

## Practice of Epidemiology

# Benchmarking Observational Analyses Before Using Them to Address Questions Trials Do Not Answer: An Application to Coronary Thrombus Aspiration

**Anthony A. Matthews\***, Issa J. Dahabreh, Ole Fröbert, Bertil Lindahl, Stefan James, Maria Feychting, Tomas Jernberg, Anita Berglund, and Miguel A. Hernán

\* Correspondence to Dr. Anthony A. Matthews, Institutet för Miljömedicin, Karolinska Institutet, Nobels väg 13, 171 65 Solna, Sweden (e-mail address: [anthony.matthews@ki.se](mailto:anthony.matthews@ki.se)).

*Initially submitted September 27, 2021; accepted for publication May 24, 2022.*

To increase confidence in the use of observational analyses when addressing effectiveness questions beyond those addressed by randomized trials, one can first benchmark the observational analyses against existing trial results. We used Swedish registry data to emulate a target trial similar to the Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia (TASTE) randomized trial, which found no difference in the risk of death or myocardial infarction by 1 year with or without thrombus aspiration among individuals with ST-elevation myocardial infarction. We benchmarked the emulation against the trial at 1 year and then extended the emulation's follow-up to 3 years and estimated effects in subpopulations underrepresented in the trial. As in the TASTE trial, the observational analysis found no differences in risk of outcomes by 1 year between groups (risk difference = 0.7 (confidence interval,  $-0.7, 2.0$ ) and  $-0.2$  (confidence interval,  $-1.3, 1.0$ ) for death and myocardial infarction, respectively), so benchmarking was considered successful. We additionally showed no difference in risk of death or myocardial infarction by 3 years, or within subpopulations by 1 year. Benchmarking against an index trial before using observational analyses to answer questions beyond those the trial could address allowed us to explore whether the observational data can be trusted to deliver valid estimates of treatment effects.

benchmarking; causal inference; observational analyses; randomized trial; target trial emulation

Abbreviations: CI, confidence interval; PCI, percutaneous coronary intervention; STEMI, ST-elevation myocardial infarction; SWEDEHEART, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies; TASTE, Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia.

Randomized trials are the preferred approach to estimate causal effects of clinical interventions. Randomized trials, however, cannot answer all clinically relevant causal questions, including those about long-term treatment effects or effects in individuals who do not enroll in trials. Analyses of observational databases are often used to complement the estimates of randomized trials, but observational analysis estimates may be confounded because differences in risk between treatment groups may be explained by differences between the individuals in each group rather than by the effect of treatment (1, 2). Therefore, causal analyses of observational data adjust for known and measured confounders, although there is no guarantee that such adjustment suffices to eliminate confounding bias (3).

One possible approach to increase confidence in observational effect estimates is benchmarking, that is, to demonstrate that the observational analysis is able to replicate an index trial's findings (e.g., effect on death by 1 year) before using the observational data to estimate what the index trial could not estimate (e.g., effect on death by 3 years if the index trial had a follow-up of 1 year, or the effect within subpopulations that were not well represented in the index trial) (4). As an example, consider the Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia (TASTE) randomized trial as our index trial. TASTE found no difference in the risk of death or myocardial infarction by 30 days or 1 year when comparing percutaneous coronary intervention (PCI) with and without thrombus aspiration among

individuals with ST-elevation myocardial infarction (STEMI) in the Nordic countries (5, 6). TASTE was designed to study the effects of thrombus aspiration by 1 year after baseline; analyses of observational data may be able to complement these results and make further inferences beyond those made by the TASTE trial. Trust in such observational analyses designed to ask a similar question as TASTE would be increased if they agreed with the 1-year trial results.

Successful agreement, however, requires sufficient adjustment for confounding, and it is possible that the structure of confounding varies before and after the publication of the trial, especially if that trial contributed to changes in the reasons for receiving treatment. Before TASTE there was evidence that thrombus aspiration improved coronary artery flow after PCI, but it was unknown whether it improved clinical endpoints such as mortality (7–9). In Sweden, before TASTE, this uncertainty resulted in some centers implementing the routine use of thrombus aspiration, while others left it to the discretion of the operator. After TASTE found no beneficial effect of thrombus aspiration, in Sweden, routine thrombus aspiration was largely reserved for patients with large thrombi in a coronary artery (10).

Here, we used observational data from the national Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART) Registry, which is the same registry in which TASTE was nested, to emulate a target trial similar to TASTE. By carrying out the observational analysis in the same registry as the trial, we ensure that the causal question was asked in the same health-care setting. After evaluating whether observational data before and after TASTE were comparable, we informally benchmarked the observational analysis results against the trial results at 1 year, then extended follow-up to 3 years and explored effects in subpopulations by 1 year.

## THE INDEX RANDOMIZED TRIAL: TASTE

### Trial design and analysis

TASTE was a multicenter, prospective, randomized, controlled, open-label clinical trial carried out between June 2010 and March 2013 (5, 6). In total, 31 PCI centers recruited participants: 29 in Sweden, 1 in Iceland, and 1 in Denmark. SWEDEHEART was used to collect information for Swedish participants. Individuals were eligible for TASTE if PCI was planned for the treatment of acute STEMI (see Table 1 for additional criteria). Individuals who accepted the invitation to participate were randomly assigned to receive PCI either with or without thrombus aspiration. The primary end point was death by any cause within 30 days of PCI, and additional analyses explored death by any cause, rehospitalization for myocardial infarction, and stent thrombosis with 1 year of PCI. Data on clinical end points were obtained from the Cause of Death and SWEDEHEART registries. The intention-to-treat analyses compared 1-year risk curves from Kaplan-Meier analyses and estimated the corresponding average hazard ratios from Cox proportional-hazards models.

## Trial results

As reported in the original TASTE publication, during the enrollment period there were 11,956 individuals with STEMI, approximately 9,420 individuals potentially eligible for enrollment in Sweden and Iceland (eligible individuals unknown in Denmark), and 7,244 individuals randomized to a treatment arm in Sweden, Iceland, and Denmark; 3,621 were assigned to thrombus aspiration and 3,623 to no thrombus aspiration (6). Web Table 1 (available at <https://doi.org/10.1093/aje/kwac098>) shows the baseline characteristics, and Table 2 shows the 1-year risks and average hazard ratios. The risk of each individual outcome did not differ between the treatment groups. The 1-year risk of death was 5.3% in individuals in the thrombus aspiration group and 5.6% in the no thrombus aspiration group, with a hazard ratio of 0.94 (95% confidence interval (CI): 0.78, 1.15). The 1-year risk of myocardial infarction was 2.7% in both groups, with a hazard ratio of 0.97 (95% CI: 0.73, 1.28). Stent thrombosis was rare; the 1-year risk was 0.7% in the thrombus aspiration group and 0.9% in the no thrombus aspiration group, with a hazard ratio of 0.84 (95% CI: 0.50, 1.40).

## THE OBSERVATIONAL ANALYSIS

Causal inference from observational data can be seen as an attempt to emulate a pragmatic randomized trial—the target trial—that would answer the question of interest. The approach for emulating a target trial has 2 steps: 1) specify the protocol of the target trial, and 2) emulate the target trial using the available observational data and appropriate methodology (11). To compare TASTE to an observational analysis that aims to ask the same questions, using data from the SWEDEHEART Registry, we first specified a protocol of a target trial similar to the protocol of TASTE, with deviations only when the observational data did not correspond to the information collected in the trial. We then emulated the target trial using the SWEDEHEART registry data. Table 1 summarizes all protocol elements from the target trial and its emulation, which we describe herein.

### The target trial protocol

*Eligibility criteria.* The eligibility criteria of the target trial would be the same as TASTE with 5 exceptions. First, the enrollment period would be September 2007 to January 2016, excluding June 2010 to March 2013, which is the period of participation in TASTE. Second, only Swedish coronary intervention centers would be included (no data are available from the Icelandic and Danish centers in the observational data). Third, the possibility of performing thrombus aspiration would not be assessed. Fourth, correspondence between electrocardiogram findings and culprit artery pathoanatomy would not be assessed. Fifth, individuals who died on the day of PCI would be excluded and identification of outcomes would start from the day after PCI as it is not possible to distinguish whether outcome events other than death (i.e., myocardial infarction) occurred before

**Table 1.** Description of TASTE Randomized Trial, Target Trial, and Target Trial Emulation Using the SWEDEHEART Registry

Protocol Component	TASTE (Index Randomized Trial)	Target Trial	Target Trial Emulation Using SWEDEHEART
Eligibility criteria	Age 18 years or older between June 15, 2010, and March 25, 2013 Diagnosis of ST-elevation myocardial infarction as defined by chest pain suggestive for myocardial ischemia for at least 30 minutes before hospital admission, time from onset of symptoms of less than 24 hours, and an ECG with new ST-segment elevation in two or more contiguous leads of greater than or equal to 0.2 mV in leads V2-V3 and/or greater than or equal to 0.1 mV in other leads or a probable new-onset left bundle branch block. Planned percutaneous coronary intervention in one of the 29 Swedish, 1 Icelandic, or 1 Danish coronary intervention centers Minimum of 50% stenosis in culprit artery by visual estimate Correspondence between ECG findings and culprit artery pathoanatomy Possibility of performing thrombus aspiration	Same as TASTE apart from: Study period September 4, 2007, to January 4, 2016, excluding June 15, 2010, to March 25, 2013, which was the period of recruitment for TASTE  Only in the Swedish coronary intervention centers  Possibility to perform thrombus aspiration not assessed Correspondence between ECG findings and culprit artery pathoanatomy not assessed Individuals excluded if died on same day as percutaneous coronary intervention	Same as target trial apart from:  No informed consent asked, so not able to exclude those who would have not been asked or who would have declined participation if asked
Treatment strategies	No emergency coronary artery bypass grafting No previous randomization in the TASTE trial Provided informed consent 1) No thrombus aspiration followed by percutaneous coronary intervention: balloon dilatation, balloon dilatation and stenting, or direct stenting to achieve antegrade flow. Post-dilatation of stents is optional. 2) Thrombus aspiration followed by percutaneous coronary intervention: thrombus aspiration with an Export aspiration catheter (Medtronic Inc., Santa Rosa, California). Continuous manual suction is performed using a proximal-to-distal approach, which is defined as active aspiration during initial passage of the lesion. In lesions that cannot initially be passed with the thrombus aspiration catheter, it is permitted to dilate the lesion with an angioplasty balloon up to a maximal nominal diameter size of 2.0 mm and attempt to advance the thrombus aspiration catheter for a second time. After thrombus aspiration, percutaneous coronary intervention is done as described above.	Same as TASTE	Same as target trial

Table continues

Table 1. Continued

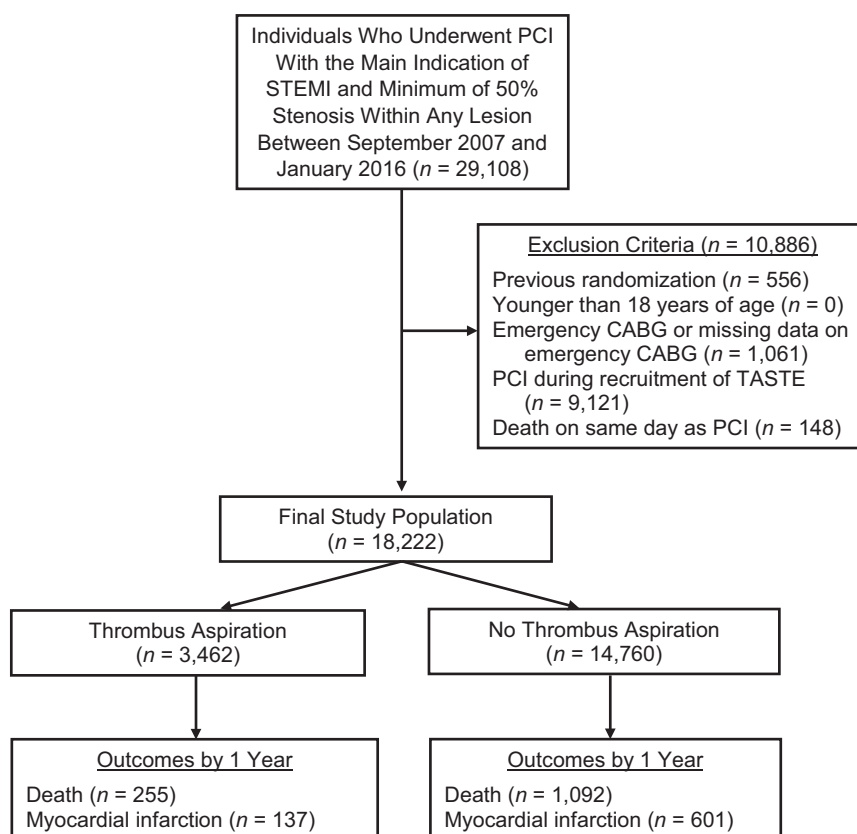
Protocol Component	TASTE (Index Randomized Trial)	Target Trial	Target Trial Emulation Using SWEDEHEART
Treatment assignment	Individuals randomized to a treatment strategy (by center)	Individuals would be randomized to a treatment strategy and were aware of the assigned strategy.	Individuals assigned to the strategy that their data were compatible with. Assignment was treated as if randomized within levels of the following baseline covariates: age, sex, hospital, diabetes, body mass index, smoking, hyperlipidemia, hypertension, previous infarction, previous percutaneous coronary intervention, previous coronary artery bypass graft, stenosis class, proportion stenosis in culprit artery, angiography finding, heart rate, systolic blood pressure, diastolic blood pressure, thrombolysis, warfarin, aspirin, clopidogrel, prasugrel, heparin, low molecular weight heparin, bivalirudin, and Gp2b3a inhibitors.
Outcomes	Death from any cause Rehospitalization for myocardial infarction Stent thrombosis	Same as TASTE	Same as target trial apart from: No stent thrombosis due to few events Outcomes identified as following: Death from any cause from the Swedish Cause of Death Register by 1 year Myocardial infarction from the SWEDEHEART Register by 1 year Same as target trial
Follow-up	Starts at treatment assignment and ends at date of first outcome (separately for analysis of each outcome), migration, or 1 year	Same as TASTE apart from: Unable to identify migration date Started from day after percutaneous coronary intervention Follow-up to 3 years	
Causal contrasts	Intention-to-treat effect, per-protocol effect	Intention-to-treat effect, per-protocol effect	Observational analogue of the per-protocol effect

Table continues

Table 1. Continued

Protocol Component	TASTE (Index Randomized Trial)	Target Trial	Target Trial Emulation Using SWEDEHEART
Statistical analysis	Kaplan-Meier plots	For intention-to-treat analyses, survival curves estimated using a pooled logistic regression outcome model with an indicator for assigned treatment group, a flexible time-varying intercept, product terms between treatment group and time, and standardization of the period-specific cumulative probabilities. Comparison of risks via differences and ratios then estimated.	Same per protocol analysis as target trial with adjustment for baseline covariates
	Treatment differences were assessed with the use of the log-rank test and Cox regression.	Per-protocol analyses use the same technique as above, but restricted to individuals who received their assigned treatment, and with the inclusion of baseline covariates in the outcome models.	
		Nonparametric bootstrapping with 200 samples used to calculate 95% confidence intervals.	

Abbreviations: ECG, electrocardiogram; Gp2b3a, glycoprotein 2b/3a; SWEDEHEART, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies; TASTE, Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia.



**Figure 1.** Flowchart of individuals eligible for an observational emulation of a target trial of thrombus aspiration versus no thrombus aspiration, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART) registry, Sweden, 2007–2016. There were 18,222 eligible individuals, of whom 3,462 were given thrombus aspiration and 14,760 were not given thrombus aspiration. CABG, coronary artery bypass graft; PCI, percutaneous coronary intervention; STEMI, ST-elevation myocardial infarction; TASTE, Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia.

or after PCI when the events occurred on the same day as the procedure.

*Treatment strategies.* The treatment strategies in the target trial would be the same as those in TASTE: PCI, either 1) with thrombus aspiration or 2) without thrombus aspiration.

*Treatment assignment.* The target trial would randomly assign eligible individuals to one of the treatment strategies, and the physicians would be aware of the strategy to which the patient had been assigned.

*Outcomes.* The outcomes in the target trial would be death from any cause, myocardial infarction, or stent thrombosis.

*Follow-up.* The target trial would follow each individual from the day after treatment assignment until the outcome of interest (separate analysis for each outcome), or either 1 year for benchmarking or 3 years for analyses with extended follow-up, whichever occurred first. It is not possible to identify migration date, so outcome data on those who migrated out of Sweden are unavailable. However, only about 0.5% of the Swedish population emigrates each year (12). We expect this proportion to be even lower among individuals eligible

for our study, who recently had a myocardial infarction and are receiving regular health care.

*Causal contrasts.* The target trial would estimate the intention-to-treat effect, which is the effect of being assigned to thrombus aspiration or no thrombus aspiration, and the per-protocol effect, which is the effect of receiving the assigned thrombus aspiration or no thrombus aspiration.

*Statistical analysis.* For the intention-to-treat analysis, we estimated the survival curves in each group defined by assigned treatment strategy via a parametric pooled logistic model with an indicator for treatment group, a flexible time-varying intercept, and product terms between treatment group and time. We compared the estimated risks (1 – survival) via differences and ratios. To estimate the total effect on myocardial infarction, individuals who die are treated as not experiencing the outcome after death rather than as censored at death (13). For the per-protocol analysis, we used the same technique as above, except that the analysis was restricted to individuals who received their assigned treatment, baseline covariates are included in the outcome models, and the estimated probabilities are standardized to

**Table 2.** Estimated Risks, Risk Differences, and Risk Ratios From the TASTE Randomized Trial and an Observational Emulation of a Target Trial of Thrombus Aspiration Versus No Thrombus Aspiration, SWEDEHEART Registry, 2007–2016

Follow-up Time and Outcome	Thrombus Aspiration		No Thrombus Aspiration		RD	95% CI	RR	95% CI
	Risk, %	95% CI	Risk, %	95% CI				
TASTE <sup>a</sup>								
1 year								
Death	5.3		5.6				0.94	0.78, 1.15
Myocardial infarction	2.7		2.7				0.97	0.73, 1.28
Observational analysis <sup>b</sup>								
1 year								
Death	8.0	6.7, 9.3	7.3	6.8, 7.9	0.7	−0.7, 2.0	1.09	0.96, 1.24
Myocardial infarction	3.9	2.9, 4.9	4.1	3.6, 4.5	−0.2	−1.3, 1.0	0.96	0.79, 1.17
3 year <sup>b</sup>								
Death	13.3	11.8, 14.7	12.4	11.7, 13.1	0.9	−0.7, 2.4	1.07	0.98, 1.17
Myocardial infarction	6.7	5.6, 7.9	6.9	6.4, 7.5	−0.2	−1.5, 1.1	0.97	0.85, 1.11

Abbreviations: CI, confidence interval; RD, risk difference; RR, risk ratio; SWEDEHEART, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies; TASTE, Thrombus Aspiration in ST-Elevation Myocardial Infarction in Scandinavia.

<sup>a</sup> Risk estimates from Kaplan-Meier analyses, and no confidence intervals were provided in the published trial results; risk ratios are hazard ratios from a Cox proportional hazards model with treatment as the only regressor.

<sup>b</sup> Adjusted at baseline for: age, sex, hospital, diabetes, body mass index, smoking, hyperlipidemia, hypertension, previous infarction, previous percutaneous coronary intervention, previous coronary artery bypass graft, stenosis class, proportion stenosis, angiography finding, heart rate, systolic blood pressure, diastolic blood pressure, thrombolysis, and use of warfarin, aspirin, clopidogrel, prasugrel, heparin, low molecular weight heparin, bivalirudin, or glycoprotein 2b/3a inhibitors.

the distribution of the baseline covariates (14). Nonparametric bootstrapping with 200 samples was used to calculate 95% confidence intervals.

### Emulating the target trial in the SWEDEHEART registry

**Data sources.** SWEDEHEART collects data from all patients hospitalized for acute coronary syndrome or undergoing coronary or valvular intervention for any indication in all relevant hospitals across Sweden (15). The registry was created by merging 4 existing cardiovascular health-care quality registries in 2009: the Register of Information and Knowledge About Swedish Heart Intensive Care Admissions (RIKSHIA), the Swedish Coronary Angiography and Angioplasty Registry (SCAAR), the Swedish Heart Surgery Registry and the National Registry of Secondary Prevention (SEPHIA), and the Swedish Heart Surgery Registry. SWEDEHEART was used to collect information for patients when they were randomized in the TASTE trial, hence the data collection process was broadly similar between the 2 studies. SWEDEHEART is also linked to the Swedish National Patient Register, which records all primary and secondary diagnoses and procedures from inpatient hospitalizations and outpatient specialist care visits across Sweden; the Swedish Cause of Death Register, which records all deaths and causes of death; and the Prescribed Drug Register, which collects information on all dispensed medications (16–18).

**Eligibility criteria.** We identified individuals in the SWEDEHEART registry who met the eligibility criteria. As in all observational emulations, no informed consent was asked, and hence we could not exclude individuals who would not have been asked or who would have declined participation if asked.

**Treatment strategies and assignment.** As treatment had already been given under routine clinical practice, we assigned eligible individuals in SWEDEHEART to the strategy their data were compatible with at baseline, and proceeded as if treatment was randomly assigned within levels of the following baseline covariates (full detail on covariates and their definitions in Web Table 2): age, sex, hospital, diabetes, body mass index, smoking, hyperlipidemia, hypertension, previous infarction, previous PCI, previous coronary artery bypass graft, stenosis class, proportion stenosis, angiography finding, heart rate, systolic blood pressure, diastolic blood pressure, thrombolysis, and use of warfarin, aspirin, clopidogrel, prasugrel, heparin, low molecular weight heparin, bivalirudin, and glycoprotein 2b3a inhibitors.

**Outcomes.** We did not use stent thrombosis as an outcome because few events had been reported. We identified deaths from the Cause of Death Register and myocardial infarctions from the SWEDEHEART Registry. See Web Table 3 for further details on outcomes and their definitions.



*Follow-up.* Follow-up was the same as the target trial.

*Causal contrasts.* It was possible to estimate only the observational analog of the per-protocol effect as SWEDEHEART collects information on the treatment an individual actually received, not what they were assigned.

*Statistical analysis.* The per-protocol analysis was the same as described above. Details of our modeling approach are presented in Web Appendix 1. We additionally stratified data by time period (before the TASTE trial began enrollment and after the TASTE trial completed enrollment), repeated the analyses, and compared estimates using data from each period to assess whether results at 1 year of follow-up were comparable, regardless of changing reasons for receiving thrombus aspiration following publication of TASTE trial results.

We carried out 9 sensitivity analyses: 1) We did not apply the eligibility criterion of 50% minimum stenosis (there was a high degree of missingness for the proportion-stenosis variable used to identify this eligibility in the period before TASTE); 2) to understand the impact of measured covariates on effect estimates, we conducted a separate analysis in which we adjusted for age and sex only, and we computed the difference between the fully adjusted risk difference and the age- and sex-adjusted risk difference; 3) we dropped all individuals with any missing data for baseline covariates (complete-case analysis); 4) we censored individuals at death in the myocardial infarction analysis; 5) we defined myocardial infarction using a 2-day gap between discharge following the initial period in hospital and the new myocardial infarction event to account for individuals that who transferred between different hospitals without a new event; 6) we additionally included a Killip class variable in the models when data were stratified into time after TASTE (Killip class was collected from June 2009, so there was a high degree of missing data before TASTE); 7) we additionally included an indicator for time period (before or after the TASTE trial) in the models; 8) we estimated the standardized risk of each outcome separately in each treatment arm to allow for all possible interactions between treatment and covariates; and 9) we adjusted for baseline covariates using inverse probability weighting.

We informally benchmarked 1-year results from the emulation against the results of the TASTE trial. If risk contrasts when comparing those with and without thrombus aspiration were similar to those from TASTE, and the same clinical decision would be made regardless of the study used to inform the decision, benchmarking was deemed successful; analyses were then replicated to estimate the 3-year risks, and data were stratified to estimate treatment effects by 1 year in subpopulations of individuals within strata of sex (female/male), age (<65 years/≥65 years), diabetes (no/yes), previous PCI (no/yes), and previous myocardial infarction (no/yes).

## RESULTS

Figure 1 shows a flowchart of selection for the target trial emulation. There were 18,222 eligible individuals, of whom

3,462 were given thrombus aspiration and 14,760 were not given thrombus aspiration. Table 3 shows the baseline characteristics of all eligible individuals. Before standardization, there were differences between groups for several variables including age, hospital, stenosis in culprit artery, and angiography finding (standardized mean differences in Web Table 4).

Table 2 shows the estimated 1-year risks, risk differences, and risk ratios for death and myocardial infarction. The estimated risk of death was 8.0% (95% CI: 6.7, 9.3) in the thrombus aspiration group and 7.3% (95% CI: 6.8, 7.9) in the group without thrombus aspiration; which results in a risk difference of 0.7% (95% CI: -0.7, 2.0) and a risk ratio of 1.09 (95% CI: 0.96, 1.24). The estimated risk of myocardial infarction was 3.9% (95% CI: 2.9, 4.9) in the thrombus aspiration group and 4.1% (95% CI: 3.6, 4.5) in the group without thrombus aspiration; which results in a risk difference of -0.2% (95% CI: -1.3, 1.0) and a risk ratio of 0.96 (95% CI: 0.79, 1.17).

Web Table 5 shows the baseline characteristics stratified by period (before and after TASTE enrollment), and Web Table 6 and Web Figure 1 show results when using these stratified data for analysis. The 1-year risk of death and myocardial infarction did not differ between the treatment groups in both time periods, so use of data from both enrollment periods for benchmarking appears justified. Web Tables 7–16 show results from sensitivity analyses; all results were broadly similar those from the primary analyses. Treatment groups were generally balanced in terms of the observed covariates after inverse probability weighting (Web Table 4).

## Benchmarking

Results of the target trial emulation at 1 year were informally benchmarked against results from the intention-to-treat analyses in TASTE (Table 2). Results appeared compatible within sampling variability: Both the estimates from TASTE and the emulated target trial were very compatible with a similar range of hazard and risk ratio values for death (TASTE 95% CI: 0.78, 1.15; emulated target trial 95% CI: 0.96, 1.24) or myocardial infarction (TASTE 95% CI: 0.73, 1.28; emulated target trial 95% CI: 0.79, 1.17) by 1 year in the groups with or without thrombus aspiration.

## Extended follow-up

Figure 2 shows the 3-year survival curves and Table 2 also shows the estimated 3-year risks, risk differences, and risk ratios. The estimated risk of death was 13.3% (95% CI: 11.8, 14.7) in the thrombus aspiration group and 12.4% (95% CI: 11.7, 13.1) in the group without thrombus aspiration, which results in a risk difference of 0.9% (95% CI: -0.7, 2.4) and a risk ratio of 1.07 (95% CI: 0.98, 1.17). The estimated risk of myocardial infarction was 6.7% (95% CI: 5.6, 7.9) in the thrombus aspiration group and 6.9% (95% CI: 6.4, 7.5) in the group without thrombus aspiration; which results in a risk difference of -0.2% (95% CI: -1.5, 1.1) and a risk ratio of 0.97 (95% CI: 0.85, 1.11).



**Table 3.** Baseline Characteristics of Eligible Individuals From an Observational Emulation of a Target Trial of Thrombus Aspiration Versus No Thrombus Aspiration, SWEDHEART Registry, 2007–2016

Characteristic	Thrombus Aspiration (n = 3,462)		No Thrombus Aspiration (n = 14,760)	
	No.	%	No.	%
Age, years <sup>a</sup>	66.0 (57.0, 74.0)		68.0 (60.0, 77.0)	
Female sex	887	25.6	4,422	30.0
Hospital				
Borås	15	0.4	143	1.0
Danderyd	134	3.9	363	2.5
Eskilstuna	67	1.9	398	2.7
Falun	211	6.1	638	4.3
Gävle	333	9.6	506	3.4
Halmstad	13	0.4	267	1.8
Helsingborg	22	0.6	90	0.6
Huddinge	28	0.8	131	0.9
Jönköping	49	1.4	679	4.6
Kalmar	84	2.4	566	3.8
Karlskrona	165	4.8	619	4.2
Karlstad	65	1.9	778	5.3
Karolinska Solna	255	7.4	1,099	7.4
Kristianstad	4	0.1	144	1.0
Linköping	251	7.3	713	4.8
Lund	785	22.7	1,929	13.1
Malmö	32	0.9	199	1.3
Sahlgrenska	164	4.7	1,387	9.4
Skövde	50	1.4	487	3.3
St Görans	42	1.2	62	0.4
Sunderbyn	19	0.5	281	1.9
Sundsvall	26	0.8	214	1.4
SÖS	170	4.9	260	1.8
Trollhättan	65	1.9	328	2.2
Umeå	33	1.0	327	2.2
Uppsala	160	4.6	810	5.5
Västerås	85	2.5	283	1.9
Örebro	125	3.6	998	6.8
Östersund	6	0.2	47	0.3
Östra sjukhuset	4	0.1	14	0.1
Stenosis class				
A	168	4.9	928	6.3
B1	882	25.5	4,550	30.8
B2	1,346	38.9	6,035	40.9
C	1,060	30.6	3,214	21.8
Other	6	0.2	33	0.2

Table continues

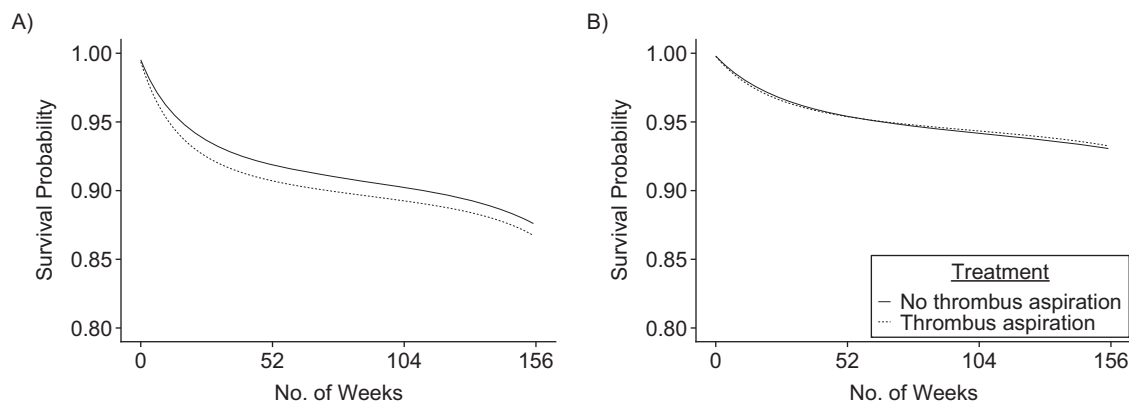
Table 3. Continued

Characteristic	Thrombus Aspiration (n = 3,462)		No Thrombus Aspiration (n = 14,760)	
	No.	%	No.	%
Stenosis in culprit artery				
50%–69%	34	1.0	198	1.3
70%–89%	118	3.4	1,081	7.3
90%–99%	510	14.7	4,406	29.9
100%	2,800	80.9	9,075	61.5
Angiography finding				
Normal	2	0.1	16	0.1
1 vessel	1,957	56.5	7,260	49.2
2 vessels	931	26.9	4,271	28.9
3 vessels	450	13.0	2,537	17.2
Left main	117	3.4	659	4.5
Missing	5	0.1	17	0.1
BMI <sup>a,b</sup>	26.0 (24.0, 29.0)		26.0 (24.0, 29.0)	
Missing	839	24.2	3,642	24.7
Smoking status				
Never	1,157	33.4	5,596	37.9
Former smoker (>1 month)	957	27.6	4,045	27.4
Current smoker	1,038	30.0	3,999	27.1
Missing	310	9.0	1,120	7.6
Diabetes	428	12.4	2,292	15.5
Hyperlipidemia treatment	724	20.9	3,327	22.5
Hypertension treatment	1,340	38.7	6,628	44.9
Previous myocardial infarction	426	12.3	1,996	13.5
Previous percutaneous coronary intervention	368	10.6	1,563	10.6
Previous coronary artery bypass grafting	64	1.8	337	2.3
Thrombolysis	16	0.5	54	0.4
Warfarin	72	2.1	303	2.1
Aspirin	3,341	96.5	14,347	97.2
Clopidogrel or ticlopidine	2,141	61.8	6,357	43.1
Prasugrel	118	3.4	623	4.2
Heparin	2,796	80.8	12,648	85.7
Low-molecular weight heparin	311	9.0	883	6.0
Bivalirudin	1,729	49.9	7,081	48.0
Glycoprotein IIb/IIIa inhibitors	1,457	42.1	3,906	26.5
Heart rate <sup>a</sup>	74.0 (61.0, 87.0)		75.0 (63.0, 88.0)	
Missing	245	7.1	784	5.3
Systolic blood pressure <sup>a</sup>	138.0 (120.0, 157.0)		141.0 (125.0, 160.0)	
Missing	257	7.4	837	5.7
Diastolic blood pressure <sup>a</sup>	80.0 (70.0, 95.0)		84.0 (72.0, 96.0)	
Missing	457	13.2	1,475	10.0

Abbreviations: BMI, body mass index; SWEDEHEART, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies.

<sup>a</sup> Values are expressed as median (interquartile range).

<sup>b</sup> Weight (kg)/height (m)<sup>2</sup>.



**Figure 2.** Standardized survival curves for outcomes of death (A) and myocardial infarction (B) from an observational emulation of a target trial of thrombus aspiration versus no thrombus aspiration, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART) registry, Sweden, 2007–2016. The 3-year risk difference for death was 0.9% (95% confidence interval:  $-0.7, 2.4$ ), and the 3-year risk difference of myocardial infarction was  $-0.2\%$  (95% confidence interval:  $-1.5, 1.1$ ).

### Subgroup effects

Table 4 shows the 1-year risks, risk differences, and risk ratios stratified by age, sex, diabetes, previous PCI, and previous myocardial infarction. Results were generally consistent with those from our main analyses.

### DISCUSSION

We used observational data from the SWEDEHEART registry to address questions beyond those the TASTE trial could answer. The process had 3 steps. First, we used the observational data to emulate a target trial similar to TASTE, which estimated the effect of thrombus aspiration on risk of death and myocardial infarction by 1 year. Second, we informally benchmarked the observational analysis against TASTE by concluding that the same clinical decision would be made using either study because both studies estimated no benefit of thrombus aspiration. Third, in the observational analysis we extended follow-up to 3 years to also estimate no benefit, and estimated no effects by 1 year in subpopulations defined by age, sex, diabetes, previous PCI, and previous myocardial infarction.

Unmeasured confounding is always a possibility in observational analyses. We were concerned that we could not adjust for thrombus burden, a predictor of myocardial infarction and death that affects the decision whether to administer thrombus aspiration, especially after TASTE when it was only used as a bail-out for those with high thrombus burden in Sweden (10, 19). Lack of adjustment for this variable might explain the increased 1-year risk of death in the subpopulations of women and those without diabetes. However, a sensitivity analysis that additionally adjusted for Killip class, which is correlated with thrombus burden, did not considerably change the estimates (Web Table 13; analysis restricted to the time period after TASTE as there was a high degree of missingness of Killip class earlier) (20). This

implies either that Killip class does not adequately capture thrombus burden or, more likely, that there is little residual confounding due to thrombus burden.

Even in the absence of unmeasured confounding, there may be differences between the randomized trial and the observational analysis with respect to: 1) study populations, 2) definition or measurement of interventions or outcomes, and 3) causal estimands. Because these differences may affect the estimates in different directions, it is logically possible that a partial canceling out of these impacts leads to an erroneous conclusion that benchmarking was successful. We now discuss each of these differences and consider their impact on our results.

Between-study differences in effect estimates will occur if the treatment effect varies across groups that are unequally represented in each study. Eligible individuals who would not agree to enroll in a randomized trial is one such group, because those who do not enroll in trials (about 39% of individuals with STEMI in TASTE) are generally sicker, with poorer prognosis. In TASTE, the 1-year risk of death was 5.3% in those who enrolled and were randomized to thrombus aspiration, and it was 16.4% in those who did not enroll and were given thrombus aspiration under routine practice. The inclusion of these individuals in the target trial emulation meant higher absolute risks than among those who enrolled in TASTE, but this did not seem to affect the risk ratio estimates because, among those not enrolled in TASTE, the risks were similar when comparing groups with and without thrombus aspiration (16.4% and 15.7% for death; 3.8% and 3.7% for myocardial infarction) (6).

Another reason that study populations may differ is that observational data may not be detailed enough to match the eligibility criteria of the index trial. In our application, fewer individuals were eligible for the observational analysis in the period before TASTE compared with after TASTE, possibly because data on the proportion of stenosis in the culprit artery were less complete in SWEDEHEART

**Table 4.** Estimated 1-Year Risks, Risk Differences, and Risk Ratios From an Observational Emulation of a Target Trial of Thrombus Aspiration Versus No Thrombus Aspiration, Stratified by Subpopulation, SWEDEHEART Registry, 2007–2016

Subpopulation and Outcome	Thrombus Aspiration <sup>a</sup>		No Thrombus Aspiration <sup>a</sup>		RD <sup>a</sup>	95% CI	RR	95% CI <sup>a</sup>
	Risk, %	95% CI	Risk, %	95% CI				
<i>Sex</i>								
Death								
Female	13.2	10.8, 15.7	10.4	9.3, 11.6	2.8	−0.1, 5.6	1.27	1.06, 1.51
Male	5.9	4.6, 7.2	6.1	5.5, 6.7	−0.2	−1.6, 1.2	0.96	0.81, 1.15
Myocardial infarction								
Female	4.4	2.3, 6.5	4.1	3.2, 4.9	0.3	−2.0, 2.6	1.08	0.75, 1.57
Male	3.7	2.6, 4.8	4.1	3.5, 4.7	−0.4	−1.7, 0.9	0.91	0.72, 1.15
<i>Age</i>								
Death								
≥65 years	12.0	10.1, 13.9	10.9	10.0, 11.8	1.1	−0.9, 3.1	1.10	0.97, 1.25
<65 years	2.7	1.4, 3.9	2.7	2.2, 3.3	−0.1	−1.4, 1.3	0.98	0.68, 1.41
Myocardial infarction								
≥65 years	4.2	2.8, 5.5	4.5	3.8, 5.1	−0.3	−1.8, 1.2	0.93	0.73, 1.19
<65 years	3.5	2.3, 4.8	3.5	2.9, 4.2	0.0	−1.4, 1.4	1.00	0.76, 1.31
<i>Diabetes</i>								
Death								
No	7.5	6.2, 8.8	6.4	5.9, 7.0	1.0	−0.4, 2.5	1.16	1.01, 1.34
Yes	9.9	6.2, 13.5	12.6	10.9, 14.4	−2.8	−6.5, 1.0	0.78	0.58, 1.05
Myocardial infarction								
No	3.8	2.8, 4.8	3.7	3.2, 4.2	0.1	−1.0, 1.3	1.03	0.84, 1.27
Yes	3.9	1.1, 6.7	6.1	4.7, 7.4	−2.2	−5.3, 0.9	0.64	0.36, 1.12
<i>Previous Percutaneous Coronary Intervention</i>								
Death								
No	8.0	6.7, 9.3	7.4	6.8, 8.0	0.6	−0.8, 2.0	1.08	0.95, 1.23
Yes	9.6	5.6, 13.5	6.7	5.0, 8.4	2.9	−1.4, 7.2	1.43	0.98, 2.07
Myocardial infarction								
No	3.6	2.5, 4.6	3.6	3.2, 4.1	−0.1	−1.2, 1.1	0.98	0.78, 1.22
Yes	6.8	3.3, 10.3	7.7	5.8, 9.6	−0.9	−4.9, 3.2	0.89	0.58, 1.35
<i>Previous Myocardial Infarction</i>								
Death								
No	7.5	6.3, 8.7	6.8	6.2, 7.4	0.7	−0.7, 2.1	1.10	0.96, 1.27
Yes	12.1	7.3, 16.8	11.0	9.3, 12.6	1.1	−3.9, 6.1	1.10	0.80, 1.52
Myocardial infarction								
No	3.8	2.7, 4.9	3.5	3.1, 4.0	0.3	−0.9, 1.4	1.08	0.87, 1.33
Yes	4.8	1.9, 7.7	7.6	6.0, 9.1	−2.7	−6.0, 0.6	0.64	0.40, 1.02

Abbreviations: CI, confidence interval; RD, risk difference, RR, risk ratio; SWEDEHEART, Swedish Web-Based System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies.

<sup>a</sup> Adjusted at baseline for: age, sex, hospital, diabetes, body mass index, smoking, hyperlipidemia, hypertension, previous infarction, previous percutaneous coronary intervention, previous coronary artery bypass graft, stenosis class, proportion stenosis, angiography finding, heart rate, systolic blood pressure, diastolic blood pressure, thrombolysis, and use of warfarin, aspirin, clopidogrel, prasugrel, heparin, low molecular weight heparin, bivalirudin, or glycoprotein 2b/3a inhibitors.

in the earlier period (meaning that fewer people could be evaluated for the minimum of 50% stenosis criterion). However, in a sensitivity analysis in which we did not use the minimum stenosis criterion to determine eligibility,

effect estimates were broadly similar to the main results (Web Table 7).

Between-study differences in effect estimates will also occur if the measurement of interventions or outcomes

varies between studies. However, this is unlikely to occur in our application because the randomized trial, TASTE, and the observational analysis were both embedded within the SWEDEHEART Registry, and the definition and measurement of the intervention, thrombus aspiration, and the outcomes, death and myocardial infarction, were captured using the same mechanism. Use of the SWEDEHEART Registry also means the health-care system was the same in both studies.

Differences in causal estimands may also lead to between-study differences in effect estimates. In randomized trials, the main estimand is often the intention-to-treat effect (i.e., the effect of assignment to treatment). However, when using observational data, information may be available only on treatment an individual actually received, not what they were assigned or prescribed. Then, for point interventions like thrombus aspiration, the observational analysis can only estimate the per-protocol effect—that is, the effect of receiving treatment (21). Appropriate benchmarking then necessitates reanalyzing the randomized trial data to estimate the per-protocol effect, which requires adjustment for before-randomization factors to account for confounding (22, 23). In our application, it is unlikely that differences in estimands affected the comparability of the estimates because adherence to the assigned treatment was very high (94%) in TASTE, and in fact, an unadjusted comparison restricted to the adherers resulted in a hazard ratio (0.95) very similar to that of the intention-to-treat analysis (0.94) (6).

Informal benchmarking at 1 year increases confidence in the reliability of observational inferences at 3 years and within subpopulations. Because increasing follow-up increases the possibility of selection bias due to loss to follow-up, observational analyses generally require longitudinal data on joint predictors of loss to follow-up and the outcome interest. In our study, however, loss to follow-up is a minor concern because <0.5% of individuals emigrate each year (12). We also cannot think of baseline confounders that introduce bias only after 1 year. Additionally, our main analysis relies on the assumption that the measured covariates are approximately sufficient to adjust for confounding when estimating the effect in the entire study population. That is, we assume the magnitude of unmeasured confounding is, on average, small across all subgroups defined by the measured covariates. However, the magnitude of unmeasured confounding might be greater (or smaller) in certain subgroups and thus some subgroup-specific effect estimates may be more (or less) biased than the effect estimates in the entire study population.

In summary, we carried out an observational analysis that emulates a target trial, informally benchmarked its results with those from an index randomized trial, and used the observational analyses to draw causal inferences over a longer follow-up duration and within subpopulations. The agreement between estimates from TASTE and our emulated target trial suggests that the observational data can deliver approximately valid estimates of treatment effects. This example shows how high-quality observational data can complement results from randomized trials and provide additional evidence to support clinical decision making.

## ACKNOWLEDGMENTS

Author affiliations: Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden (Anthony A. Matthews, Maria Feychting, Anita Berglund, Miguel A. Hernán); CAUSALab, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States (Issa J. Dahabreh, Miguel A. Hernán); Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States (Issa J. Dahabreh, Miguel A. Hernán); Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States (Issa J. Dahabreh, Miguel A. Hernán); Department of Cardiology, Faculty of Health, Örebro University Hospital, Örebro, Sweden (Ole Fröbert); Department of Clinical Medicine, Aarhus University Health, Aarhus, Denmark (Ole Fröbert); Department of Clinical Pharmacology, Aarhus University Hospital, Aarhus, Denmark (Ole Fröbert); Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark (Ole Fröbert); Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden (Bertil Lindahl, Stefan James); and Department of Clinical Sciences, Danderyd University Hospital-Karolinska Institutet, Stockholm, Sweden (Tomas Jernberg).

This work was supported by a grant from the Swedish Research Council (2018-03028). A.A.M. received funds from the Strategic Research Program in Epidemiology at Karolinska Institutet, Forte (2020-00029), and Swedish Research Council (2021-02236). I.J.D. received funds from the Patient-Centered Outcomes Research Institute (PCORI; ME-1502-27794) during the conduct of the study.

Pseudonymized personal data were obtained from national Swedish Registry holders after ethical approval and secrecy assessment. According to Swedish laws and regulations, sensitive personal data can only be made available for researchers who fulfill legal requirements for access to personal sensitive data. All analysis code is available at: <https://github.com/tonymatthews/taste>.

This study was approved by the Regional Ethical Review Board in Stockholm (2012/60-31/2).

The views expressed in this article are those of the authors and do not reflect those of the funders.

O.F. reports grants from Sanofi Pasteur during the conduct of the study; M.A.H. reports personal fees from Cytel and ProPublica during the conduct of the study. The other authors report no conflicts.

## REFERENCES

- Collins R, Bowman L, Landray M, et al. The magic of randomization versus the myth of real-world evidence. *N Engl J Med*. 2020;382(7):674–678.
- Gerstein HC, McMurray J, Holman RR. Real-world studies no substitute for RCTs in establishing efficacy. *Lancet*. 2019; 393(10168):210–211.
- VanderWeele TJ. Principles of confounder selection. *Eur J Epidemiol*. 2019;34(3):211–219.

4. Dahabreh IJ, Robins JM, Hernán MA. Benchmarking observational methods by comparing randomized trials and their emulations. *Epidemiology*. 2020;31(5):614–619.
5. Fröbert O, Lagerqvist B, Olivecrona GK, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. *N Engl J Med*. 2013;369(17):1587–1597.
6. Lagerqvist B, Fröbert O, Olivecrona GK, et al. Outcomes 1 year after thrombus aspiration for myocardial infarction. *N Engl J Med*. 2014;371(12):1111–1120.
7. Sardella G, Mancone M, Bucciarelli-Ducci C, et al. Thrombus aspiration during primary percutaneous coronary intervention improves myocardial reperfusion and reduces infarct size: the EXPIRA (Thrombectomy With Export Catheter in Infarct-Related Artery During Primary Percutaneous Coronary Intervention) prospective, randomized trial. *J Am Coll Cardiol*. 2009;53(4):309–315.
8. Svilaas T, Vlaar PJ, van der Horst IC, et al. Thrombus aspiration during primary percutaneous coronary intervention. *N Engl J Med*. 2008;358(6):557–567.
9. Ikari Y, Sakurada M, Kozuma K, et al. Upfront thrombus aspiration in primary coronary intervention for patients with ST-segment elevation acute myocardial infarction: report of the VAMPIRE (VAcuum asPiration thrombus REmoval) trial. *J Am Coll Cardiol Interv*. 2008;1(4):424–431.
10. Buccheri S, Sarno G, Fröbert O, et al. Assessing the nationwide impact of a registry-based randomized clinical trial on cardiovascular practice. *Circ Cardiovasc Interv*. 2019;12(3):e007381.
11. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758–764.
12. Statistics Sweden. Emigration from Sweden 2020 and 2019 by sex and top 20 countries of birth 2020. <http://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-composition/population-statistics/pong/tables-and-graphs/yearly-statistics--the-whole-country/emigration-from-sweden-2017-and-2018-by-sex-and-top-20-countries-of-birth/>. (Accessed July 16, 2021).
13. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, et al. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med*. 2020;39(8):1199–1236.
14. Hernán MA, Robins JM. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC; 2020.
15. Jernberg T, Attebring MF, Hambraeus K, et al. The Swedish Web-system for Enhancement and Development of Evidence-based care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART). *Heart*. 2010;96(20):1617–1621.
16. Brooke HL, Talbäck M, Hörnblad J, et al. The Swedish Cause of Death Register. *Eur J Epidemiol*. 2017;32(9):765–773.
17. Ludvigsson JF, Andersson E, Ekblom A, et al. External review and validation of the Swedish National Inpatient Register. *BMC Public Health*. 2011;11(1):450.
18. Wettermark B, Hammar N, Fored CM, et al. The new Swedish Prescribed Drug Register—opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf*. 2007;16(7):726–735.
19. Sianos G, Papafaklis MI, Daemen J, et al. Angiographic stent thrombosis after routine use of drug-eluting stents in ST-segment elevation myocardial infarction: the importance of thrombus burden. *J Am Coll Cardiol*. 2007;50(7):573–583.
20. Tanboga IH, Topcu S, Aksakal E, et al. Determinants of angiographic thrombus burden in patients with ST-elevation myocardial infarction. *Eur Heart J*. 2014;20(7):716–722.
21. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med*. 2017;377(14):1391–1398.
22. Lodi S, Phillips A, Logan R, et al. Comparative effectiveness of immediate antiretroviral therapy versus CD4-based initiation in HIV-positive individuals in high-income countries: observational cohort study. *Lancet HIV*. 2015;2(8):e335–e343.
23. Lodi S, Phillips A, Lundgren J, et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. *Am J Epidemiol*. 2019;188(8):1569–1577.