not the patient had been admitted to another center before being transferred to the intervention center [8, 9]. Stroke onset was defined as the time point when the sudden appearance of stroke symptoms was witnessed by the patient or an observer. In cases the time of first symptoms was unknown, onset was defined as the moment the patient was last seen well.

### Quality of care indicators

Quality of care indicators were defined using Donabedian's framework comprising indicators of structure, process, and outcome [4]. Both center volume and year of admission reflect the experience of a center with EVT and were used as structure indicators. Center volume was defined as the percentage of all EVT-patients treated in each center relative to all EVT-patients treated in the

Netherlands in the study period. In stroke care, high center volume was found to be associated with lower stroke-related mortality [10]. In other studies, higher volume stroke centers showed better outcomes on average [11–13]. Since EVT is a relatively new treatment in stroke care, we hypothesized (overall) performance to increase with calendar year and therefore added year of admission as an additional structure indicator.

Two process indicators were defined: time from arrival at the emergency department of the intervention center to groin puncture and the use of general anesthesia (yes/no). A significant negative association between ‘time-to-groin’ and outcome was found in previously published research, indicating that time delays before initiation of EVT have a negative effect on the likelihood of independent functional recovery at 90 days [8, 13–17]. So far, no differences in outcomes between general anesthesia and conscious sedation were observed in previous randomized controlled trials [18–20]. In several meta-analyses of observational studies, conscious sedation was associated with better outcomes than general anesthesia [21–23]. Although general anesthesia reduces the risk of patient agitation, unnecessary use of general anesthesia increases time delay in the total process of care, intra-procedural complications, and may result in cerebral hypoperfusion (e.g. through fluctuations in blood pressure on general anesthesia induction and abnormal cerebral auto-regulation) [24, 25]. Overall, we expected the use of general anesthesia to influence patient outcome, and therefore defined this as an additional process indicator.

We used the modified Rankin Scale (mRS) score as the outcome indicator [26]. The mRS score is a commonly used measure of patients’ functional outcome after ischemic stroke, and ranges from 0 (no symptoms) to 6 (death). The mRS score was assessed at 90 days after EVT ( $\pm$  14 days). Good functional outcome, defined as mRS 0–2, was used as secondary outcome (see below).

## Statistical analyses

### Descriptive analyses

We used Pearson’s chi-square statistic and the non-parametric Kruskal Wallis test for a univariable comparison of centers on case-mix and quality of care indicators. The predicted probability of good functional outcome can be considered an overall measure of each centers’ case-mix. To calculate this, we first fitted an individual-level logistic regression model including all case-mix indicators as predictors and yes/no good functional outcome (mRS 0–2 at 90 days) as the dependent variable. The predicted patient-level probabilities by this model were then used to calculate the median predicted probability per center.

### Random effect regression models

In order to adjust for random variation and assess the effects of adjusting for case-mix and performance on structure and

process indicators on between-center variation in outcome, we used random effect proportional odds regression modeling. A random center effect (intercept) accounts for the fact that the observed outcomes for lower-volume centers can take extreme values due to random variation. A proportional odds model exploits the full ordinal nature of the mRS as an outcome scale with more than two possible categories [27].

In all analyses, we used the inverse of the mRS score for each patient. Doing so allows us to interpret the estimates as the effects on the likelihood of a more favorable outcome, since a higher score on the inverse of the mRS means more favorable outcome (see above). We estimated common odds ratios with 95% confidence intervals on the patient level using four proportional odds regression models [27]. First, we fitted an ‘empty’, unadjusted model including only a random center effect, providing insight in between-center variation in outcome accounting only for random variation. In the second model, in addition to the random center effect, we adjusted for the individual-level fixed effects of case-mix indicators on outcome. In the third model, we added the fixed effects of the structure indicators to the model. Finally, the fourth model also contains the individual-level fixed effects of the process indicators.

### Between-center variation

To assess the relative impact of adjusting for case-mix and performance on structure and process indicators on between-center variation in outcome, we compared the variance of the random center effect ( $\tau^2$ ) across models. Essentially,  $\tau^2$  reflects the amount of between-center variation in outcome. In addition, separately for each model, we constructed forest plots to visualize this variation using estimates of center-specific outcome (i.e. the random center effects). The predictive power of the four models was compared using Akaike’s Information Criterion (AIC), in which a lower AIC value indicates a higher predictive power for outcome [28].

### Missing data

Multiple imputation was used to deal with missing data, which ranged between 0.7% (previous diabetes) to 6.3% (collateral grade) [5]. We fitted regression imputation models [29, 30] and imputed data five times, using the following variables: age, sex, medical history, pre-stroke mRS score, location of occlusion, collateral grade, baseline NIHSS score, whether patient transferred from other hospital, time intervals from onset to arrival at the ED, center volume per year, year of admission, time intervals from onset to groin puncture, and use of general anesthesia. Each imputed dataset was analyzed separately, after which the results were pooled.

All statistical analyses were performed with R statistical software version 3.4.3 (R Foundation for Statistical Computation, Vienna, Austria), using the *clmm* module

in the *ordinalimputation* package. Statistical significance was assessed at  $P < 0.05$  in all analyses.

## Results

### Descriptive analyses

At the center level, the median patient age ranged from 68 to 77 years, with statistically significant differences between centers (Table 1). Differences between centers were also statistically significant for previous stroke (range 0–26%), atrial fibrillation (13–37%), peripheral arterial disease (4–22%), hypertension (41–67%), hypercholesterolemia (15–50%), pre-stroke mRS, location of occlusion, collateral grade, baseline NIHSS score (range median score per center 13–17) and percentage of transferred patients from another hospital (0–77%). Median time of stroke onset to arrival at the ED of the intervention center ranged from 52 to 160 min across centers ( $P < 0.001$ ). Importantly, the median predicted probability of good functional outcome (mRS 0–2 at 90 days), which can be interpreted as an overall measure of a center's case-mix, varied between 17 and 50% across centers ( $P = 0.004$ ) (Table 1 and Supplementary Table 1). The patient-level effect estimates of case-mix variables on outcome shown in Supplementary Figure 3 are comparable to the results of prior research [31].

Relative to the total number of patients in our data, the number of EVT-patients varied significantly between centers across the 4 years (Fig. 1). Median time from arrival at the ED of the intervention center to groin puncture also varied substantially: between 74 and 125 min for non-transferred patients ( $P < 0.001$ ) and between 20 and 60 min for transferred patients ( $P < 0.001$ ). Variation in the use of general anesthesia (0–99%) was also statistically significant. Crude differences in outcome were statistically significant ( $P < 0.001$ ) across centers for mRS values 0–6: no symptoms (1–18%), no significant disability (7–26%), slight disability (4–29%), moderate disability (4–20%), moderately severe disability (7–31%), severe disability (2–11%), and death (21–36%) (Table 2 and Supplementary Table 2).

### Predictive power for outcome

Model 1 generated an AIC of 11,158, which dropped to 10,050 after adding case-mix variables in model 2, suggesting a considerably improved predictive power (Table 3). Adding structure (AIC = 10,043) and process indicators (AIC = 10,012) only slightly improved the predicted power further (Table 3).

### Between-center variation in outcome

In model 1, which only adjusts for random variation, the amount of the between-center variation in outcome ( $\tau^2$ ) was 0.040 (Table 3). The  $\tau^2$  represents the amount of variability in outcome between centers. This estimate more than doubled after adding case-mix indicators (model 2,  $\tau^2$ : 0.086), while adding structure indicators (model 3)

and process indicators (model 4) left it almost unaffected ( $\tau^2$ : 0.081 and 0.089, respectively). This indicates that only a small amount of between-center variation was explained by variation in performance on structure and process indicators. This finding is also reflected in the forest plots (Fig. 2), which for each model show the estimated effects of all 17 centers on the likelihood of favorable outcome. Between-center variation increased particularly after case-mix indicators were added (compare Fig. 2b with Fig. 2a), while variation remained rather constant after adding structure (Fig. 2c) and process (Fig. 2d) indicators.

## Discussion

This study focused on assessing variation in outcome between centers treating ischemic stroke with EVT in the Netherlands, and the impact of differences in case-mix and in performance on structure and process indicators. Our results show that differences in case-mix have a much larger impact on between-center differences in outcome for stroke patients treated with EVT than differences in structure and processes of care across centers. Therefore, an (unadjusted) measure of functional outcome may not be a valid indicator for quality of care when comparing stroke centers.

### Outcome indicators as measures of quality of care

Benchmarking initiatives increasingly use outcome measures to compare quality of care between centers. However, as found in this study, significant differences in performance on evidence- and consensus-based measures of (processes of) care may not be reflected in between-center differences in outcome. An important explanation for this is the observational nature of the data that are typically used in such benchmarking initiatives. When assessing the effect of interventions or processes of care on outcome in such data, confounding by indication is a major issue. For example, physicians might treat more severely affected patients faster than less severely affected patients [32, 33]. As disease severity is likely to influence outcome, in observational data it can confound the estimated relation with outcome insofar severity is insufficiently accounted for. Although we adjusted for disease severity in several ways using measured case-mix variables, we cannot preclude the possibility that our estimated effect of time to treatment on outcome is to some extent biased. This could partly be because of a residual unmeasured case-mix effect, as well as unaccounted for technical parameters. Knowing the association of center-level outcomes with process indicators and structural characteristics allows decision-makers to understand the determinants of performance and implement improvement strategies; in other words, all indicators should be retained to get an empirical (i.e., context specific) evaluation of the quality of care. If a center is an “outlying provider” even after accounting for patient case mix and

**Table 1** Case-mix characteristics of patients treated in intervention centers in MR CLEAN Registry

	N (%) / Median (IQR) in total population (missing excluded)	Center-level range (median or percentage)	Total number of patients	Missing N (%)	P-value*
Age (years)	72 (61–80)	68–77	3279	0	0.001
Men	1696 (52)	39–55	3279	0	0.790
Medical History					
Previous Stroke	546 (17)	0–26	3252	27 (0.8)	< 0.001
Atrial Fibrillation	772 (24)	13–37	3236	43 (1.3)	< 0.001
Myocardial Infarction	453 (14)	4–18	3212	67 (2.0)	0.16
Peripheral Arterial Disease	301 (9)	4–22	3211	68 (2.1)	< 0.001
Hypertension	1688 (53)	41–67	3213	66 (2.0)	< 0.001
Diabetes	532 (16)	12–25	3255	24 (0.7)	0.09
Hypercholesterolemia	967 (31)	15–50	3136	143 (4.4)	< 0.001
Pre-stroke modified Rankin Scale score			3207	72 (2.2)	< 0.001
0	2170 (68)	45–87			
1	424 (13)	3–19			
2	241 (8)	0–16			
≥ 3	372 (12)	5–25			
Location of occlusion			3119	160 (4.99)	< 0.001
M1	1815 (58)	45–70			
M2	455 (15)	4–31			
Intracranial ICA	161 (5)	0–10			
ICA-T	663 (21)	9–27			
Other (M3/anterior)	25 (1)	0–4			
Collateral grade			3072	207 (6.3)	< 0.001
0 (Absent)	187 (6)	0–9			
1 (< 50%)	1100 (36)	17–46			
2 (> 50% but < 100%)	1190 (39)	33–61			
3 (100%)	595 (19)	10–37			
Baseline NIHSS score	16 (11–20)	13–17	3224	55 (1.7)	< 0.001
Transferred patients from other hospital	1783 (54)	0–77	3279	0	< 0.001
Time from onset to arrival at the ED (min)	135 (65–195)	52–160	3155	124 (3.8)	< 0.001
Predicted probability of good functional outcome <sup>a</sup>	41 (20–62)	17–50	2550	729 (22.2)	0.004

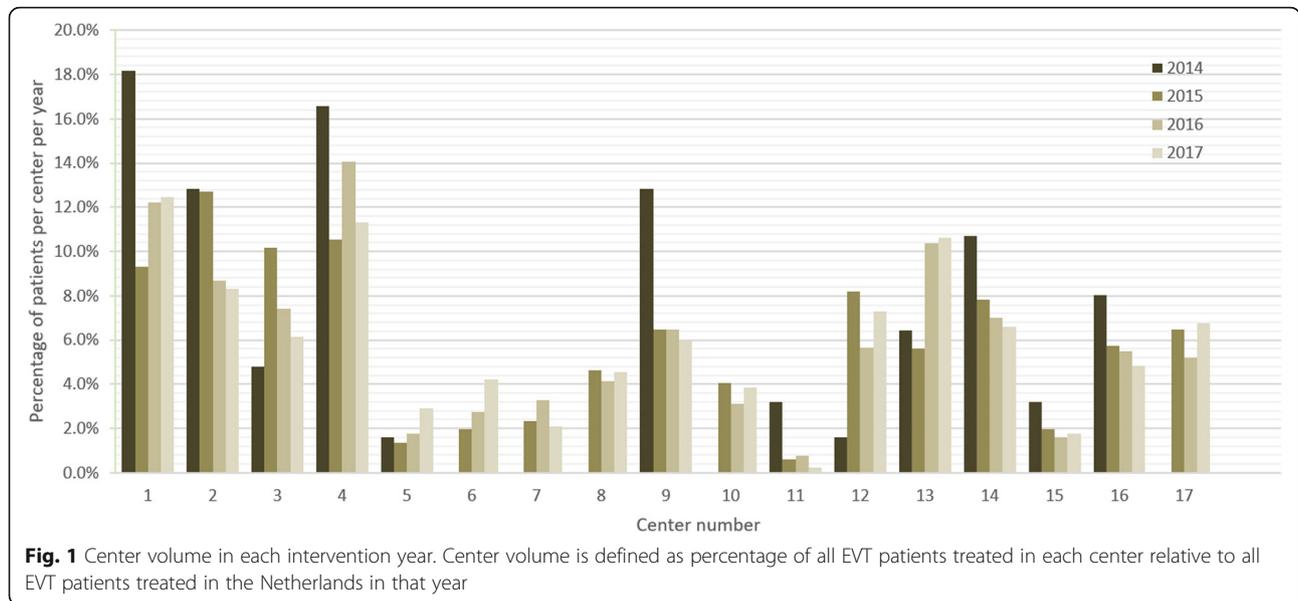
IQR Interquartile range, ICA-T internal carotid artery terminus, M1/M2 middle cerebral artery, NIHSS National Institutes of Health Stroke Scale, ED Emergency department

<sup>a</sup>Predicted probability (%) of good functional outcome (modified Rankin Scale score 0–2 at 90 days) is based on an individual-level logistic regression model predicting good functional outcome from the case-mix variables included in this table. See Supplementary Table 1 for details of case-mix characteristics of patients treated in each intervention center

\*P-value is based on comparison between 17 centers using a non-parametric Kruskal Wallis test for continuous variables or Pearson's chi-square statistic for categorical variables

process/structural characteristics, in-depth analyses and audit activities should be put in place to understand the reasons for this center's outlying performance. Given that detecting a beneficial effect of effective interventions on outcome can be difficult even in large methodologically sound randomized trials, observing effects of evidence- and consensus-based indicators of (processes of) care on between-center differences in outcome may be even harder in benchmarking initiatives using observational registry

data. In other words, when observational data are used for benchmarking, which in practice is typically the case, good performance on structure and process indicators may not be reflected in favorable outcome due to unmeasured confounding factors. Therefore, identifying and benchmarking performance on indicators with a proven contribution to favorable outcomes should be an important future direction for stroke care quality assessment and improvement initiatives.



**Table 2** Quality of care indicators of all 17 intervention centers in the Netherlands

	N (%) / Median (IQR) in total population (missing excluded)	Center-level range (median or percentage)	Total number of patients	Missing N (%)	P-value*
Structure					
Center volume			3279	0	
2014	187 (6 <sup>a</sup> )	0–18			< 0.001
2015	822 (25)	1–13			< 0.001
2016	1131 (34)	1–14			< 0.001
2017	1139 (35)	0.3–13			< 0.001
Processes					
Time from arrival at the ED to groin puncture in non-transferred patients (min)	90 (70–119)	74–125	1292	203 (6.2)	< 0.001
Time from arrival at the ED to groin puncture in transferred patients (min)	39 (26–57)	20–60	1722	61 (1.9)	< 0.001
Use of general anesthesia	778 (25)	0–99	3082	197 (6.0)	< 0.001
Outcome					
Modified Rankin Scale (mRS) score at 90 days			3065	214 (6.5)	< 0.001
mRS 0 (No symptoms)	209 (7)	1–18			
mRS 1 (No significant disability)	471 (15)	7–26			
mRS 2 (Slight disability)	561 (18)	4–29			
mRS 3 (Moderate disability)	404 (13)	4–20			
mRS 4 (Moderately severe disability)	366 (12)	7–31			
mRS 5 (Severe disability)	168 (6)	2–11			
mRS 6 (Dead)	886 (29)	21–36			

IQR Interquartile range, ED Emergency department, EVT endovascular treatment

<sup>a</sup>Percentage of all EVT patients treated in the Netherlands in that year (see Fig. 1 for details)

\*P-value is based on comparison between 17 centers using a non-parametric Kruskal Wallis test for continuous variables or Pearson's chi-square statistic for categorical variables

See Supplementary Table 1 for details of quality of care indicators in each intervention center

**Table 3** Results from the random effect proportional odds regression analysis using the inverse of the modified Rankin Scale at 90 days as the dependent variable

	Model 1	Model 2	Model 3	Model 4
Center volume [OR (95% CI)]	–	–	1.01 (0.98–1.04)	1.01 (0.98–1.05)
Year of admission [OR (95% CI)]	–	–		
2014			ref	ref
2015			1.32 (0.95–1.85)	1.20 (0.86–1.67)
2016			1.67 (1.22–2.28)	1.36 (0.98–1.90)
2017			1.60 (1.17–2.19)	1.27 (0.91–1.77)
Time from arrival at the ED of intervention center to groin puncture (every 30 min) [OR (95% CI)]	–	–	–	0.87 (0.82–0.92)
Use of general anesthesia [OR (95% CI)]	–	–	–	0.72 (0.57–0.90)
Variance of random center intercept [ $\tau^2$ (95% CI)] <sup>#</sup>	0.040 (0.012–0.113)	0.086 (0.042–0.261)	0.081 (0.028–0.227)	0.089 (0.033–0.254)
AIC	11,158	10,050	10,043	10,012

ED Emergency department, AIC Akaike’s Information Criterion. A lower AIC value indicates a better model fit

Model 1 (‘unadjusted model’) includes a random center intercept only

Model 2 (‘case-mix adjusted model’) includes a random center intercept and the case-mix indicators (Table 1). See Supplementary Figure 3 for the estimated fixed effects of each case-mix indicator

Model 3 (‘case-mix and structure indicators adjusted model’) includes a random center intercept, case-mix indicators, and structure indicators

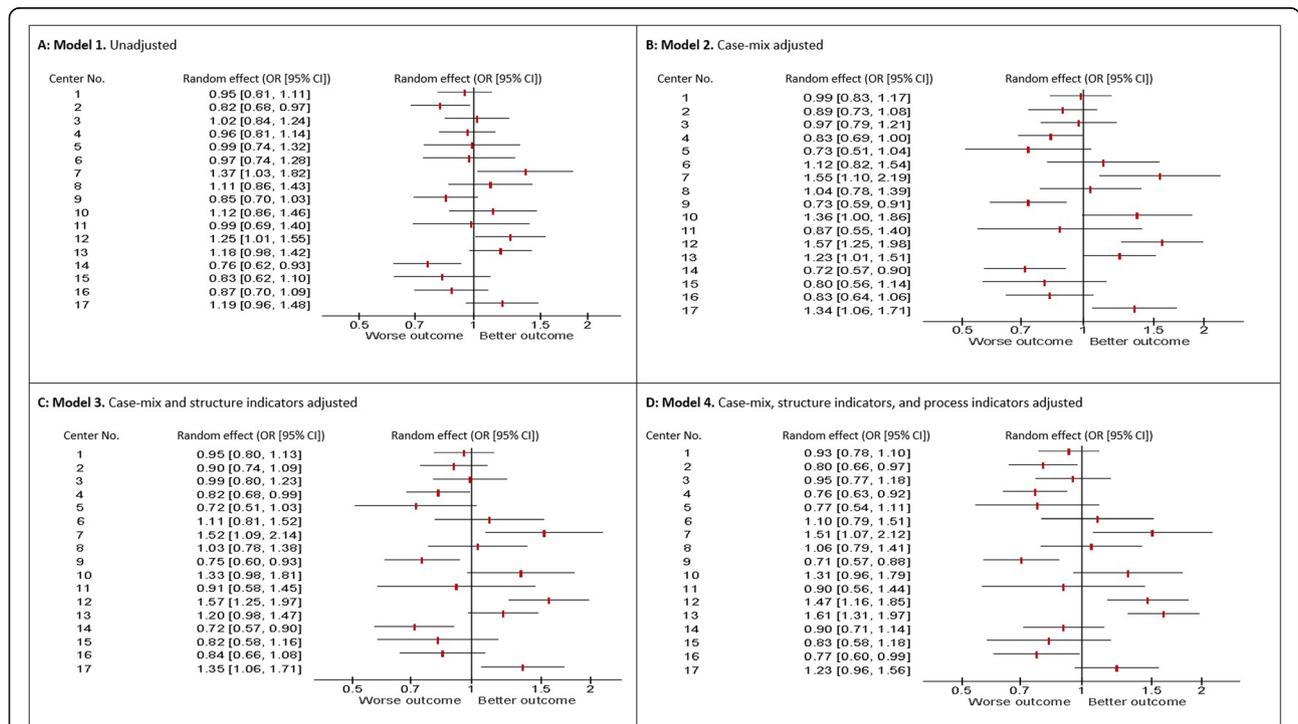
Model 4 (‘case-mix, structure and process indicators adjusted model’) includes a random center intercept, case-mix indicators, structure indicators, and process indicators

<sup>#</sup>95% CIs around  $\tau^2$  were estimated using single imputed data with bootstrap

**Between-center differences in outcome after case-mix adjustment**

The observed difference of the unadjusted and adjusted estimated center effect is because of statistical consideration and is a result of both imbalance and stratification [34, 35].

Good outcomes will generally be more difficult to achieve for patients who are more severely affected at baseline. Therefore, the observed outcomes of centers with a relatively ‘severe’ case-mix will on average be less favorable (i.e. biased downwards) as compared to those of centers with a



**Fig. 2** Forest plots reporting random center effect (odds ratios and 95% confidence intervals) on inverse of modified Rankin Scale at 90 days in four models using random effect proportional odds regression analysis. **a:** Model 1 (unadjusted model); **b:** Model 2 (case-mix adjusted model); **c:** Model 3 (case-mix and structure indicators adjusted model); **d:** Model 4 (case-mix, structure and process indicators adjusted model)

relatively ‘mild’ case-mix (i.e. biased upwards). In general, adjusting for the imbalance in case-mix would then reduce observed variation in outcome between centers. However, in this study an opposite pattern is observed. After adjusting for case-mix, actual between-center differences in outcome become visible. Apparently, centers with a more severe case-mix tend to have relatively good outcomes and vice versa. Although counterintuitive, this still underlines the necessity of appropriate case-mix adjustment in benchmarking quality performance using observational data. Besides the effect of adjusting for the imbalance, more extreme center effect after adjustment could be a result of the stratification effect. Although “stratification” usually refers to conditioning on categorical subgroups (e.g. on sex), we also use this term when continuous variables are involved, for example, age. Adjustment will generally increase standard errors, and the stratification (adjustment) effect will lead to more extreme effect estimates [34].

### Strengths and limitations

In previous studies [6, 7], a dichotomized version of the mRS was used (mRS  $\geq 3$  was considered as ‘poor outcome’) and analyzed using binary logistic regression models. In order to exploit the full ordinal nature of the mRS score, we used proportional odds regression analysis. In addition, the use of random effect analyses allowed us to estimate (the variance of) center outcomes adjusted for random variation and various other factors (i.e. case-mix, structure, and process indicators).

A first limitation of this study is the unavailability of other potentially contributing factors to between-center outcome differences, e.g. unmeasured patient characteristics, care processes, and center characteristics. A second limitation is that missing values may have introduced some bias, although we believe to have mitigated this issue considerably using multiple imputation, which is the preferred method over complete case analysis [36, 37]. A third limitation is low number of second-level units, with resulting lack of precision in  $\tau_{\text{au}}^2$  estimates. A final limitation is that we only analyzed one outcome. Although the mRS is used as an outcome in virtually all modern stroke trials, our conclusions might have been different if we had used other outcomes that are relevant to stroke patients treated with EVT, like patient-reported outcomes such as quality of life. In addition, even though the mRS is an appropriate tool for assessing patient disability after stroke care, it may not be easily transferrable to outcomes research. After all, in many countries other than the Netherlands, the mRS is not routinely collected for all acute stroke patients. However, if we would have used clinical outcomes that are commonly registered in administrative databases, such as short-term mortality or readmission, we would have less certainty about our estimations because of the much smaller number of events in the context of acute stroke

treatment. Moreover, from a clinical perspective these outcome measures are far less relevant in the stroke context compared to the mRS. Therefore, we used the mRS as the most powerful and clinically relevant outcome for our study.

### Conclusions

In this study, we have demonstrated that between-center differences in performance on structure and process indicators have a small impact on functional outcome of ischemic stroke patients treated with EVT, while differences in case-mix affects this variation substantially. Thus, outcome indicators may not be valid and useful for comparing and improving the quality of stroke care based on observational data. Since variation in performance on structure and process indicators captures real quality improvement potential, these indicators should be used in future benchmarking initiatives. This is especially true when a strong association exists between those indicators and outcome, as is the case for time to treatment in ischemic stroke.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12913-020-05841-y>.

**Additional file 1: Table S1.** Case-mix characteristics of patients treated in intervention centers in the MR CLEAN Registry. **Table S2.** Quality of care indicators characteristics of the intervention centers in the MR CLEAN Registry database. **Figure S1.** Specialized EVT centers in the Netherlands. Adapted from MR CLEAN Registry study (<https://www.mrclean-trial.org/mr-clean-registry/participating-centers.html>). **Figure S2.** Flowchart of patient selection in the MR CLEAN Registry. **Figure S3** Forest plot reporting odds ratios with 95% confidence intervals of the fixed effects of the case-mix variables on the inverse of the modified Rankin Scale at 90 days using random effect proportional odds regression analysis in the case-mix adjusted model.

### Abbreviations

EVT: Endovascular treatment; ICA: Internal carotid artery; ICA-T: Internal carotid artery terminus; M1/M2: Middle cerebral artery; A1/A2: Anterior cerebral artery; CTA: Computed tomography angiography; mRS: Modified Rankin Scale; NIHSS: National Institute of Health Stroke Scale; ED: Emergency department; AIC: Akaike's Information Criterion; IQR: Interquartile range; PAD: Peripheral arterial disease; MI: Myocardial infarction; HCL: Hypercholesterolemia

### Acknowledgements

We thank the MR CLEAN Registry Investigators-group authors: Executive committee.

Diederik W.J. Dippel<sup>1</sup>, Aad van der Lugt<sup>2</sup>, Charles B.L.M. Majoie<sup>3</sup>, Yvo B.W.E.M. Roos<sup>4</sup>, Robert J. van Oostenbrugge<sup>5</sup>, Wim H. van Zwam<sup>6</sup>, Jelis Boiten<sup>14</sup>, Jan Albert Vos<sup>8</sup>. Study coordinators.

Josje Brouwer<sup>4</sup>, Sanne J. den Hartog<sup>1,2,40</sup>, Wouter H. Hinsenveld<sup>5,6</sup>, Manon Kappelhof<sup>3</sup>, Kars C.J. Compagne<sup>2</sup>, Robert-Jan B. Goldhoorn<sup>5,6</sup>, Maxim J.H.L. Mulder<sup>1,2</sup>, Ivo G.H. Jansen<sup>3</sup>.

Local principal investigators.

Diederik W.J. Dippel<sup>1</sup>, Bob Roozenbeek<sup>1</sup>, Aad van der Lugt<sup>2</sup>, Adriaan C.G.M. van Es<sup>2</sup>, Charles B.L.M. Majoie<sup>3</sup>, Yvo B.W.E.M. Roos<sup>4</sup>, Bart J. Emmer<sup>3</sup>, Jonathan M. Coutinho<sup>4</sup>, Wouter J. Schonewille<sup>7</sup>, Jan Albert Vos<sup>8</sup>, Marieke J.H. Wermer<sup>9</sup>, Marianne A.A. van Walderveen<sup>10</sup>, Julie Staals<sup>5</sup>, Robert J. van Oostenbrugge<sup>5</sup>, Wim H. van Zwam<sup>6</sup>, Jeannette Hofmeijer<sup>11</sup>, Jasper M. Martens<sup>12</sup>, Geert J. Lycklama à Nijeholt<sup>13</sup>, Jelis Boiten<sup>14</sup>, Sebastiaan F. de

Bruijn<sup>15</sup>; Lukas C. van Dijk<sup>16</sup>; H. Bart van der Worp<sup>17</sup>; Rob H. Lo<sup>18</sup>; Ewoud J. van Dijk<sup>19</sup>; Hieronymus D. Boogaarts<sup>20</sup>; J. de Vries<sup>22</sup>; Paul L.M. de Kort<sup>21</sup>; Julia van Tuij<sup>21</sup>; Jo Jo P. Peluso<sup>26</sup>; Puck Franssen<sup>22</sup>; Jan S.P. van den Berg<sup>22</sup>; Boudewijn A.A.M. van Hasselt<sup>23</sup>; Leo A.M. Aerden<sup>24</sup>; René J. Dallinga<sup>25</sup>; Maarten Uyttenboogaart<sup>26</sup>; Omid Eschgi<sup>29</sup>; Reinoud P.H. Bokkers<sup>29</sup>; Tobien H.C.M.L. Schreuder<sup>30</sup>; Roel J.J. Heijboer<sup>31</sup>; Koos Keizer<sup>32</sup>; Lonneke S.F. Yo<sup>33</sup>; Heleen M. den Hertog<sup>22</sup>; Emiel J.C. Sturm<sup>35</sup>; Paul Brouwers<sup>34</sup>.

#### Imaging assessment committee.

Charles B.L.M. Majoie<sup>3</sup> (chair); Wim H. van Zwam<sup>6</sup>; Aad van der Lugt<sup>2</sup>; Geert J. Lycklama à Nijeholt<sup>13</sup>; Marianne A.A. van Walderveen<sup>10</sup>; Marieke E.S. Sprengers<sup>3</sup>; Sjoerd F.M. Jenniskens<sup>27</sup>; René van den Berg<sup>3</sup>; Albert J. Yoo<sup>38</sup>; Ludo F.M. Beenen<sup>3</sup>; Alida A. Postma<sup>6</sup>; Stefan D. Roosendaal<sup>3</sup>; Bas F.W. van der Kallen<sup>13</sup>; Ido R. van den Wijngaard<sup>13</sup>; Adriaan C.G.M. van Es<sup>2</sup>; Bart J. Emmer<sup>3</sup>; Jasper M. Martens<sup>12</sup>; Lonneke S.F. Yo<sup>33</sup>; Jan Albert Vos<sup>8</sup>; Joost Bot<sup>36</sup>; Pieter-Jan van Doormaal<sup>2</sup>; Anton Meijer<sup>27</sup>; Elyas Ghariq<sup>13</sup>; Reinoud P.H. Bokkers<sup>29</sup>; Marc P. van Proosdij<sup>37</sup>; G. Menno Krietemeijer<sup>33</sup>; Jo P. Peluso<sup>26</sup>; Hieronymus D. Boogaarts<sup>20</sup>; Rob Lo<sup>18</sup>; Dick Gerrits<sup>35</sup>; Wouter Dinkelaar<sup>2</sup>; Auke P.A. Appelman<sup>29</sup>; Bas Hammer<sup>16</sup>; Sjoert Pegge<sup>27</sup>; Anouk van der Hoorn<sup>29</sup>; Saman Vinke<sup>20</sup>.

#### Writing committee.

Diederik W.J. Dippel<sup>1</sup> (chair); Aad van der Lugt<sup>2</sup>; Charles B.L.M. Majoie<sup>3</sup>; Yvo B.W.E.M. Roos<sup>4</sup>; Robert J. van Oostenbrugge<sup>5</sup>; Wim H. van Zwam<sup>6</sup>; Geert J. Lycklama à Nijeholt<sup>13</sup>; Jelis Boiten<sup>14</sup>; Jan Albert Vos<sup>8</sup>; Wouter J. Schonewille<sup>2</sup>; Jeannette Hofmeijer<sup>11</sup>; Jasper M. Martens<sup>12</sup>; H. Bart van der Worp<sup>17</sup>; Rob H. Lo<sup>18</sup>.

#### Adverse event committee.

Robert J. van Oostenbrugge<sup>5</sup> (chair); Jeannette Hofmeijer<sup>11</sup>; H. Zwenneke Flach<sup>23</sup>.  
Trial methodologist.  
Hester F. Lingsma<sup>40</sup>.

#### Research nurses / local trial coordinators.

Naziha el Ghannouti<sup>1</sup>; Martin Sterrenberg<sup>1</sup>; Corina Puppels<sup>7</sup>; Wilma Pellikaan<sup>7</sup>; Rita Sprengers<sup>4</sup>; Marjan Elfrink<sup>11</sup>; Michelle Simons<sup>11</sup>; Marjolein Vossers<sup>12</sup>; Joke de Meris<sup>14</sup>; Tamara Vermeulen<sup>14</sup>; Annet Geerlings<sup>19</sup>; Gina van Vemde<sup>22</sup>; Tiny Simons<sup>30</sup>; Cathelijan van Rijswijk<sup>21</sup>; Gert Messchendorp<sup>28</sup>; Nynke Nicolaij<sup>28</sup>; Hester Bongenaar<sup>32</sup>; Karin Bodde<sup>24</sup>; Sandra Kleijn<sup>34</sup>; Jasmijn Lodico<sup>34</sup>; Hanneke Droste<sup>34</sup>; Maureen Wollaert<sup>2</sup>; Sabrina Verheesen<sup>5</sup>; D. Jeurissen<sup>5</sup>; Erna Bos<sup>9</sup>; Yvonne Drabbe<sup>15</sup>; Michelle Sandiman<sup>15</sup>; Marjan Elfrink<sup>11</sup>; Nicoline Aaldering<sup>11</sup>; Berber Zweedijk<sup>17</sup>; Mostafa Khalilzadeh<sup>15</sup>; Jocova Vervoort<sup>21</sup>; Hanneke Droste<sup>34</sup>; Nynke Nicolaij<sup>2</sup>; Michelle Simons<sup>11</sup>; Eva Ponjee<sup>22</sup>; Sharon Romviel<sup>19</sup>; Karin Kanselaar<sup>19</sup>; Erna Bos<sup>9</sup>; Denn Barning<sup>10</sup>.

#### PhD / Medical students.

Esmee Venema<sup>40</sup>; Vicky Chalos<sup>1,40</sup>; Ralph R. Geuskens<sup>3</sup>; Tim van Straaten<sup>19</sup>; Saliha Ergezen<sup>1</sup>; Roger R.M. Harmsma<sup>1</sup>; Daan Muijers<sup>1</sup>; Anouk de Jong<sup>1</sup>; Olvert A. Berkhemer<sup>1,3,6</sup>; Anna M.M. Boers<sup>3,39</sup>; J. Huguet<sup>3</sup>; P.F.C. Groot<sup>3</sup>; Marieke A. Mens<sup>3</sup>; Katinka R. van Kranendonk<sup>3</sup>; Kilian M. Treurniet<sup>3</sup>; Ivo G.H. Jansen<sup>3</sup>; Manon L. Tolhuisen<sup>3,39</sup>; Heitor Alves<sup>3</sup>; Annick J. Weterings<sup>3</sup>; Eleonora L.F. Kirkels<sup>3</sup>; Eva J.H.F. Voogd<sup>11</sup>; Lieve M. Schupp<sup>3</sup>; Sabine Collette<sup>28,29</sup>; Adrien E.D. Groot<sup>4</sup>; Natalie E. LeCouffe<sup>4</sup>; Praneeta R. Konduri<sup>39</sup>; Haryadi Prasetya<sup>39</sup>; Nerea Arrarte-Terreros<sup>39</sup>; Lucas A. Ramos<sup>39</sup>.

#### List of affiliations.

Department of Neurology<sup>1</sup>, Radiology<sup>2</sup>, Public Health<sup>40</sup>, Erasmus MC University Medical Center;  
Department of Radiology and Nuclear Medicine<sup>3</sup>, Neurology<sup>4</sup>, Biomedical Engineering & Physics<sup>39</sup>, Amsterdam UMC, University of Amsterdam, Amsterdam;  
Department of Neurology<sup>5</sup>, Radiology<sup>6</sup>, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM);  
Department of Neurology<sup>7</sup>, Radiology<sup>8</sup>, Sint Antonius Hospital, Nieuwegein;  
Department of Neurology<sup>9</sup>, Radiology<sup>10</sup>, Leiden University Medical Center;  
Department of Neurology<sup>11</sup>, Radiology<sup>12</sup>, Rijnstate Hospital, Arnhem;  
Department of Radiology<sup>13</sup>, Neurology<sup>14</sup>, Haaglanden MC, the Hague;  
Department of Neurology<sup>15</sup>, Radiology<sup>16</sup>, Haga Hospital, the Hague;  
Department of Neurology<sup>17</sup>, Radiology<sup>18</sup>, University Medical Center Utrecht;  
Department of Neurology<sup>19</sup>, Neurosurgery<sup>20</sup>, Radiology<sup>27</sup>, Radboud University Medical Center, Nijmegen;  
Department of Neurology<sup>21</sup>, Radiology<sup>26</sup>, Elisabeth-TweeSteden ziekenhuis, Tilburg;  
Department of Neurology<sup>22</sup>, Radiology<sup>23</sup>, Isala Klinieken, Zwolle;  
Department of Neurology<sup>24</sup>, Radiology<sup>25</sup>, Reinier de Graaf Gasthuis, Delft;  
Department of Neurology<sup>28</sup>, Radiology<sup>29</sup>, University Medical Center Groningen;  
Department of Neurology<sup>30</sup>, Radiology<sup>31</sup>, Atrium Medical Center, Heerlen;  
Department of Neurology<sup>32</sup>, Radiology<sup>33</sup>, Catharina Hospital, Eindhoven;  
Department of Neurology<sup>34</sup>, Radiology<sup>35</sup>, Medical Spectrum Twente, Enschede;

Department of Radiology<sup>36</sup>, Amsterdam UMC, Vrije Universiteit van Amsterdam, Amsterdam; Department of Radiology<sup>37</sup>, Noordwest Ziekenhuisgroep, Alkmaar;  
Department of Radiology<sup>38</sup>, Texas Stroke Institute, Texas, United States of America.

#### Authors' contributions

HFL, BR and WJD designed the study. MA undertook analyses and interpretation of study findings and wrote the first draft of the manuscript. NvL, HFL, FE, BR, WJD contributed to the interpretation of study findings, reviews, and revision of the manuscript. MJHLM, WS, GLN, WHH, RJBG, PJD, SJ, JH contributed to the findings review and revision of the manuscript. All the authors critically reviewed the various versions of the full paper and approved the final manuscript.

#### Funding

No funding was received in support of this study.

#### Availability of data and materials

The dataset analyzed during the current study is available from the MR CLEAN Registry upon request as a study proposal. Requests should be directed to the executive committee ([mrcclean@erasmusmc.nl](mailto:mrcclean@erasmusmc.nl)). Detailed analytic methods and study materials, including log files of statistical analyses, will be made available to other researchers on request to the corresponding author ([m.amini@erasmusmc.nl](mailto:m.amini@erasmusmc.nl)).

#### Ethics approval and consent to participate

Not required.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Public Health, Erasmus MC University Medical Center, P.O. Box 2040, 3000, CA, Rotterdam, The Netherlands. <sup>2</sup>Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands. <sup>3</sup>Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. <sup>4</sup>Department of Neurology, St. Antonius Hospital, Nieuwegein, The Netherlands. <sup>5</sup>Department of Radiology, Medical Center Haaglanden, The Hague, The Netherlands. <sup>6</sup>Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), Maastricht, The Netherlands. <sup>7</sup>Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands. <sup>8</sup>Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands.

Received: 9 June 2020 Accepted: 20 October 2020

Published online: 31 October 2020

#### References

- Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. 2016;387(10029):1723–31.
- Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med*. 2018;378(8):708–18.
- Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2018;378(1):11–21.
- Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988; 260(12):1743–8.
- Jansen IGH, Mulder M, Goldhoorn RB, Investigators MCR. Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (MR CLEAN Registry). *BMJ*. 2018;360:k949.
- Lingsma HF, Dippel DW, Hoeks SE, Steyerberg EW, Franke CL, van Oostenbrugge RJ, et al. Variation between hospitals in patient outcome after

- stroke is only partly explained by differences in quality of care: results from the Netherlands stroke survey. *J Neurol Neurosurg Psychiatry*. 2008;79(8):888–94.
7. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte OP, Reimer WJ, Van Houwelingen HC, et al. Comparing and ranking hospitals based on outcome: results from the Netherlands stroke survey. *QJM*. 2010;103(2):99–108.
  8. Mulder M, Jansen IGH, Goldhoorn RB, Venema E, Chalos V, Compagne KCJ, et al. Time to endovascular treatment and outcome in acute ischemic stroke: MR CLEAN registry results. *Circulation*. 2018;138(3):232–40.
  9. Venema E, Groot AE, Lingsma HF, Hinsenveld W, Treurniet KM, Chalos V, et al. Effect of Interhospital transfer on endovascular treatment for acute ischemic stroke. *Stroke*. 2019;50(4):923–30.
  10. Saposnik G, Baibergenova A, O'Donnell M, Hill MD, Kapral MK, Hachinski V. Hospital volume and stroke outcome. Does it matter? *Neurology*. 2007; 69(11):1142–51.
  11. Gupta R, Horev A, Nguyen T, Gandhi D, Wisco D, Glenn BA, et al. Higher volume endovascular stroke centers have faster times to treatment, higher reperfusion rates and higher rates of good clinical outcomes. *J Neurointerv Surg*. 2013;5(4):294–7.
  12. Tsugawa Y, Kumamaru H, Yasunaga H, Hashimoto H, Horiguchi H, Ayanian JZ. The association of hospital volume with mortality and costs of care for stroke in Japan. *Med Care*. 2013;51(9):782–8.
  13. Lorenzo R, Waleed B, Alejandro AR. Transfer to high-volume centers associated with reduced mortality after endovascular treatment of acute stroke. *Stroke*. 2017;48(5):1316–21.
  14. Aghaebrahim A, Granja MF, Agnoletto G, Aguilar-Salinas P, Cortez GM, Santos R, et al. Workflow optimization for ischemic stroke in a community-based stroke center. *World Neurosurg*. 2019;129:e273–8.
  15. Franssen PS, Berkhemer OA, Lingsma HF, Beumer D, van den Berg LA, Yoo AJ, et al. Time to reperfusion and treatment effect for acute ischemic stroke: a randomized clinical trial. *JAMA Neurol*. 2016;73(2):190–6.
  16. Goyal M, Almekhlafi MA, Fan L, Menon BK, Demchuk AM, Yeatts SD, et al. Evaluation of interval times from onset to reperfusion in patients undergoing endovascular therapy in the interventional Management of Stroke III trial. *Circulation*. 2014;130(3):265–72.
  17. Saver JL, Goyal M, van der Lugt A, Menon BK, Majorie CB, Dippel DW, et al. Time to treatment with endovascular Thrombectomy and outcomes from ischemic stroke: a meta-analysis. *JAMA*. 2016;316(12):1279–88.
  18. Lowhagen Henden P, Rentzos A, Karlsson JE, Rosengren L, Leiram B, Sundeman H, et al. General anesthesia versus conscious sedation for endovascular treatment of acute ischemic stroke: the AnStroke trial (anesthesia during stroke). *Stroke*. 2017;48(6):1601–7.
  19. Schonenberger S, Uhlmann L, Hacke W, Schieber S, Mundiyanapurath S, Purucker JC, et al. Effect of conscious sedation vs general anesthesia on early neurological improvement among patients with ischemic stroke undergoing endovascular Thrombectomy: a randomized clinical trial. *JAMA*. 2016;316(19):1986–96.
  20. Simonsen CZ, Yoo AJ, Sorensen LH, Juul N, Johnsen SP, Andersen G, et al. Effect of general anesthesia and conscious sedation during endovascular therapy on infarct growth and clinical outcomes in acute ischemic stroke: a randomized clinical trial. *JAMA Neurol*. 2018;75(4):470–7.
  21. Brinjikji W, Murad MH, Rabinstein AA, Cloft HJ, Lanzino G, Kallmes DF. Conscious sedation versus general anesthesia during endovascular acute ischemic stroke treatment: a systematic review and meta-analysis. *AJNR Am J Neuroradiol*. 2015;36(3):525–9.
  22. Wan TF, Xu R, Zhao ZA, Lv Y, Chen HS, Liu L. Outcomes of general anesthesia versus conscious sedation for stroke undergoing endovascular treatment: a meta-analysis. *BMC Anesthesiol*. 2019;19(1):69.
  23. Jing R, Dai HJ, Lin F, Ge WY, Pan LH. Conscious sedation versus general anesthesia for patients with acute ischemic stroke undergoing endovascular therapy: a systematic review and meta-analysis. *Biomed Res Int*. 2018;2018:2318489.
  24. John N, Mitchell P, Dowling R, Yan B. Is general anaesthesia preferable to conscious sedation in the treatment of acute ischaemic stroke with intra-arterial mechanical thrombectomy? A review of the literature. *Neuroradiology*. 2013;55(1):93–100.
  25. Berkhemer OA, van den Berg LA, Franssen PSS, Beumer D, Yoo AJ, Lingsma HF, et al. The effect of anesthetic management during intra-arterial therapy for acute stroke in MR CLEAN. *Neurology*. 2016;87(7):656–64.
  26. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*. 1988;19(5):604–7.
  27. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care*. 2011;15(3):R127.
  28. Akaike H. Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory; 1973.
  29. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med*. 1991;10(4):585–98.
  30. Schenker N. Inference with imputed conditional means AU - Schafer, Joseph L. *J Am Stat Assoc*. 2000;95(449):144–54.
  31. Venema E, Mulder M, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials. *BMJ*. 2017;357:j1710.
  32. Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2010;63(1):64–74.
  33. Sjoding MW, Luo K, Miller MA, Iwashyna TJ. When do confounding by indication and inadequate risk adjustment bias critical care studies? A simulation study. *Crit Care*. 2015;19:195.
  34. Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am Heart J*. 2000; 139(5):745–51.
  35. van Leeuwen N, Walgaard C, van Doorn PA, Jacobs BC, Steyerberg EW, Lingsma HF. Efficient design and analysis of randomized controlled trials in rare neurological diseases: an example in Guillain-Barre syndrome. *PLoS One*. 2019;14(2):e0211404.
  36. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*. 2006;59(10):1102–9.
  37. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. *J Clin Epidemiol*. 2007;60(9):979.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

