

# Big data: Large scale retrospective study - Bias control by matching and weighting

양 필 성

차의과대학 분당 차병원, 심장내과

# 연구 방법들

- Prospective random study

---

- Multi- or single center prospective registry

- PCI registry, HF registry, RFCA registry

- Retrospective study

- 병원 EMR을 이용한 single center retrospective study

- 보건의료 빅데이터, 청구데이터를 이용한 large scale retrospective study

# 보건의료 빅데이터를 이용하여 가능한 연구

- **Epidemiologic study**
  - incidence, prevalence, cost effectiveness, Etc.
- **Risk factor Study**
  - Risk factors for disease occurrence
  - Risk factors for disease prognosis
- **Treatment effect study**

# Treatment effect study (치료효과 연구)

- Prospective random control study - Golden standard
    - 치료를 받을지 말지에 대한 random
    - Randomization을 통해 치료 여부 결정에 대한 bias control
    - 문제: 시간, 비용, 시스템
- 

- 보건 의료 빅데이터를 이용한 observational clinical research
  - RCT를 대체할 수 있을까? → No
  - RCT를 흉내 낼 수 있을까? RCT와 근접한 결과를 얻을 수 있을까?  
치료 여부 결정에 대한 bias control을 어떻게 할까?

# Mimics of Randomized Clinical Trial

## Randomized Clinical Trial

## Observational clinical research

Study design

- Initiator of treatment
- Active-comparator design
- New-users design, Etc.

Pre-design for potential bias minimizing

Target population

Enrollment with inclusion / exclusion criteria

Inclusion / exclusion criteria

target population

Propensity methods

- Propensity matching
- propensity score weighting
  - ✓ IPTW  $\pm$  trimming
  - ✓ Overlap Weighting

Equipoise

Covariate balance

Randomization

Precision

Outcome analysis

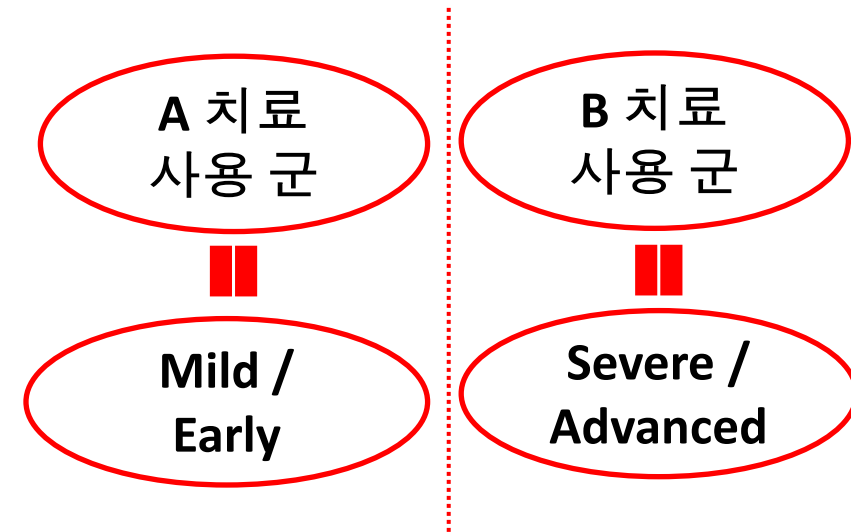
Doubly robust

Falsification outcome

Outcome analysis

# Potential bias (1) : Uncontrollable initiator of treatment

새로 나온 치료 방법 A, 이전 치료 방법 B



국내 청구 데이터로는 해당 질환의 severity를 구분할 수 없다면?

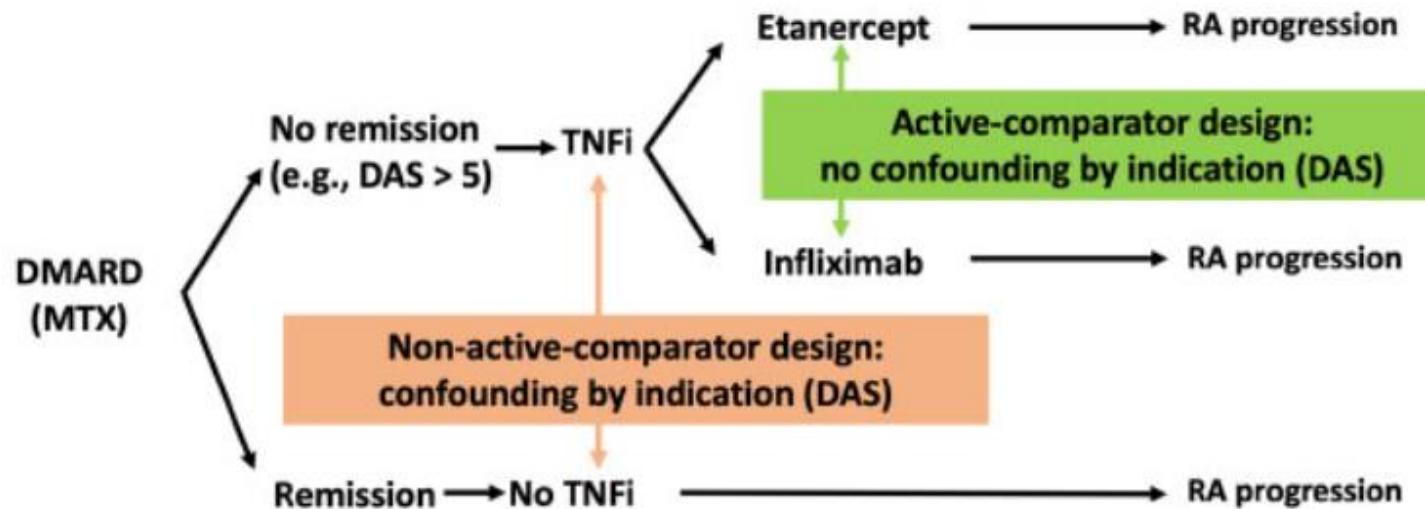
- 국내 청구데이터를 이용하여 해당 질환 환자를 A 치료 군과 B 치료군을 나누고 조사 가능한 모든 변수에 대해서 PS matching 시행한 후 outcome 차이를 비교하였다.
- A 치료 군이 B 치료군 보다 outcome 이 더 좋았다.
- 해당 질환 환자에서는 A 치료를 사용하는 것이 좋다 ???

두 치료의 initiator를 data에서 확인 및 처리 할 수 없다면 두 치료의 비교 연구는 해서는 안된다.

# Potential bias (2) : Try to make “Active comparator design”

- **Active comparator design**

- ✓ Identifying initiators of the treatment of interest and initiators of an alternative treatment for the same indication.
- ✓ Restricting study population to patients with the same indication for treatment and without contraindications



치료 받을 조건이 전혀 다른 두 치료의 비교 연구 (non-active comparator design)는 하지 말자.

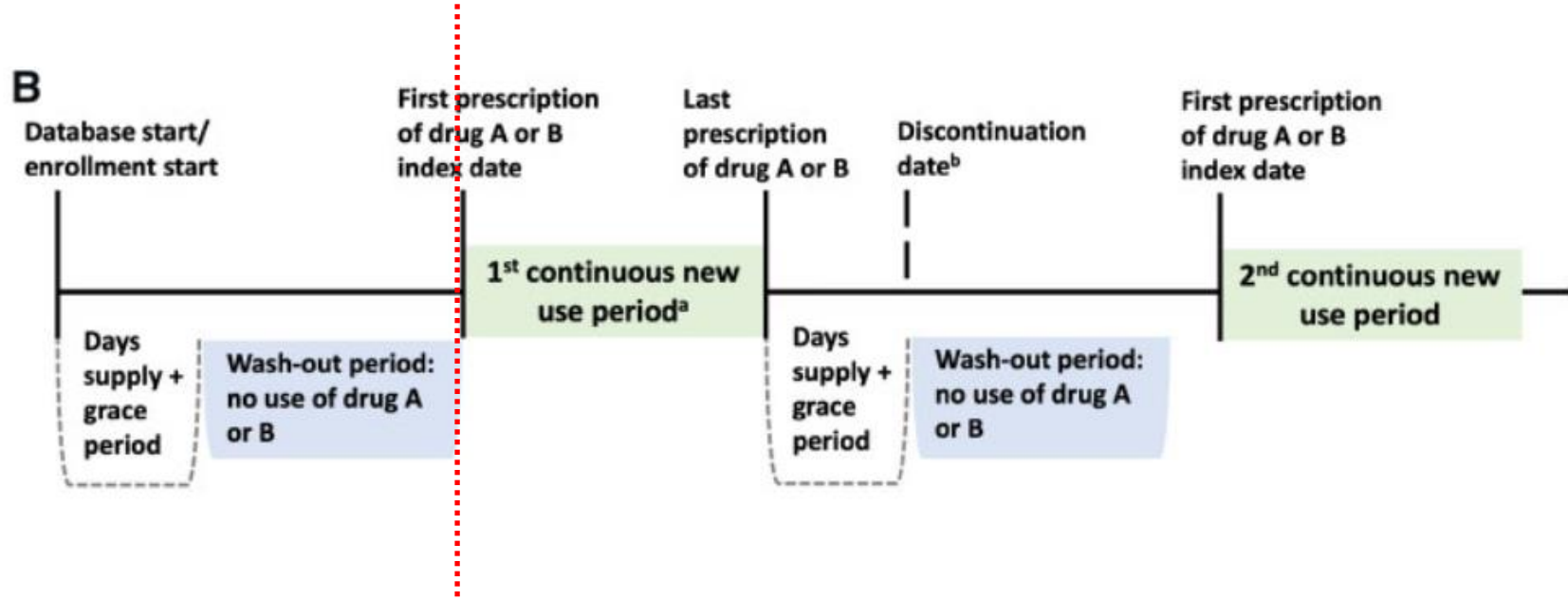
*Rheumatology 2020;59:1425*

# Potential bias (3) : Prevalent user related bias, Immortal time bias

## Try to make “New-user design”

- **New-user design**

- Identifies all patients initiating specific treatment in a defined population after a certain length of time free of the treatment (washout period)
- Solves issues of **prevalent user related bias** and **immortal time bias**
- New users do not necessarily need to be drug naïve: they are only required to be naïve for the treatments compared during the wash-out period (eg, one year)



Database 시작을 index  
date으로 하지 말고  
치료의 시작일을 index  
date으로 하자.



# Potential bias (4) : Considering treatment changes after initiation

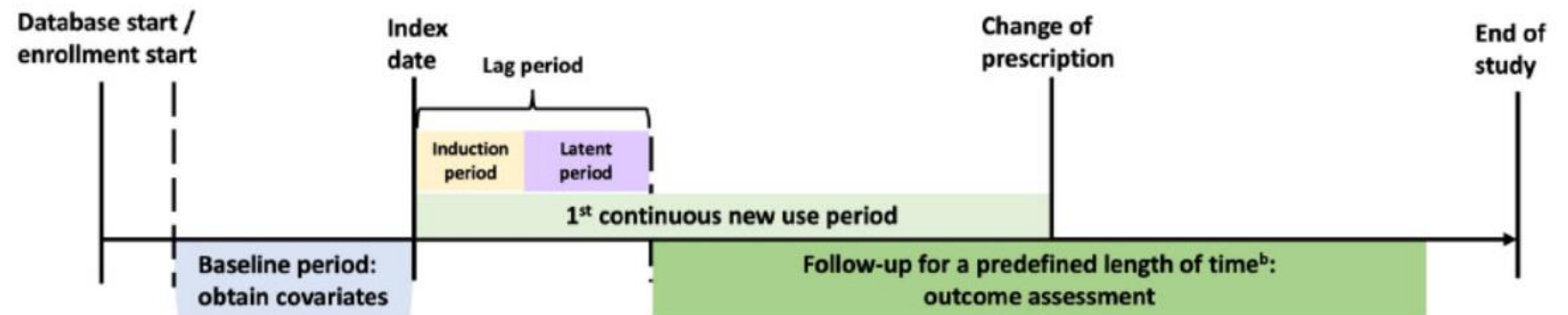
## As-treated analysis $\approx$ per-protocol analysis in RCT

- Advantage: It considers periods at actual risk due to the treatment.
- Disadvantage: Censoring patients stopping the initial treatment can introduce selection bias because changing treatments is usually due to a lack of effectiveness or side effects that are very likely to affect outcome risk.



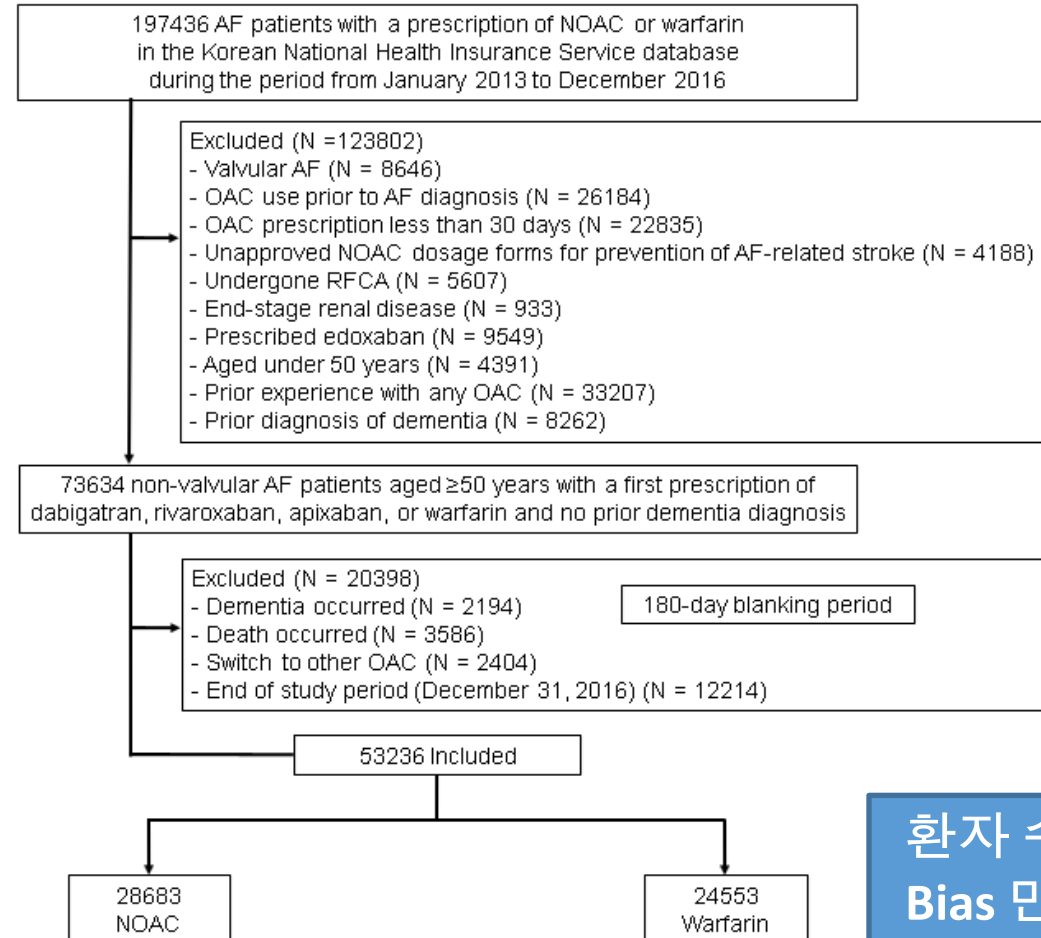
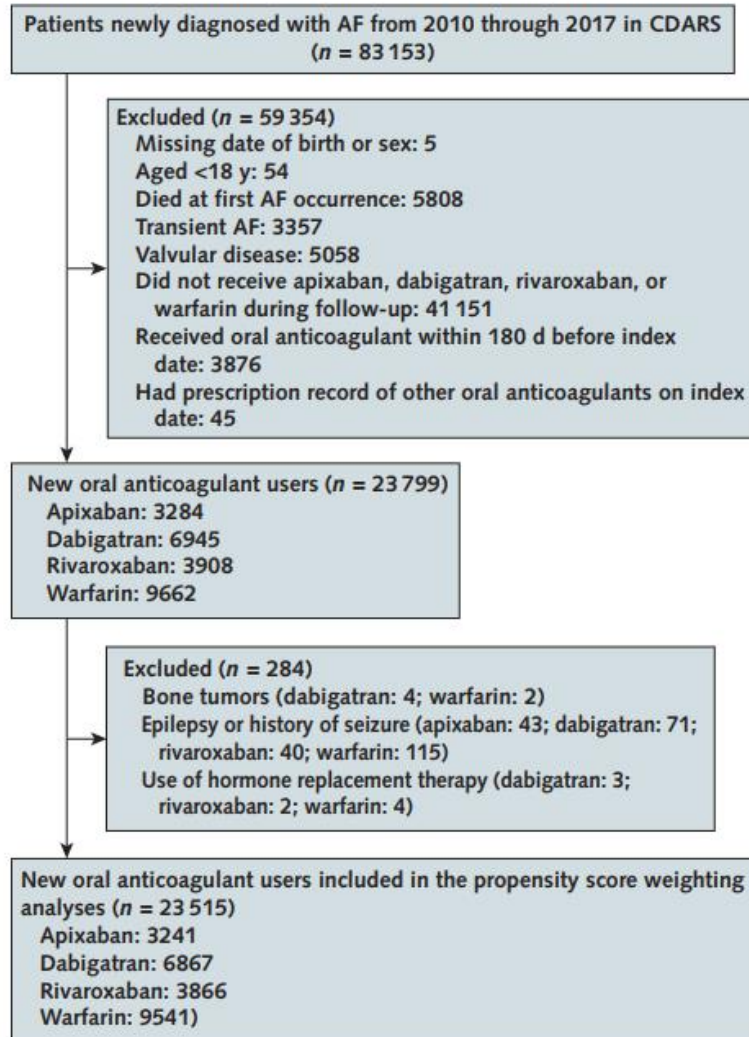
## Initial-treatment analysis $\approx$ intent-to-treat analysis in RCT

- Advantage: it protects against selection bias introduced by conditioning on continuous treatment.
- Disadvantage: It introduces bias due to increasing misclassification of exposure of treatment.



*Rheumatology 2020;59:1425*

# Target population: exclusion



환자 수에 연연하지 말고  
Bias 만들 수 있는 환자를  
최대한 exclusion.

# Propensity methods

- **Propensity methods**



- Large-scale Propensity score
- High-dimensional propensity score

## PS matching vs. PS weighting

- **Propensity score matching**
- **propensity score weighting**
  - ✓ Inverse probability of treatment weighting (IPTW)  $\pm$  trimming
  - ✓ IPTW with stabilized weights
  - ✓ Entropy balance weighting
  - ✓ **Overlap Weighting**

# PS matching vs. PS weighting

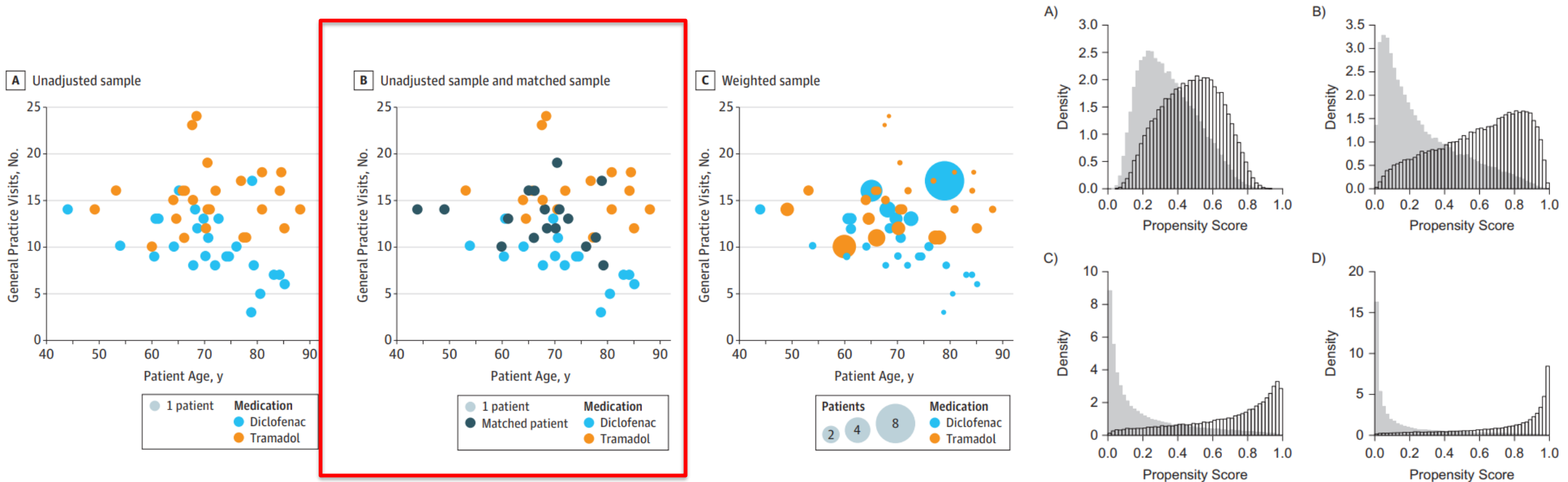
## CENTRAL ILLUSTRATION Comparison of Propensity Score Methods and Covariate Adjustment

Primary study analysis method	 Pros	 Cons
Traditional covariate adjustment	<ul style="list-style-type: none"> <li>• Performed well</li> <li>• Provides prognostic model for outcome of interest</li> </ul>	<ul style="list-style-type: none"> <li>• May not be suitable with many covariates in smaller studies</li> </ul>
Propensity score (PS) stratification	<ul style="list-style-type: none"> <li>• Retains data from all study participants</li> <li>• Opportunity to explore interactions between treatment and PS on outcome risk</li> <li>• Provides effect estimates for every stratum</li> </ul>	<ul style="list-style-type: none"> <li>• Performs less well in datasets with few outcomes, particularly when the number of strata is large</li> <li>• May not account for strong confounding</li> </ul>
PS matching	<ul style="list-style-type: none"> <li>• Reliable; provides excellent covariate balance in most circumstances</li> <li>• Simple to analyze, present and interpret</li> </ul>	<ul style="list-style-type: none"> <li>• Some patients are unmatched leading to information excluded from the analysis</li> <li>• Less precise</li> </ul>
PS inverse probability weighting	<ul style="list-style-type: none"> <li>• Retains data from all study participants</li> <li>• Easy to implement</li> <li>• Creates a pseudo population with perfect covariate balance</li> </ul>	<ul style="list-style-type: none"> <li>• Can be unstable when extreme weights occur</li> </ul>
PS covariate adjustment (use of PS as a covariate)	<ul style="list-style-type: none"> <li>• Performed well</li> </ul>	<ul style="list-style-type: none"> <li>• Adding the PS as an additional covariate produced results very similar (and not necessarily superior) to traditional covariate adjustment</li> </ul>

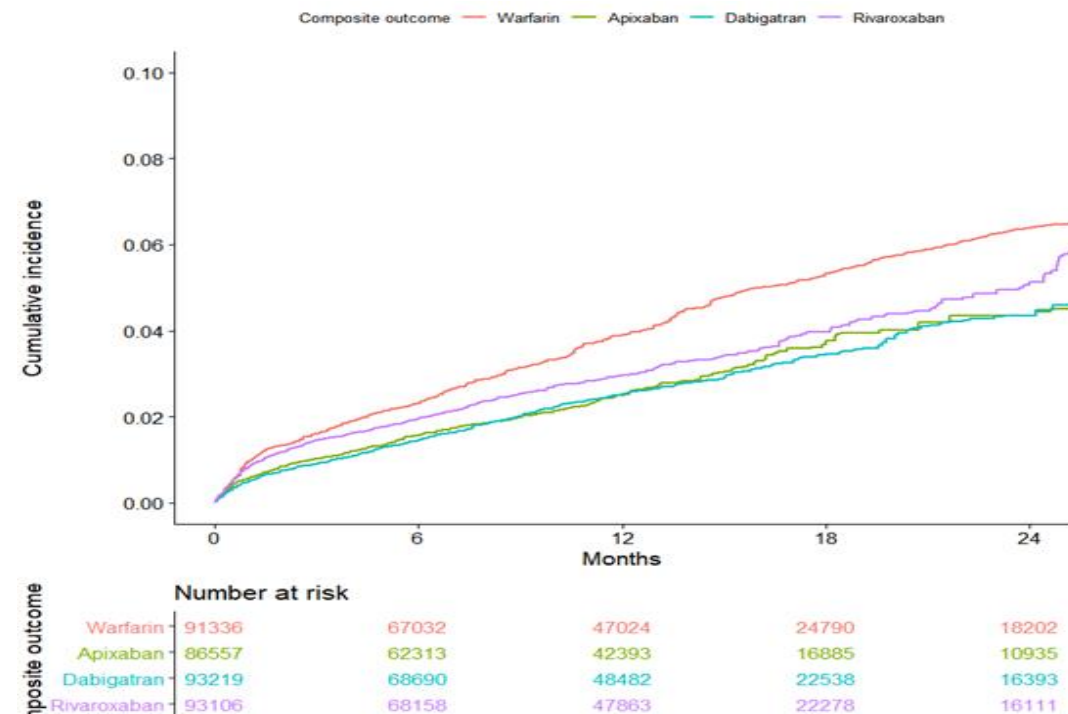
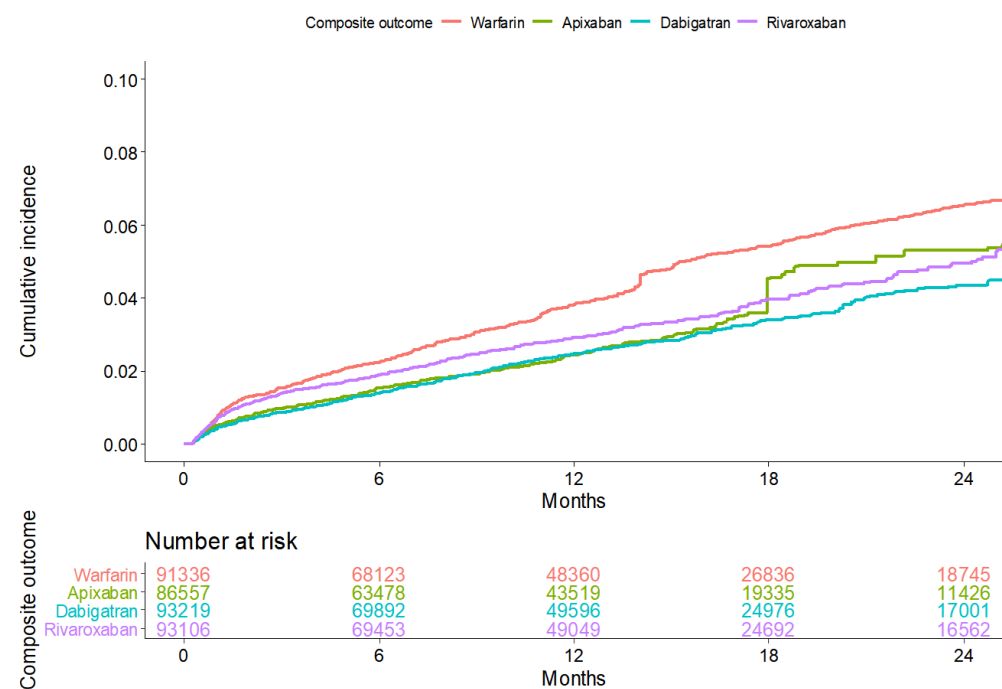
*J Am Coll Cardiol.* 2017;69(3):345–57.

# Target population: Propensity Score matching can distort Target Populations

- **Target population in RCT:**
  - depends on inclusion, exclusion criteria, and patients recruitment
- **Target population in observational study:**
  - depends on inclusion, exclusion criteria, and **choice of propensity score method.**



# PS weighting can be unstable when extreme weights occur

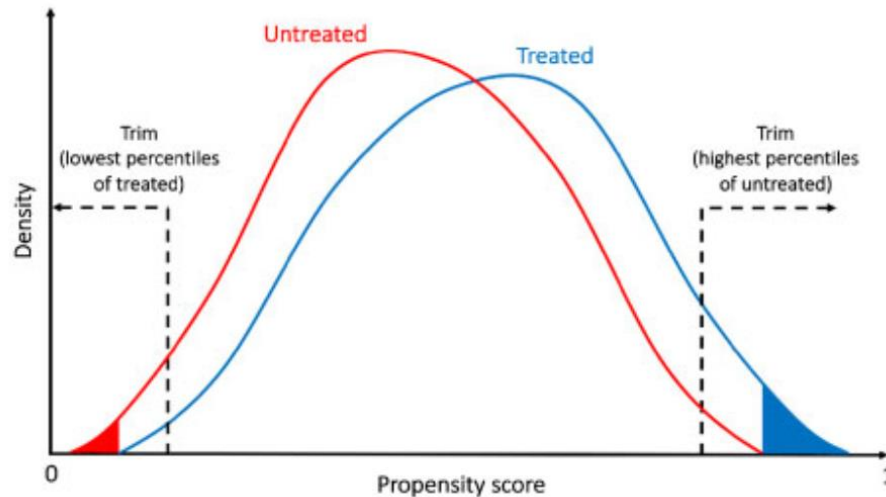


PERSON_ID	W	duration	event
12455255	186.8876	17.45753	1
44452457	98.00849	13.61096	1
48044454	33.97705	2.59726	1
11983271	28.12081	12.92055	1



# Try to overcome limitations of PS methods : trimming

- **propensity score weighting**
  - ✓ Inverse probability of treatment weighting (IPTW)  $\pm$  **trimming**
  - ✓ IPTW with **stabilized weights**
  - ✓ Entropy balance weighting
  - ✓ **Overlap Weighting**



Trimming both tails of the overlapping propensity score distribution will remove some of the patients treated contrary to prediction and thus tend to reduce unmeasured confounding.

Schematic of asymmetric propensity score trimming

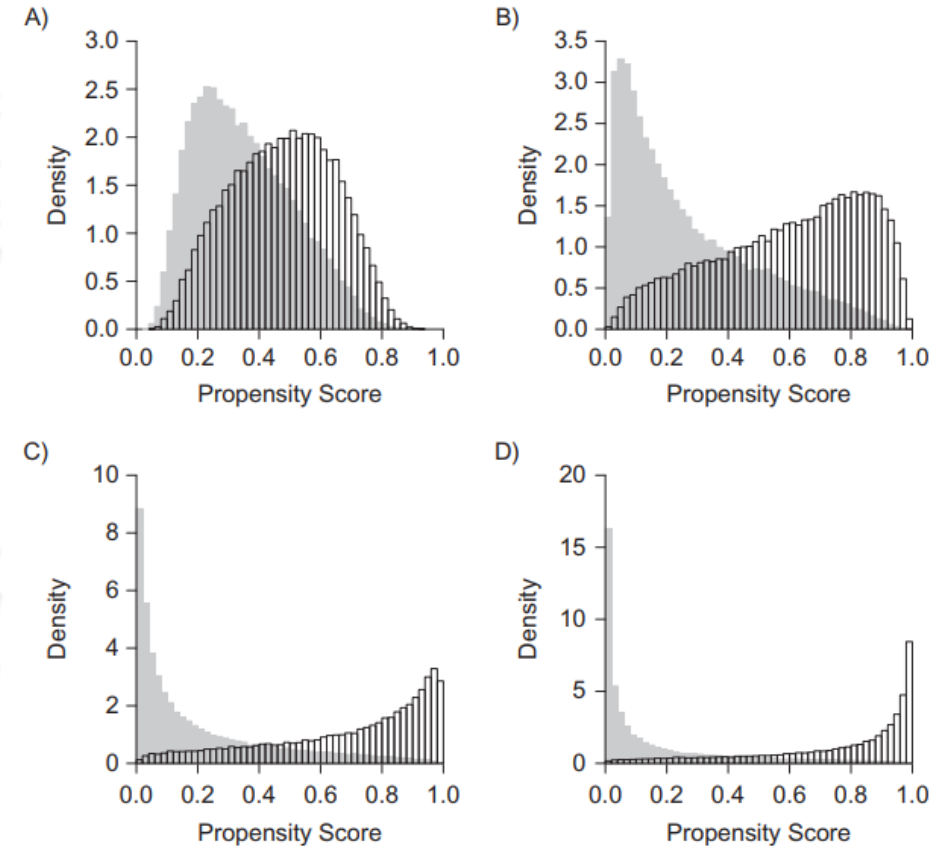
# Try to overcome limitations of PS methods : Overlap Weighting

## Overlap PS weighting

The overlap weight was calculated as  $1 - \text{propensity score for the treatment patients (treatment group)}$ , and  $\text{the propensity score for the control patients (control group)}$ . This weight is used to calculate the average treatment effect for the overlap population. This approach minimizes the asymptotic variance of the treatment effect, while also possessing a desirable exact balance property.

## Inverse probability of treatment (IPT) propensity-score weighting

We assigned patients who underwent “treatment A” a weight of  $1/(\text{propensity score})$  (treatment group) and those who underwent “treatment B” a weight of  $1/(1 - \text{propensity score})$  (control group). To reduce the variability in the inverse probability of treatment– weighted models, we used stabilized weights.

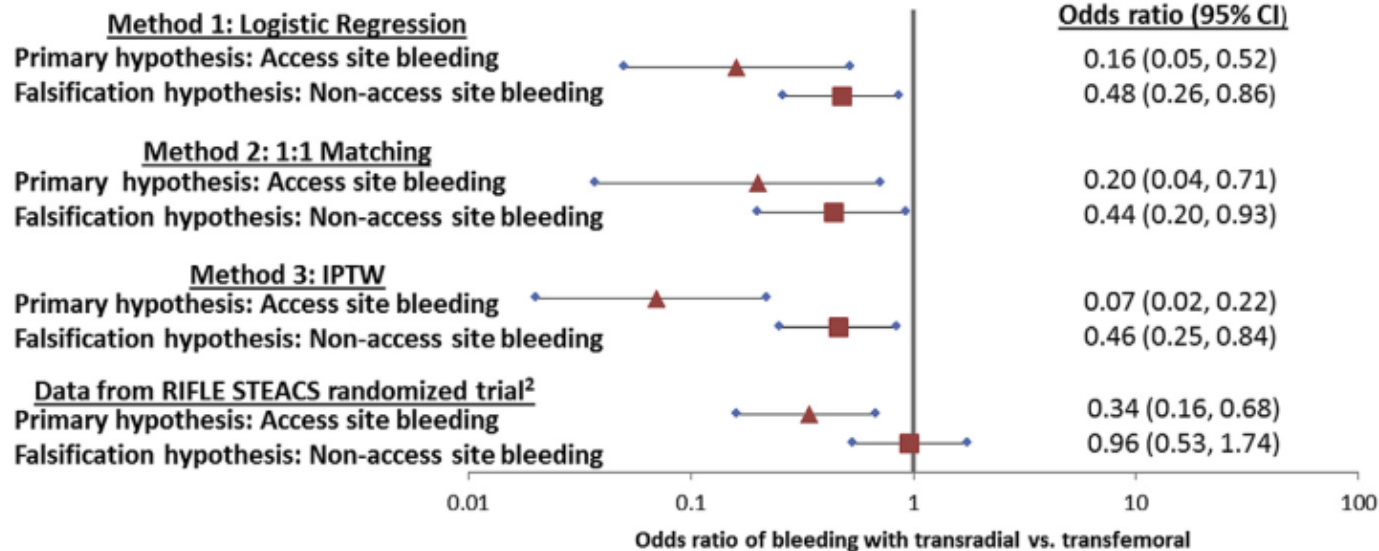




# Falsification outcome

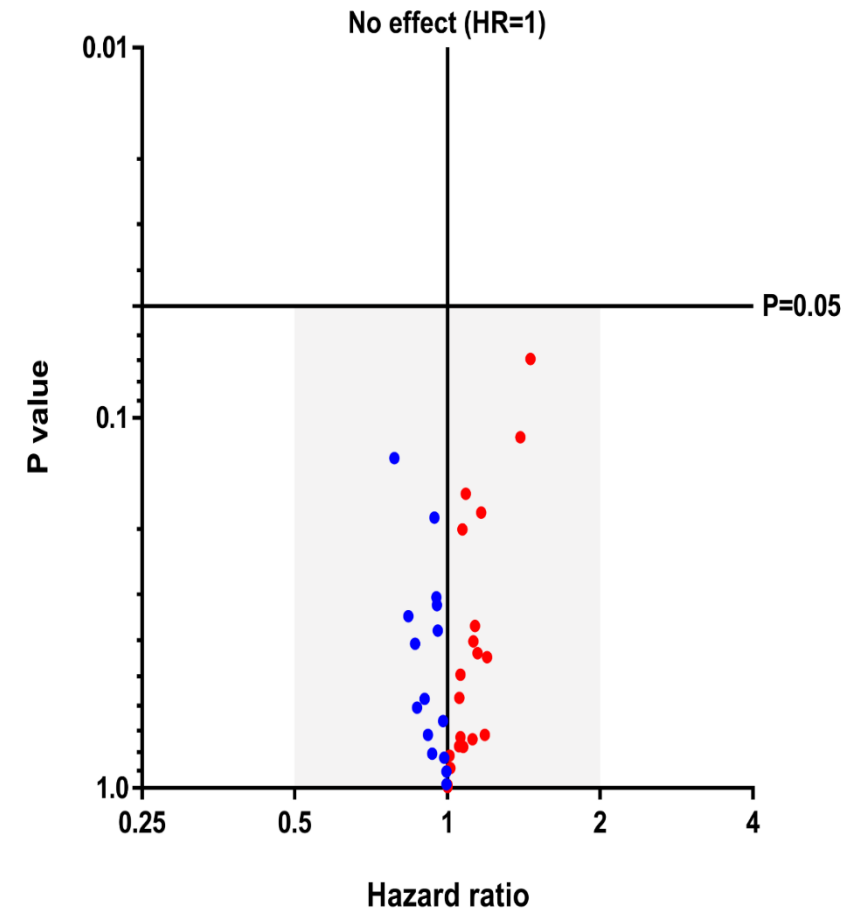
## Comparison of Transradial Versus Transfemoral Percutaneous Coronary Intervention in Routine Practice

Evidence for the Importance of “Falsification Hypotheses” in Observational Studies of Comparative Effectiveness



*JACC. Volume 62, Issue 22, December 2013*

## Risk of 35 falsification endpoints for propensity-matched patients



# 예시 논문



European Society  
of Cardiology

European Heart Journal (2019) 40, 1257–1264  
doi:10.1093/eurheartj/ehz085

**FASTTRACK CLINICAL RESEARCH**

Atrial fibrillation

## Atrial fibrillation ablation in practice: assessing CABANA generalizability

Peter A. Noseworthy<sup>1,2\*</sup>, Bernard J. Gersh<sup>2</sup>, David M. Kent<sup>3,4</sup>, Jonathan P. Piccini<sup>5</sup>, Douglas L. Packer<sup>2</sup>, Nilay D. Shah<sup>1,6,7</sup>, and Xiaoxi Yao<sup>1,6</sup>

Propensity score overlap weighting was used to account for the differences in baseline characteristics between patients who underwent ablation and those who were treated with medical therapy alone. A propensity score, the probability of undergoing ablation, was estimated using logistic regression based on socio-demographics, medical history, concurrent medication use, previous treatment with AADs or rate control drugs, the year of the index date, and the length of baseline period (variables in [Supplementary material online, Table S3](#)). The distribution of propensity scores is shown in [Supplementary material online, Figure S2](#). The overlap weight was calculated as  $1 - \text{propensity score for the ablated patients}$ , and the propensity score for the drug-treated patients. This weight is used to calculate the average treatment effect for the overlap population. This approach minimizes the asymptotic variance of the treatment effect, while also possessing a desirable exact balance property.<sup>18</sup>

## Sensitivity analyses

First, we performed subgroup analyses for the primary outcome stratified by age, sex, race, CHA<sub>2</sub>DS<sub>2</sub>-VASc, hypertension with left ventricular hypertrophy, heart failure, cardiomyopathy, sleep apnoea, and prior thromboembolism. Second, one-to-one propensity score matching was used instead of propensity score weighting. Third, we conducted a stratified analysis based on whether the drug-treated patients were treated with AADs or with rate control drugs only. Fourth, we conducted a stratified analysis based on the medication adherence in the drug-treated patients. Fifth, we explored the impact of potential protocol deviation in the trial on the treatment effect, assuming 30% of the medical therapy cohort crossed over to the ablation cohort and 10% of the ablation cohort did not receive the procedure.<sup>13</sup>

We assessed residual confounding using two methods. First, we tested three falsification endpoints that are unlikely to be a result of ablation but might be related to unmeasured confounders such as frailty: chronic obstructive pulmonary disease, pneumonia, and fracture.<sup>21</sup> Second, we used the method outlined by Lin *et al.*<sup>22</sup> to assess whether the observed difference could be fully explained by an unmeasured confounder.

A two-sided *P*-value of  $<0.05$  was considered to indicate statistical significance. No adjustment for multiple testing was performed. All the analyses except those related to the primary outcome were considered to be exploratory.

# 방법론을 이해하는데 읽어 보면 좋은 저널들

Review > J Am Coll Cardiol. 2017 Jan 24;69(3):345-357. doi: 10.1016/j.jacc.2016.10.060.

## Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies

Markus C Elze<sup>1</sup>, John Gregson<sup>2</sup>, Usman Baber<sup>3</sup>, Elizabeth Williamson<sup>2</sup>, Samantha Sartori<sup>3</sup>, Roxana Mehran<sup>3</sup>, Melissa Nichols<sup>4</sup>, Gregg W Stone<sup>4</sup>, Stuart J Pocock<sup>5</sup>

JAMA Guide to Statistics and Methods

January 10, 2020

## Using Propensity Score Methods to Create Target Populations in Observational Clinical Research

Laine Thomas, PhD<sup>1</sup>; Fan Li, PhD<sup>2</sup>; Michael Pencina, PhD<sup>3</sup>

> Author Affiliations | Article Information

JAMA. 2020;323(5):466-467. doi:10.1001/jama.2019.21558

> JAMA. 2013 Jan 16;309(3):241-2. doi: 10.1001/jama.2012.96867.

## Prespecified falsification end points: can they validate true observational associations?

Vinay Prasad<sup>1</sup>, Anupam B Jena

RHEUMATOLOGY

Rheumatology 2020;59:14-25  
doi:10.1093/rheumatology/kez320

## Real World Data: special section

### Methodological considerations when analysing and interpreting real-world data

Til Stürmer<sup>1</sup>, Tiansheng Wang<sup>1</sup>, Yvonne M. Golightly<sup>1,2,3,4</sup>, Alex Keil<sup>1</sup>, Jennifer L. Lund<sup>1</sup> and Michele Jonsson Funk<sup>1</sup>

> JAMA. 2020 May 5. doi: 10.1001/jama.2020.7819. Online ahead of print.

## Overlap Weighting: A Propensity Score Method That Mimics Attributes of a Randomized Clinical Trial

Laine E Thomas<sup>1 2</sup>, Fan Li<sup>2 3</sup>, Michael J Pencina<sup>1 2</sup>

- 경청해 주셔서 감사합니다.