Estimation of treatment effects in observational stroke care data: comparison of statistical approaches

Marzyeh Amini[1*], Nikki van Leeuwen[1], Frank Eijkenaar[2], Rob van de Graaf[3,4], Noor Samuels[1,3,4], Robert van Oostenbrugge[5], Ido R. van den Wijngaard[6], Pieter Jan van Doormaal[3], Yvo B. W. E. M. Roos[7], Charles Majoie[8], Bob Roozenbeek[3,4], Diederik Dippel[4], James Burke[9], Hester F. Lingsma[1] and on behalf of the, MR CLEAN Registry Investigators

Abstract

Introduction: Various statistical approaches can be used to deal with unmeasured confounding when estimating treatment effects in observational studies, each with its own pros and cons. This study aimed to compare treatment effects as estimated by different statistical approaches for two interventions in observational stroke care data.

Patients and methods: We used prospectively collected data from the MR CLEAN registry including all patients ($n = 3279$) with ischemic stroke who underwent endovascular treatment (EVT) from 2014 to 2017 in 17 Dutch hospitals. Treatment effects of two interventions – i.e., receiving an intravenous thrombolytic (IVT) and undergoing general anesthesia (GA) before EVT – on good functional outcome (modified Rankin Scale ≤ 2) were estimated. We used three statistical regression-based approaches that vary in assumptions regarding the source of unmeasured confounding: individual-level (two subtypes), ecological, and instrumental variable analyses. In the latter, the preference for using the interventions in each hospital was used as an instrument.

Results: Use of IVT (range 66–87%) and GA (range 0–93%) varied substantially between hospitals. For IVT, the individual-level (OR ~ 1.33) resulted in significant positive effect estimates whereas in instrumental variable analysis no significant treatment effect was found (OR 1.11; 95% CI 0.58–1.56). The ecological analysis indicated no statistically significant different likelihood ($\beta = -0.002\%$; $P = 0.99$) of good functional outcome at hospitals using IVT 1% more frequently. For GA, we found non-significant opposite directions of points estimates the treatment effect in the individual-level (ORs ~ 0.60) versus the instrumental variable approach (OR = 1.04). The ecological analysis also resulted in a non-significant negative association (0.03% lower probability).

Discussion and conclusion: Both magnitude and direction of the estimated treatment effects for both interventions depend strongly on the statistical approach and thus on the source of (unmeasured) confounding. These issues should be understood concerning the specific characteristics of data, before applying an approach and interpreting the results. Instrumental variable analysis might be considered when unobserved confounding and practice variation is expected in observational multicenter studies.

*Correspondence: m.amini@erasmusmc.nl
[1] Department of Public Health, Erasmus University Medical Center, Erasmus MC, P.O. Box 2040, CA, Rotterdam, the Netherlands
Full list of author information is available at the end of the article

Amini *et al. BMC Medical Research Methodology*   (2022) 22:103

Page 2 of 12

Introduction

Assignment of treatment mostly depends on both measured and unmeasured patient characteristics (i.e., prognostic factors) that are associated with outcome [1, 2]. As a consequence, estimates of treatment effects in observational data tend to be biased because they reflect, at least in part, the effects of (unmeasured) confounding variables rather than the true effect of treatment [1, 2]. In randomized clinical trials (RCTs), such (unmeasured) confounding is avoided because randomization ensures all prognostic factors to be equally balanced between treatment groups [3]. However, RCTs can be expensive and challenging to conduct. In addition, for ethical and practical reasons executing an RCT could even be infeasible, or it could provide results with considerable delay. Furthermore, the external validity of RCTs might be limited due to often stringent inclusion criteria [2–4].

In practice, therefore, observational studies constitute the main alternative in absence of RCTs to approximate the true treatment effects of medical interventions. Several approaches – e.g. covariate adjustment, ecological analysis, and instrumental variable analysis – have been proposed to detect or control for (un) measured confounding in such studies [4–10]. With covariate adjustment, it is possible to directly adjust the treatment effect estimate for differences in measured prognostic factors in patients (e.g., disease severity) as well as in hospital factors (e.g., admission characteristics or practice style) that may influence the outcome [5]. An important downside of this approach is that it does not allow for controlling for differences of unmeasured factors that may influence both treatment allocation and outcome, causing confounding by indication. Moreover, sufficient sample sizes are required to allow for adequate statistical modeling of multiple confounding variables [5]. The ecological analysis approach is based on the analysis of differences in treatment decisions between groups of patients, i.e., at the hospital-level [4, 7, 8, 11]. Although ecological studies may reduce the effect of unmeasured confounding, this approach introduces new problems that are not present in individual-level analyses, such as the ecological fallacy, collinearity, and lack of power [11]. A third approach is instrumental variable analysis [6, 8]. This approach takes advantage of both reduced confounding by indication in the ecological analysis and more accurate specification of individual outcome and potential confounders using the individual patient data (just like when applying covariate adjustment). With instrumental variable analysis, a measured variable (i.e., the instrument) is used that influences treatment decisions but that is assumed to be unrelated to (both measured and unmeasured) patient characteristics that influence the outcome. When substantial hospital-level variation in the use of a specific intervention exists, the variable 'hospital' can be used as an instrument to study the effect of this intervention on the outcome [6, 10]. Choosing an appropriate instrument is the main challenge when using this approach.

In stroke care, endovascular treatment (EVT) has become standard care for patients with acute ischemic stroke caused by an intracranial large vessel occlusion of the anterior circulation, when the procedure is available and can be performed in a timely fashion [12–15]. All major guidelines recommend intravenous thrombolysis (IVT) in eligible patients before EVT [16, 17]. Patients receive IVT with alteplase as standard care, unless they have a contraindication for IVT. In addition, it might be necessary to perform the EVT procedure under a form of anesthesia, like conscious sedation or general anesthesia (GA), in case the patient is agitated, or a secured airway is required [18]. The effect of IVT and GA on neurological outcomes after EVT were evaluated in previous RCTs [18–25] and observational studies [16, 18, 26], but the results were inconclusive.

Given the large between-hospital variability in IVT and GA utilization before EVT [27], this study aimed to compare treatment effects of both IVT and GA estimated using observational data derived from a nationwide stroke care registry, to provide further insight into the influence of different analysis approaches and source of (un) measured confounding.

Methods

Study design and patients

For the current study, we used data collected between March 2014 and November 2017 from the MR CLEAN Registry, a prospective, observational study in all 17 hospitals that perform EVT in the Netherlands [28]. We included patients adhering to the following criteria: treatment at age 18 years and older, treatment in a hospital that participated in the MR CLEAN trial, clinical diagnosis of acute stroke with a deficit on the National Institute of Health Stroke Scale (NIHSS) of at least 2 points, CT or MRI ruling out intracranial hemorrhage, the possibility to start treatment within 6 h of onset, and proximal intracranial vessel occlusion in the anterior circulation (internal carotid artery, internal carotid artery terminus,

middle (M1/M2) cerebral artery, or anterior (A1/A2) cerebral artery), as shown by computed tomography angiography. Details on the study design and objectives of the MR CLEAN Registry have been described elsewhere [28]. Overall, data from 3279 patients were available for analysis.

Variables
Outcome measure
Good functional outcome as measured by a score of 0–2 on the modified Rankin Scale (mRS) was used as the outcome [29]. The mRS is a commonly used measure of patients' functional outcome after ischemic stroke care and ranges from 0 (no symptoms) to 6 (death). The mRS score was assessed at 90 days after EVT (± 14 days).

Treatment variables
We assessed two treatment variables. First, yes/no IVT administration before EVT was studied. Second, yes/no GA during EVT was assessed, no GA meaning either no sedation or conscious sedation.

Case-mix variables
Time between stroke onset and arrival at the emergency department (ED) of the EVT hospital and patients' age, sex, relevant medical history (i.e. previous stroke, atrial fibrillation, hypertension, hypercholesterolemia), and baseline score on the NIHSS were considered potential confounders of the estimated association between administration of IVT or GA and good functional outcome. The latter four variables were selected based on clinical knowledge and previous studies [30, 31]. The time between stroke onset and arrival at the ED was used because it cannot be influenced by hospitals while it may have an impact on the outcome.

Statistical analysis
All analyses were carried out using SPSS version 25.0 (IBM Corporation, Armonk, NY, USA) and Stata version 13.0 (StataCorp, College Station, TX, USA). Case-mix variables and outcome were described using summary statistics and measures of spread, and differences therein among hospitals and patient groups (i.e., yes/no IVT and GA) were tested on statistical significance ($P < 0.05$). Groups were compared using a non-parametric Kruskal Wallis test for continuous variables or Pearson's chi-square statistic for categorical variables. Next, separately for IVT and GA, we applied three statistical approaches to estimate and compare treatment effects, both unadjusted and case-mix adjusted: individual-level, ecological, and instrumental variable analysis.

Individual-level analysis
For the individual-level analysis we ran two separate models for each of the two interventions, with good functional outcome after 90 days (i.e., mRS 0–2) as the dependent variable: standard logistic regression and generalized estimating equations (GEE).

Logistic regression assumes independent observations for every patient. But in multicenter studies, it is not necessarily valid when individual observations (here patients) and their characteristics are 'clustered' in higher-level entities (here hospitals) [32]. Therefore, given differences among hospitals in the way they treat patients, correlation is likely within-hospital observations. The GEE model allows for accounting for this correlation (differences in the use of IVT or GA between hospitals) to appropriately adjust the estimation of effect sizes and confidence intervals [32]. In the case of significant within-hospital correlation, broader confidence intervals are estimated due to fewer independent observations. We accounted for clustering by hospitals using a compound symmetry correlation structure (correlation assumed to be the same for all within-hospital comparisons and all between-hospital comparisons). This means that we assumed the treatment effects were the same within the same cluster (hospital), and there was no variance between patients. Given our binary outcome variable, we used a binomial distribution and logistic link function for the GEE model [33].

Ecological analysis
Ecological analysis exploits differences in preferences for the use of IVT or GA between groups of patients (i.e., hospitals). The key assumption is that differences in the use of interventions across hospitals are mainly driven by hospitals' practice styles rather than differences in patients' prognostic factors. While practice style is difficult to detect at the individual patient level, its impact can be measured ecologically. Linear regression at the hospital-level was used to estimate the association between the percentage of patients with a good functional outcome (dependent variable) and the percentage of patients who received IVT or GA. To let high-volume hospitals contribute more to the analysis than low-volume hospitals, we used the number of patients treated in each hospital as weights. In the adjusted analysis, case-mix variables were measured at the hospital-level (i.e., mean age, the proportion of patients with the male sex, proportion of patients with a previous stroke, atrial fibrillation, hypertension, and hypercholesterolemia, mean NIHSS score, and mean time from onset to ED-arrival). This analysis provides insight into the change in absolute probability

of a good functional outcome for every additional 1% of patients receiving IVT or GA.

Instrumental variable analysis

The advantages of considering the hospital's practice style in the ecological analysis and the potentially more accurate specification of individual-level outcome and potential confounders in the individual-level analysis, were combined in the instrumental variable analysis. Specifically, preference for IVT or GA per hospital was used as an instrument, defined as the proportion of patients who received IVT or GA within each hospital [34, 35]. By using this instrument, we relied on three key assumptions for instrumental variable analysis (Fig. 1): (1) the instrument is associated with exposure to the treatment; (2) the instrument only has an effect on the outcome *through* treatment; and [3] the instrument is unrelated to (un) measured prognostic factors. Using the Stata command "*ivregress gmm*" [36] we used two-stage logistic regression to estimate the treatment effect on good functional outcome (Supplementary information). The first stage comprised running a case-mix adjusted logistic regression model to estimate the predicted probability of receiving IVT or GA given the preference to use those interventions at that hospital. In the second stage, the predicted probability of the first stage was used as a covariate in another logistic regression model at the individual patient level adjusted for case-mix variables, with yes/no good functional outcome as a dependent variable. In addition, the cluster effect of hospitals was added to the analysis and the standard errors were adjusted for 17 clusters. This model yields an estimate of the effect of IVT or GA on good functional outcome. Then, we assessed the validity of the instrumental variable analysis. Specifically, based on the first abovementioned assumption, the strength of the instrument was assessed using the *F*-statistic calculated as $[R^2 \times (n - 1 - K)]/[(1 - R^2) \times K]$, where '$R^2$' represents the partial variance from the first-stage, 'n' represents the sample size, and 'K' represents the number of instrumental variables included in the model [36, 37]. As a rule of thumb, the *F*-statistic for the significance of the instrument in the first stage should exceed 10. The second assumption was tested using the Durbin-Wu-Hausman χ^2 test to examine the existence of endogeneity in the estimated treatment effects. This test compares an estimate of the average treatment effect assuming no unmeasured confounding, to an estimate of the average treatment effect using an instrument that allows for unmeasured confounding [38]. The third assumption underlying instrumental variable analysis we tested was whether the instrumental variable was associated with case-mix variables.

Results

Descriptive analyses

The number of patients receiving IVT varied significantly between hospitals (range 66–87%, $P < 0.01$) and GA (range 0–93%, $P < 0.01$) (Fig. 2). The number of EVT patients also varied widely from 23 to 405 patients per hospital (Table 1). There were significant differences ($P < 0.05$) in case-mix variables between patients who received the interventions of interest (i.e., IVT and GA) and patients who did not (Table 1).

IVT intervention

In the adjusted logistic regression model, receiving IVT was significantly associated with higher odds of good functional outcome (Fig. 3A-1). As the within-hospital correlation between patients was very small (0.0053), the estimated odds ratios and width of the confidence
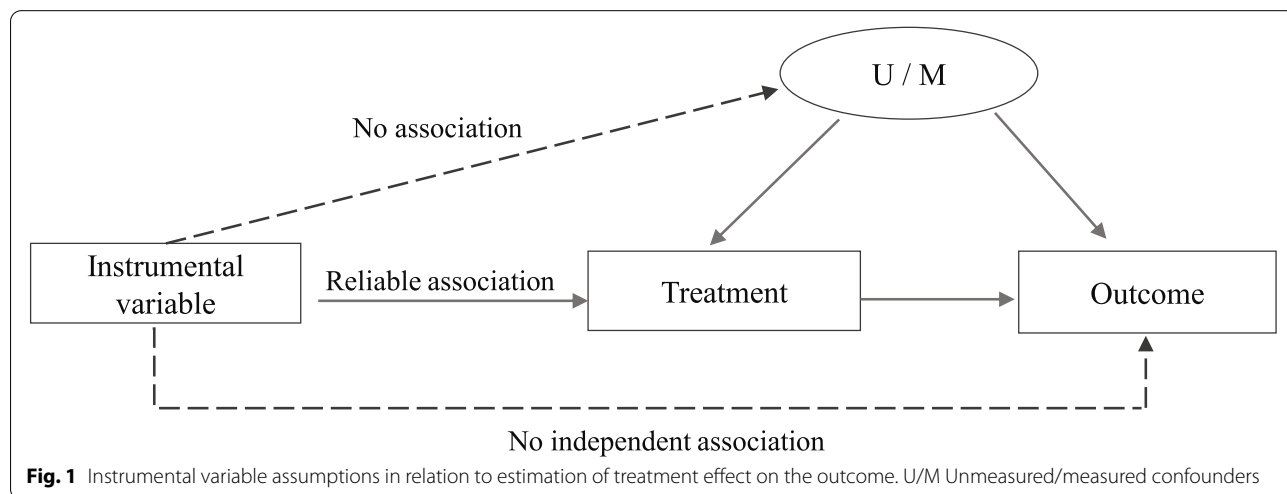

Fig. 1 Instrumental variable assumptions in relation to estimation of treatment effect on the outcome. U/M Unmeasured/measured confounders
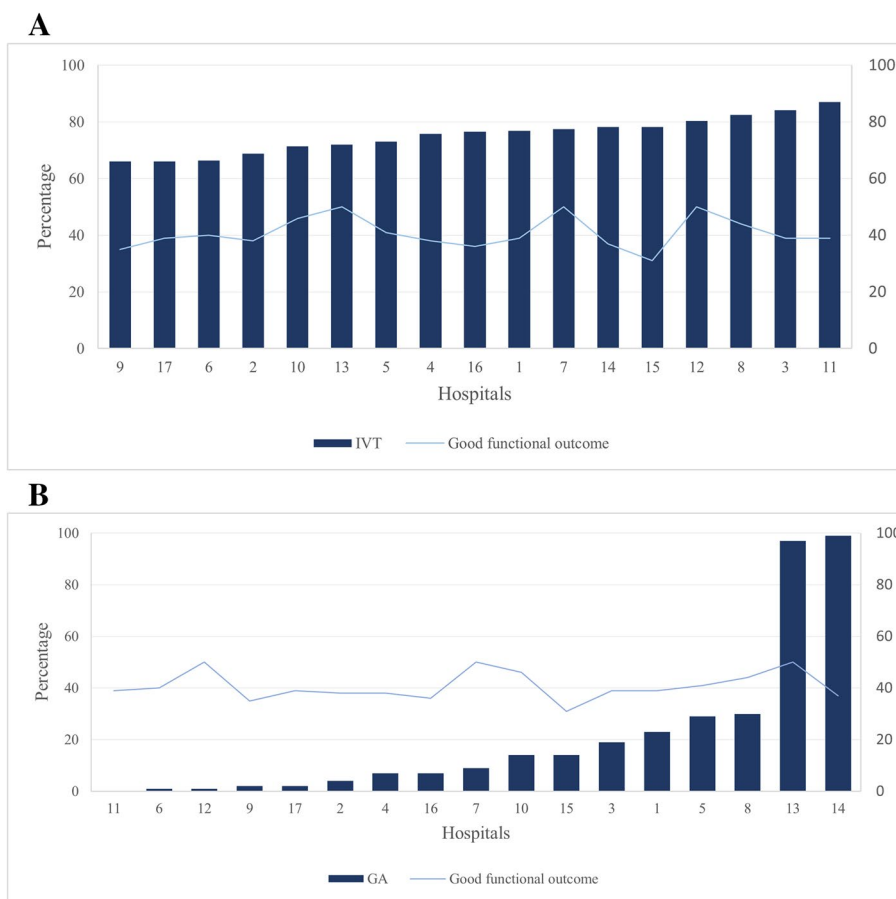
Fig. 2 Differences in the percentage of patients receiving IVT (**A**) and undergoing general anesthesia (**B**) before EVT and good functional outcome in 17 EVT hospitals in the Netherlands. *Note*: Good functional outcome is defined as mRS 0–2 at 90 days

intervals in the GEE model were very similar to those in the logistic regression model (Fig. 3A-2). The same pattern was noticed in the unadjusted models, in which larger treatment effect estimates were estimated, as compared to the adjusted models (Figs. 3A-1 and A-2).

In the ecological analysis at the hospital-level, no statistically significant different likelihood (β = − 0.002%; P = 0.99) of good functional outcome at hospitals using IVT 1% more frequently was estimated after case-mix adjustment (Fig. 3B). In the case-mix adjusted instrumental variable analysis, no significant treatment effect was found when patients treated in hospitals were using IVT more frequently (OR 1.11; 95% CI 0.58–1.56; Fig. 3C). These results were not significantly different from the individual-level estimates from the logistic and GEE analyses (Durbin-Wu-Hausman $\chi^2 = 0.073$, *P-value* for difference = 0.75), and no endogeneity was noticed. The instrument explained 1.6% of the variation in the use of IVT in the adjusted model. The derived *F*-statistics was 41.2 in the adjusted model, which indicates that the instrument has sufficient power. However, we did find

the instrument to be associated with some confounders including age, previous stroke, previous atrial fibrillation, and baseline NIHSS score (Supplementary Table 1). Thus, the assumptions underlying an appropriate instrumental variable analysis were not fully met.

GA intervention
The adjusted logistic regression analyses of patients undergoing GA on good functional outcome showed no statistically significant effect (OR 0.87; 95% CI 0.71–1.07; Fig. 4A-1). The within-hospital correlation was 0.022 and in the adjusted GEE model the treatment effect was statistically significant (OR 0.60; 95% CI 0.44–0.80; Fig. 4A-2) and larger compared to the effect in the logistic regression model.

The ecological analysis showed no statistically significant different likelihood of a good functional outcome in the adjusted model, following a 1% increase in the use of GA (Fig. 4B).

In contrast to the individual-level results, the instrumental variable analysis showed that the use of GA has

Amini *et al. BMC Medical Research Methodology*   (2022) 22:103

Page 6 of 12

Table 1 Case-mix characteristics of patients treated in all 17 EVT hospitals in the Netherlands; stratified by whether patients received IVT or not and whether patients underwent GA or not

	hospitals range (n = 17)	P-values[b]	IVT + (n = 2450)	IVT - (n = 817)	P-values[b]	GA + (n = 778)	GA - (n = 2304)	P-values[c]
Age (years) [Median (IQR)]	68–77	0.001	71 (60–80)	73 (63–82)	< 0.001	70 (59–79)	73 (62–81)	< 0.001
Baseline NIHSS score [Median (IQR)]	13–17	< 0.001	16 (11–19)	16 (11–20)	0.048	17 (12–20)	16 (11–19)	0.011
Time from onset to arrival at the ED (min) [Median (IQR)]	52–160	< 0.001	135 (65–185)	145 (67–240)	< 0.001	140 (72–190)	135 (64–195)	0.538
Men [N (%)]	39–55%	0.790	1294 (53)	394 (48)	0.024	420 (54)	1178 (51)	0.171
Medical History [N (%)]								
Previous Stroke	0–26%	< 0.001	330 (14)	212 (26)	< 0.001	127 (16)	389 (17)	0.739
Atrial Fibrillation	13–37%	< 0.001	414 (17)	356 (44)	< 0.001	157 (20)	567 (25)	0.012
Hypertension	41–67%	< 0.001	1234 (50)	447 (55)	0.037	346 (45)	1251 (54)	< 0.001
Hypercholesterolemia	15–50%	< 0.001	696 (28)	269 (33)	0.023	207 (27)	720 (31)	< 0.001
Good functional outcome[a] [N (%)]	31–50%	0.004	985 (40)	251 (30)	< 0.001	272 (35)	896 (39)	0.309
IVT utilization [N (%)]	66–87%	< 0.001						
GA utilization [N (%)]	0–99%	< 0.001						
Hospital volume [n]	23–405							

IVT Intravenous thrombolysis, *GA* general anesthesia, *NIHSS* National institutes of health stroke scale, *ED* Emergency department

[a] Good functional outcome is defined as mRS 0–2 at 90 days

[b] *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

[c] *P-value* is based on comparison between groups (receiving IVT/GA vs. non-receiving IVT/GA) using a non-parametric Kruskal Wallis test for continuous variables or Pearson's chi-square statistic for categorical variables
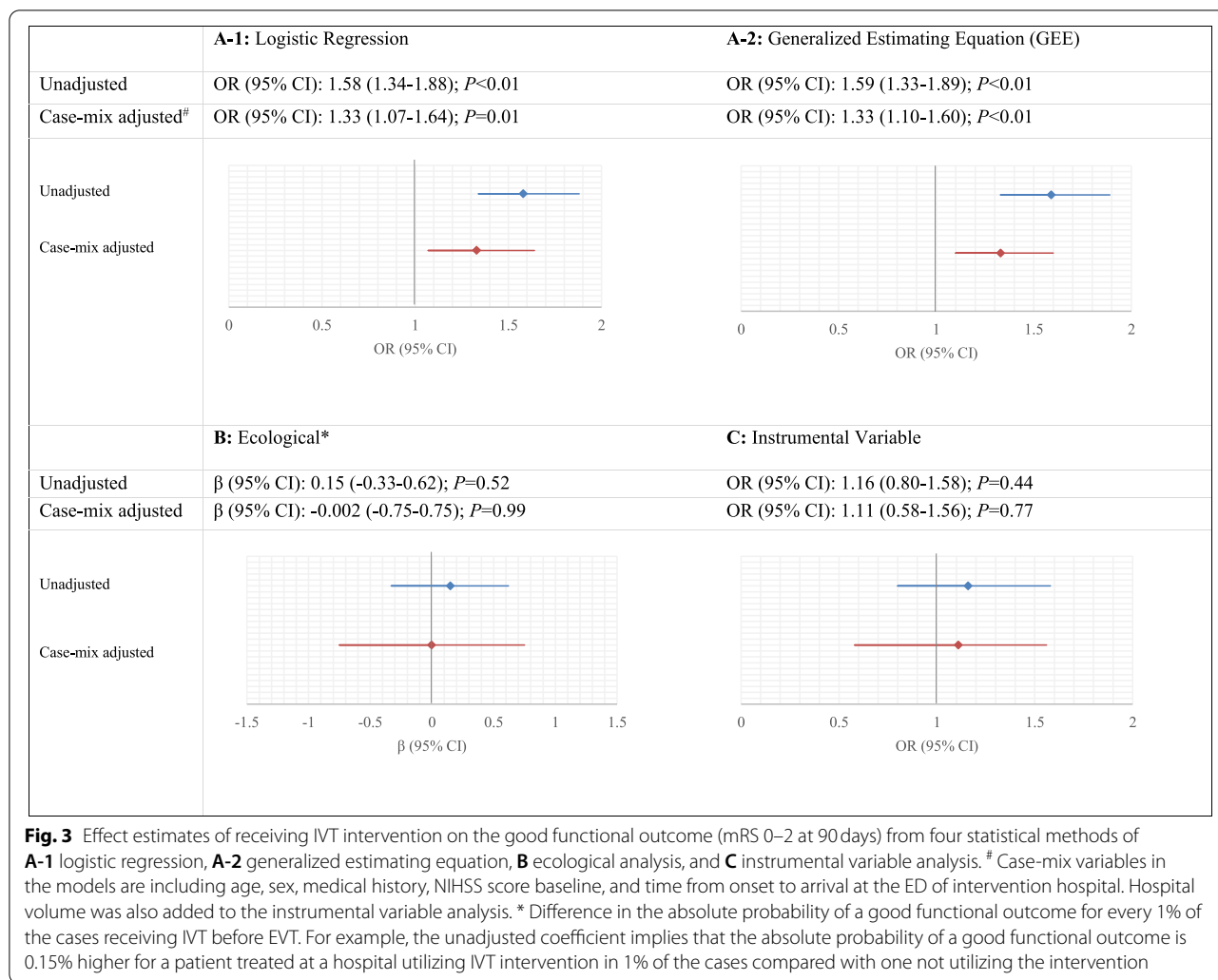
a positive but non-significant effect on a good functional outcome (OR 1.04; 95% CI 0.95–1.15; Fig. 4C). Durbin-Wu-Hausman χ^2 was significant (17.6, $P < 0.01$) in the adjusted model, which indicates that one of the assumptions underlying the instrumental variable is met. The instrumental variable explained about 5.9% of the variation in the use of GA, regardless of case-mix adjustment. The *F*-statistic was 11,243.5, indicating a sufficiently powerful instrument. In addition, associations between the instruments and measured confounders were generally small (Supplementary Table 1), which indicates that another assumption underlying the instrumental variable was met.

Discussion

The challenge when using observational data to estimate treatment effects is to overcome unmeasured confounding [8, 39, 40]. In the current study, we applied and compared various analytical approaches that have been advocated to reduce confounding in observational data, to assess the effect of IVT and GA on the probability of good functional outcome after ischemic stroke due to a large vessel occlusion in patients treated with EVT. We found that the size and direction of estimated effects strongly depend on the statistical approach used and

therefore on the source of unmeasured confounding, i.e. individual patient prognostic factor or hospital practice.

In observational studies, regression analysis adjusted for patient characteristics is commonly used to estimate treatment effects of interventions. Our results of these analyses showed that patients who received IVT had significantly better outcomes than patients who did not receive IVT. This was consistent with the recent RCTs [22, 41] and inconsistent with other trials [18, 24, 25] results. Our adjusted regression analysis showed that patients receiving GA had no statistically significant higher odds of good functional outcome. Previous RCTs showed no conclusive results on the effect of GA versus other types of anesthesia [20, 21, 23]. In those trials, the comparator was different than the comparator in our study. In our data 'no GA' included patients with either no sedation or conscious sedation, but in those trials, GA was compared with conscious sedation only or local anesthesia. Generally speaking, individual-level regression analysis could provide unbiased estimates of treatment effects if all relevant confounders are known, measured, and adjusted for [5, 6]. But because of the difficulty to assess medical indications and underlying disease severity and prognosis, confounding by indication is often an impassable problem in multivariable regression analysis

	A-1: Logistic Regression	**A-2:** Generalized Estimating Equation (GEE)
Unadjusted	OR (95% CI): 1.58 (1.34-1.88); *P*<0.01	OR (95% CI): 1.59 (1.33-1.89); *P*<0.01
Case-mix adjusted[#]	OR (95% CI): 1.33 (1.07-1.64); *P*=0.01	OR (95% CI): 1.33 (1.10-1.60); *P*<0.01

	B: Ecological*	**C:** Instrumental Variable
Unadjusted	β (95% CI): 0.15 (-0.33-0.62); *P*=0.52	OR (95% CI): 1.16 (0.80-1.58); *P*=0.44
Case-mix adjusted	β (95% CI): -0.002 (-0.75-0.75); *P*=0.99	OR (95% CI): 1.11 (0.58-1.56); *P*=0.77

Fig. 3 Effect estimates of receiving IVT intervention on the good functional outcome (mRS 0–2 at 90 days) from four statistical methods of **A-1** logistic regression, **A-2** generalized estimating equation, **B** ecological analysis, and **C** instrumental variable analysis. [#] Case-mix variables in the models are including age, sex, medical history, NIHSS score baseline, and time from onset to arrival at the ED of intervention hospital. Hospital volume was also added to the instrumental variable analysis. * Difference in the absolute probability of a good functional outcome for every 1% of the cases receiving IVT before EVT. For example, the unadjusted coefficient implies that the absolute probability of a good functional outcome is 0.15% higher for a patient treated at a hospital utilizing IVT intervention in 1% of the cases compared with one not utilizing the intervention

at the individual-level [5, 6]. In our observational data, some important prognostic variables were available (e.g., age, comorbidity, the severity of disease) and treatment groups were unbalanced concerning these variables. Still, it is possible that we have missed other important unmeasured prognostic factors, which might be associated with outcome. For example, many patients received IVT in a primary stroke center and were transferred to an intervention center. These patients were treated with EVT following no response to IVT. Also, many patients in the dataset were not eligible for IVT since they were outside the time window for treatment or because of other contraindications. Therefore, the treatments effects of IVT or GA as estimated in our individual-level analyses may be biased. This also could explain the different effect estimates in our analyses compared to the results in the RCTs. Additionally, an alternative approach for the measured confounding factor adjustment is the propensity score statistical methods, such as propensity score matching or inverse probability of treatment weighting, that may put different results [42, 43].

In addition, in multicenter studies at the individual-level, treatment effects are likely to be biased when observations are not independent but clustered at the hospital-level. To deal with this and to adjust for the potential effects of unmeasured hospital characteristics, GEE can be used [44]. For the IVT intervention, the within-hospital correlation between patients was very small and thus the GEE and logistic regression model yielded very similar results. In contrast, for GA the within-hospital correlation was moderate, resulting in no statistically significant effect in the logistic regression model compared to a significantly lower odds of good functional outcome in the GEE model. Similar to GEE, other approaches like random effect analysis which deal with dependent observations can put a similar argument (Supplementary Fig. 1 and Supplementary Fig. 2). In the presence of hospital-level clustering and unmeasured
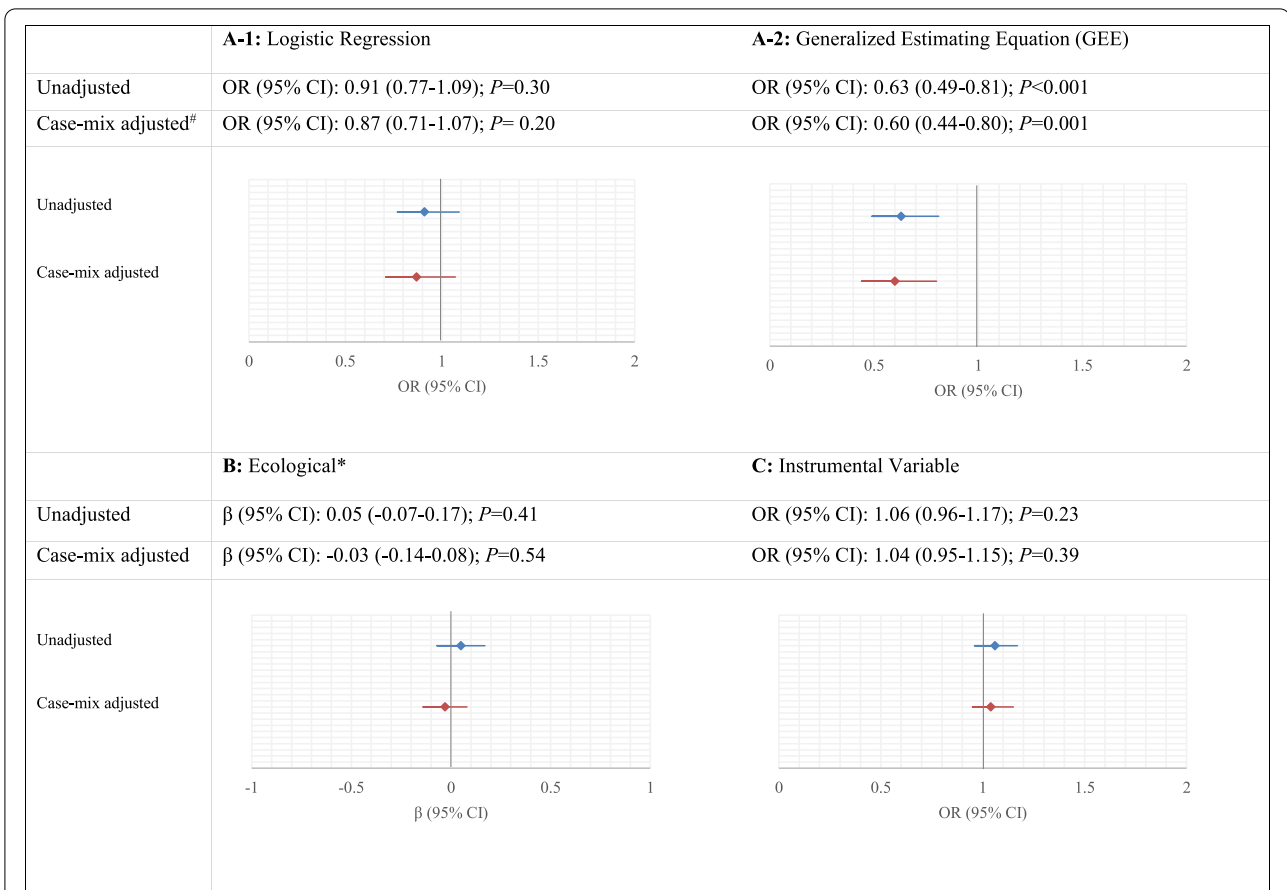
	A-1: Logistic Regression	**A-2:** Generalized Estimating Equation (GEE)
Unadjusted	OR (95% CI): 0.91 (0.77-1.09); *P*=0.30	OR (95% CI): 0.63 (0.49-0.81); *P*<0.001
Case-mix adjusted#	OR (95% CI): 0.87 (0.71-1.07); *P*= 0.20	OR (95% CI): 0.60 (0.44-0.80); *P*=0.001

	B: Ecological*	**C:** Instrumental Variable
Unadjusted	β (95% CI): 0.05 (-0.07-0.17); *P*=0.41	OR (95% CI): 1.06 (0.96-1.17); *P*=0.23
Case-mix adjusted	β (95% CI): -0.03 (-0.14-0.08); *P*=0.54	OR (95% CI): 1.04 (0.95-1.15); *P*=0.39

Fig. 4 Effect estimates of undergoing general anesthesia on the good functional outcome (mRS 0–2 at 90 days) from four statistical methods of **A-1** logistic regression, **A-2** generalized estimating equation, **B** ecological analysis, and **C** instrumental variable analysis. # Case-mix variables in the models are including age, sex, medical history, NIHSS score baseline, and time from onset to arrival at the ED of intervention hospital. Hospital volume was also added to the instrumental variable analysis. * Difference in the absolute probability of a good functional outcome for every 1% of the cases receiving general anesthesia. For example, the unadjusted coefficient implies that the absolute probability of a good functional outcome is 0.05% higher for a patient treated at a hospital utilizing general anesthesia in 1% of the cases compared with one not utilizing the intervention

hospital characteristics influencing the outcome, GEE or other approaches which deal with clustering will be superior to conventional regression analysis for treatment effect estimation. However, this type of analysis might still be biased in the case of a small number of hospitals and/or small within-hospital sample size [44].

In an ecological analysis, aggregated data instead of individual-level data is used. Only the proportion of patients who received the intervention at a given hospital and aggregated information on case-mix was included. This approach is adjusted for confounding by hospital practice, especially in the case of a large variation in the use of the intervention between hospitals (as in the current study) [7]. For both interventions, observed differences in treatment effect estimates in the hospital-level versus the individual-level analyses suggest unmeasured confounding by hospital practice to be present to some extent. An ecological analysis at the level of the hospital

bypasses the issue of confounding by indication. Patients' prognostic factors are no longer placed in the model—only the aggregate values for a hospital's treated population are considered. The portion of cases treated with interventions at a given hospital is more likely to be influenced by practice style rather than the prognostic factors of the case mix, particularly because some hospitals do not utilize the interventions. Yet, given its hospital-level approach, ecological analysis is prone to bias. For example, some hospitals treated more patients who were transferred from other hospitals, suggesting differences in some patient characteristics, which may have resulted in a change in the probability of a good functional outcome. The intuitive solution of adding these factors to the model (if available) would weaken rather than strengthen the analysis. In our analysis, the model had just 17 observations and already 8 independent variables, and adding additional variables would (further) increase the potential

for collinearity and reduce power [11]. Another potential issue is the ecological fallacy [11], which could lead to effect estimates that do not exist or are in the opposite direction of the effects at the individual-level. Thus, when individual-level data are unavailable, this approach does not allow for conclusions about individual-level treatment effects. Combined with the biased outcome measure due to possibly missing data at certain hospitals [11], this warrants caution in interpreting treatment effects derived from this type of analysis [45]. In general, the ecological analysis would only be superior to individual-level analysis when the influence of hospital practice on treatment is independent of prognostic factors and the number of observations is sufficiently high.

Instrumental variable analysis has become increasingly popular in epidemiology as it can exploit the advantages of both individual-level and hospital-level approaches and can provide unbiased treatment effect estimates [46–48]. This is underscored by the fact that for IVT, our instrumental variable analysis results are in line with recent RCTs; among patients with acute large vessel occlusion stroke, these trials showed noninferiority of combined IVT and EVT relative to EVT alone regarding good functional outcome [18, 24, 25]. This was in contrast to other RCTs that failed to demonstrate noninferiority [22, 41]. In this analysis, the assumption of endogeneity due to unobserved confounding was violated, but still, in comparison to RCTs, instrumental variable analysis is superior to the conventional regression method. In addition, the results of previous RCTs have not been conclusive on the impact of the type of anesthesia on neurological outcome after EVT for stroke patients [20, 21, 49, 50]. However, comparing our results with those RCTs might not be valid since the comparator was different [51]. No violations of the instrumental variable assumption were noticed in the GA intervention analysis. This demonstrates that our instrumental variable analysis for this intervention was more efficient than for the IVT intervention and its results might be more valid than those of the conventional approaches. In instrumental variable analysis, a reliable instrumental variable must satisfy criteria related to the instrument and sample size to allow for adequate estimation of the treatment effect [52]. A comparison of the precision of the estimates was assessed by looking at the widths of the confidence intervals around effect estimates. Larger widths are interpreted as a less precise effect estimate. Weak instruments might produce large confidence intervals which result in imprecise (and biased) results. For the GA intervention, the intervals around the estimates were narrower using the instrumental variable approach than conventional approaches. For the IVT intervention, the intervals were broader, as a result of the weak instrument. For the IVT intervention, using hospital

as an instrument might not be valid, since the receiving IVT is not very hospital dependent but more related to (unmeasured) contraindications at the patient level. In addition, the instrument was associated with some measured confounders. The adjusted instrument, by definition, is independent of those factors. On the other hand, the fact that we have measured confounders that predict the instrument makes it likely that other unmeasured predictors also predict treatment, that we cannot account for. Therefore, for reliability and validity of the instrumental variable analysis, it is essential that a suitable instrument is used to overcome unmeasured confounding. Reliability is mentioned as the most important limitation of instrumental variable analysis. Boef et al. [53] states that the statistical power of instrumental variable analysis is dependent on instrument strength and the strength of unmeasured confounding, but will usually be large given the typical moderate instrument strength in medical research. However, the increasing availability of data (sources) and existence of large registry databases, can accommodate the lack of statistical power of this method.

In summary, we recommend that investigators analyzing observational studies consider the sources of (un) measured confounding that may affect their results. The standard analysis should not be blindly accepted before considering what type of (un) measured confounding may contribute to the results. Several approaches can be used in different phases of study to account for unmeasured confounding [4]. Instrumental variable analysis might be used in conjunction with other approaches, such as multivariate regression analysis, propensity score methods, or ecological analysis to better describe the effect of treatment.

Strengths and limitations

A major strength of our study is that we used the prospective data from a nationwide registry of all 17 EVT hospitals with various relevant case-mix variables which enable us to compare four approaches and demonstrate the influence of the analytical approach on the estimated effect.

The first limitation of the study is that we were not able to compare our treatment effect estimates to an estimate derived from an RCT with the same comparator. As a result, we cannot prove one approach as being superior to the others given our observational data. The second limitation might be that we only analyzed one clinical dichotomous outcome and no other outcomes with possibly other scales (i.e., ordinal, continuous). Thus, it is uncertain whether the results of this study are also generalizable to other outcomes. The third limitation is that we only used three approaches to compare the treatment effect. There are other approaches with probability of different

Amini *et al. BMC Medical Research Methodology*   (2022) 22:103

Page 10 of 12

results like the propensity score calibration and two-stage calibration statistical methods [54–56] which can be an alternative approach for instrumental variable analysis; or propensity score matching and inverse probability of treatment weighting [42, 43] as an alternative to covariate adjustment. The fourth potential limitation is that for some of the models we compared (e.g. GEE), we made specific analytical choices that may have affected (the statistical significance of) our effect estimates. For example, in the GEE model, we assumed a compound symmetry correlation structure. While this is a common assumption in this type of research [57], it may not (fully) hold. Finally, it should be recognized that we deliberately included a limited number of covariates to keep consistency over different models, while adjusting for other patient or hospital characteristics, e.g., the time from arrival at the emergency department to EVT could have led to improved model performance and different treatment effect estimates.

Conclusion

Using observational data, the choice of statistical approach has important consequences for the size, direction, and statistical significance of estimated treatment effects. Each approach has its merits and limitations and deals differently with the potential influence of (un) measured confounding. Before applying an approach and interpreting the results, researchers should understand these issues in relation to the specific characteristics of their data. While it is not desirable to appoint one approach as being always preferable, the instrumental variable analysis should be considered when unobserved confounding and hospital practice variation are expected in observational multicenter studies.

Abbreviations

ED: Emergency department; GA: General anesthesia; GEE: Generalized estimating equation; IVT: Intravenous thrombolysis; mRS: Modified Rankin Scale; NIHSS: National Institute of Health Stroke Scale.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-022-01590-0.

Additional file 1.

Amini *et al. BMC Medical Research Methodology*   (2022) 22:103

Page 11 of 12

List of affiliations.
Department of Neurology[4], Radiology[3], Public Health[1], Erasmus MC University Medical Center;
Department of Radiology and Nuclear Medicine[8], Neurology[7], Biomedical Engineering & Physics[42], Amsterdam UMC, University of Amsterdam, Amsterdam;
Department of Neurology[5], Radiology[10], Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM);
Department of Neurology[11], Radiology[12], Sint Antonius Hospital, Nieuwegein;
Department of Neurology[13], Radiology[14], Leiden University Medical Center;
Department of Neurology[15], Radiology[16], Rijnstate Hospital, Arnhem;
Department of Radiology[6], Neurology[17], Haaglanden MC, the Hague;
Department of Neurology[18], Radiology[19], HAGA Hospital, the Hague;
Department of Neurology[20], Radiology[21], University Medical Center Utrecht;
Department of Neurology[22], Neurosurgery[23], Radiology[30], Radboud University Medical Center, Nijmegen;
Department of Neurology[24], Radiology[29], Elisabeth-TweeSteden ziekenhuis, Tilburg;
Department of Neurology[25], Radiology[26], Isala Klinieken, Zwolle;
Department of Neurology[27], Radiology[28], Reinier de Graaf Gasthuis, Delft;
Department of Neurology[31], Radiology[32], University Medical Center Groningen;
Department of Neurology[33], Radiology[34], Atrium Medical Center, Heerlen;
Department of Neurology[35], Radiology[36], Catharina Hospital, Eindhoven;
Department of Neurology[37], Radiology[38], Medical Spectrum Twente, Enschede;
Department of Radiology[39], Amsterdam UMC, Vrije Universiteit van Amsterdam, Amsterdam; Department of Radiology[40], Noordwest Ziekenhuisgroep, Alkmaar; Department of Radiology[41], Texas Stroke Institute, Texas, United States of America.

Authors' contributions
HFL, BR, and DD designed the study. MA undertook analyses and interpretation of study findings and wrote the first draft of the manuscript. NvL, HFL, FE, BR, DD contributed to the interpretation of study findings, reviews, and revision of the manuscript. RG, NS, RO, IRW, PJD, YR, CHM, JB 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 data that support the findings of this study are available from the MR CLEAN Registry executive committee but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the MR CLEAN Registry executive committee.

Declarations

Ethics approval and consent to participate
This was an analysis of an existing data set. Ethics for the original MR CLEAN Registry was approved by the medical ethics committee of the Erasmus University MC, Rotterdam, the Netherlands (MEC-201-235). Then, informed consent for study participation was obtained from all subjects in MR CLEAN Registry previously. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Public Health, Erasmus University Medical Center, Erasmus MC, P.O. Box 2040, CA, Rotterdam, the Netherlands. [2]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands. [3]Department of Radiology and Nuclear Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [4]Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. [5]Maastricht University Medical Center, Maastricht, the Netherlands. [6]Department of Neurology, Haaglanden Medical Center, the Hague, the Netherlands. [7]Amsterdam University Medical Centers, location AMC, Amsterdam, the Netherlands. [8]Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, location AMC, Amsterdam, the Netherlands. [9]Department of Neurology, University of Michigan, Ann Arbor, MI, USA.

References

1. Joseph KS, Mehrabadi A, Lisonkova S. Confounding by indication and related concepts. Curr Epidemiol Rep. 2014;1(1):1–8.
2. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. Jama. 2016;316(17):1818–9.
3. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. Hepatology (Baltimore, Md). 2006;44(5):1075–82.
4. Uddin MJ, Groenwold RH, Ali MS, de Boer A, Roes KC, Chowdhury MA, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. Int J Clin Pharm. 2016;38(3):714–23.
5. Cnossen MC, van Essen TA, Ceyisakar IE, Polinder S, Andriessen TM, van der Naalt J, et al. Adjusting for confounding by indication in observational studies: a case study in traumatic brain injury. Clin Epidemiol. 2018;10:841–52.
6. 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.
7. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin Epidemiol. 1999;52(1):7–12.
8. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. Pharmacoepidemiol Drug Saf. 2003;12(7):551–8.
9. Sedgwick P. Ecological studies: advantages and disadvantages. BMJ : British Med J. 2014;348:g2979.
10. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology (Cambridge, Mass). 2006;17(4):360–72.
11. Sedgwick P. Ecological studies: advantages and disadvantages. BMJ (Clinical research ed). 2014;348:g2979.
12. 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 (London, England). 2016;387(10029):1723–31.
13. Mueller-Kronast NH, Zaidat OO, Froehler MT, Jahan R, Aziz-Sultan MA, Klucznik RP, et al. Systematic evaluation of patients treated with Neurothrombectomy devices for acute ischemic stroke: primary results of the STRATIS registry. Stroke. 2017;48(10):2760–8.
14. Sheth SA, Jahan R, Gralla J, Pereira VM, Nogueira RG, Levy EI, et al. Time to endovascular reperfusion and degree of disability in acute stroke. Ann Neurol. 2015;78(4):584–93.
15. Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie 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.
16. Chalos V, LeCouffe NE, Uyttenboogaart M, Lingsma HF, Mulder M, Venema E, et al. Endovascular treatment with or without prior intravenous Alteplase for acute ischemic stroke. J Am Heart Assoc. 2019;8(11):e011592.
17. Powers WJ, Derdeyn CP, Biller J, Coffey CS, Hoh BL, Jauch EC, et al. 2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early Management of Patients with Acute Ischemic Stroke Regarding Endovascular Treatment: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2015;46(10):3020–35.
18. Yang P, Zhang Y, Zhang L, Zhang Y, Treurniet KM, Chen W, et al. Endovascular Thrombectomy with or without intravenous Alteplase in acute stroke. N Engl J Med. 2020;382(21):1981–93.

19. Albers GW. Thrombolysis before Thrombectomy - to be or DIRECT-MT? N Engl J Med. 2020;382(21):2045–6.
20. Kim C, Kim S-E, Jeon JP. Influence of anesthesia type on outcomes after endovascular treatment in acute ischemic stroke: Meta-analysis. Neurointervention. 2019;14(1):17–26.
21. Simonsen CZ, Yoo AJ, Sørensen 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.
22. Suzuki K, Matsumaru Y, Takeuchi M, Morimoto M, Kanazawa R, Takayama Y, et al. Effect of mechanical Thrombectomy without vs with intravenous thrombolysis on functional outcome among patients with acute ischemic stroke: the SKIP randomized clinical trial. Jama. 2021;325(3):244–53.
23. Wan T-F, Zhang J-R, Liu L. Effect of general anesthesia vs. conscious sedation on the outcomes of acute ischemic stroke patients after endovascular therapy: a Meta-analysis of randomized clinical trials. Front Neurol 2019;10:1131.
24. Zi W, Qiu Z, Li F, Sang H, Wu D, Luo W, et al. Effect of endovascular treatment alone vs intravenous Alteplase plus endovascular treatment on functional Independence in patients with acute ischemic stroke: the DEVT randomized clinical trial. Jama. 2021;325(3):234–43.
25. Neale Td. MR CLEAN-NO IV: No advantage to skipping tPA before stroke Thrombectomy. https://www.tctmdcom/news/mr-clean-no-iv-no-advantage-skipping-tpa-stroke-thrombectomy (under publication). 2021.
26. 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.
27. Amini M, van Leeuwen N, Eijkenaar F, Mulder MJHL, Schonewille W, Lycklama à NG, et al. Improving quality of stroke care through benchmarking center performance: why focusing on outcomes is not enough. BMC Health Serv Res. 2020;20(1):998.
28. 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 (Clinical research ed) 2018;360:k949.
29. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Inter-observer agreement for the assessment of handicap in stroke patients. Stroke. 1988;19(5):604–7.
30. 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.
31. 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.
32. Christopher JWZ. Generalized estimating equation Models for correlated data: a review with applications. Am J Polit Sci. 2001;45(2):470–90.
33. Parzen M, Ghosh S, Lipsitz S, Sinha D, Fitzmaurice GM, Mallick BK, et al. A generalized linear mixed model for longitudinal binary data with a marginal logit link function. Ann Appl Stat. 2011;5(1):449–67.
34. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat. 2007;3(1):Article 14.
35. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology (Cambridge, Mass). 2006;17(3):268–75.
36. Nonrecursive Models: Endogeneity, reciprocal relationships, and feedback loops. Thousand Oaks, California2011. Available from: https://methods.sagepub.com/book/nonrecursive-models.
37. Andrews I, Stock JH, Sun L. Weak instruments in instrumental variables regression: theory and practice. Annual Rev Econ. 2019;11(1):727–53.
38. Guo Z, Cheng J, Lorch SA, Small DS. Using an instrumental variable to test for unmeasured confounding. Stat Med. 2014;33(20):3528–46.
39. Bigby M. Challenges to the hierarchy of evidence: does the emperor have no clothes? Arch Dermatol. 2001;137(3):345–6.
40. McMahon AD. Observation and experiment with the efficacy of drugs: a warning example from a cohort of nonsteroidal anti-inflammatory and ulcer-healing drug users. Am J Epidemiol. 2001;154(6):557–62.
41. Fischer U, Gralla J. Direct mechanical thrombectomy fails to show non-inferiority compared with IV t-PA plus thrombectomy for LVOs. https://neuronewsinternationalcom/swift-direct-trial-results-esoc-2021/. 2021.
42. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol. 2005;58(6):550–9.
43. Laborde-Castérot H, Agrinier N, Thilly N. Performing both propensity score and instrumental variable analyses in observational studies often leads to discrepant results: a systematic review. J Clin Epidemiol. 2015;68(10):1232–40.
44. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? Stat Med. 2009;28(2):221–39.
45. Naylor CD. Ecological analysis of intended treatment effects: caveat emptor. J Clin Epidemiol. 1999;52(1):1–5.
46. John ER, Abrams KR, Brightling CE, Sheehan NA. Assessing causal treatment effect estimation when using large observational datasets. BMC Med Res Methodol. 2019;19(1):207.
47. Vansteelandt S, Didelez V. Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators. Scand J Stat. 2018;45(4):941–61.
48. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014;33(13):2297–340.
49. 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.
50. Abou-Chebl A, Yeatts SD, Yan B, Cockroft K, Goyal M, Jovin T, et al. Impact of general anesthesia on safety and outcomes in the endovascular arm of interventional Management of Stroke (IMS) III trial. Stroke. 2015;46(8):2142–8.
51. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. Circulation. 2021;143(10):1002–13.
52. Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. J Econ Perspect. 2001;15(4):69–85.
53. Boef AG, Dekkers OM, Vandenbroucke JP, le Cessie S. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. J Clin Epidemiol. 2014;67(11):1258–64.
54. Lin HW, Chen YH. Adjustment for missing confounders in studies based on observational databases: 2-stage calibration combining propensity scores from primary and validation data. Am J Epidemiol. 2014;180(3):308–17.
55. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. Am J Epidemiol. 2005;162(3):279–89.
56. Su CC, Yang YK, Lai EC, Hsieh CY, Cheng CL, Chen CH, et al. Comparative safety of antipsychotic medications in elderly stroke survivors: a nationwide claim data and stroke registry linkage cohort study. J Psychiatr Res. 2021;139:159–66.
57. Wang L. GEE analysis of clustered binary data with diverging number of covariates. Ann Stat. 2011;39(1):389–417. 29.

Publisher's Note