

Challenges and Opportunities in Observational Studies

Adin-Cristian Andrei, PhD
Associate Professor
Northwestern University
a-andrei@northwestern.edu

Multidisciplinary Clinical Research Center
Clinical and Translational Research Incubator Seminar

20 February, 2018

Performance Outcome Measures (POMs) in Scientific Research

- ▶ POMs are **formal** tools that permit valid comparisons across study groups, healthcare providers, devices, etc.
- ▶ There are general POMs, such as hospital length of stay, 30-day mortality, 30-day hospital readmission, etc.
- ▶ Each disease area has its own set of relevant POMs
- ▶ For example, in cardiology and cardiac surgery, there is interest in reoperation-free survival

Assessment of POMs

- ▶ In many disease areas, it is a high priority to develop evidence-based Clinical Practice Guidelines
- ▶ ACC/AHA in cardiology: high blood pressure in adults (2017), valvular heart disease (2017), atrial fibrillation (2014)
- ▶ American Cancer Society: breast cancer screening (2017), prostate cancer early detection (2016)
- ▶ Society for Vascular Surgery: management of diabetic foot (2016)
- ▶ Supporting evidence behind guidelines ranges from expert opinion (little data) to level 1 (large RCTs)

POMs in Rheumatology (Suter et al., Arthritis Care & Research, 2016)

American College of Rheumatology White Paper on Performance Outcome Measures in Rheumatology

LISA G. SUTER,¹ CLAIRE E. BARBER,² JEPH HERRIN,³ AMYE LEONG,⁴ ELENA LOSINA,⁵ AMY MILLER,⁶ ERIC NEWMAN,⁷ MARK ROBBINS,⁸ HEATHER TORY,⁹ AND JINOOS YAZDANY¹⁰

- ▶ **Data source:** detailed, reproducible information
- ▶ **Measure cohort (denominator):** inclusion/exclusion criteria
- ▶ **Reporting period and at-risk period:** timeframe
- ▶ **Measure outcome:** clear definition, feasible, meaningful
- ▶ **Outcome attribution:** causality worth discussing
- ▶ **Risk adjustment:** critical in most studies
- ▶ **Reliability/validity testing:** valid, reproducible data and conclusions

Assessment of POMs

- ▶ Every aspect outlined is important and adds scientific rigor
- ▶ Using examples, I will illustrate aspects relevant for most of these areas
- ▶ In nearly every disease area there are untapped opportunities for utilizing novel statistical methodologies
- ▶ This may lead to new insights and/or strengthen scientific evidence
- ▶ At the core of **Data source** stands **study design** (randomized or observational)

Definitions for An Observational Study

- ▶ *"... an empiric investigation [in which]...the objective is to elucidate **cause-and-effect relationships**...[in which] it is not feasible to use controlled experimentation, in the sense of being able to impose the procedures or treatments whose effects it is desired to discover, or to assign subjects at random to different procedures"* – William Cochran (JRSS A, 1965)

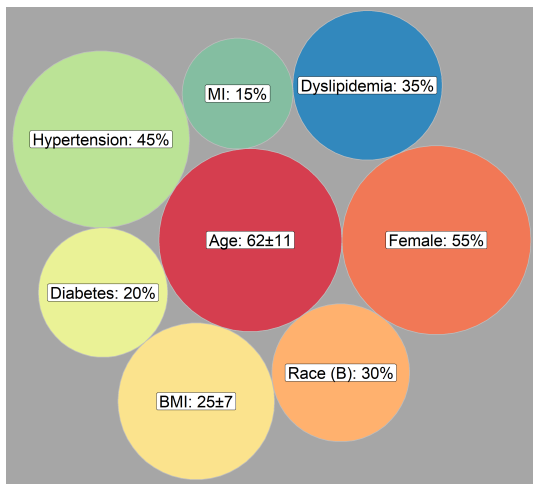
Definitions for An Observational Study

- ▶ *"An observational study is an empiric investigation of **effects caused by treatments** when randomized experimentation is unethical or infeasible"* – Paul Rosenbaum (2010)
- ▶ *"... an observational study draws inferences from a sample to a population where the independent variable is not under the control of the researcher"* – Wikipedia (2015)

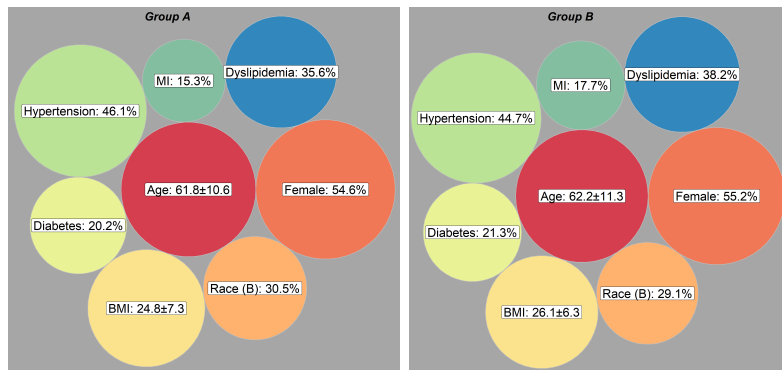
Observational Studies Are Ubiquitous

- ▶ Observational studies are increasingly common and complex
- ▶ Cardiology and Rheumatology: cohort studies, registries
- ▶ Health services: symptom management studies, survey data
- ▶ Pharmaceutical: post-approval studies
- ▶ Biotechnology: medical devices, health economics, post-marketing
- ▶ Online advertising: user behaviour and satisfaction

Draw Groups A and B From A Hypothetical Population of Interest



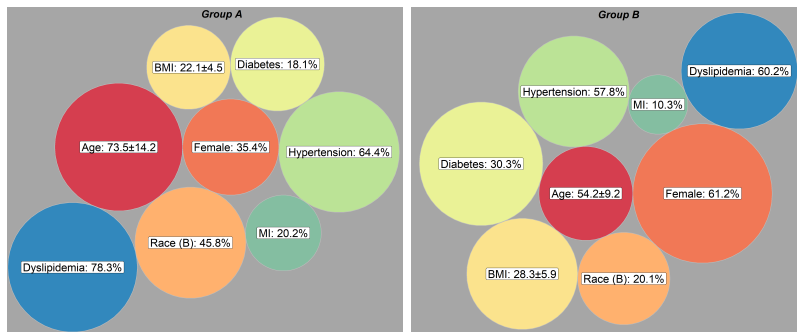
Data Anticipated In a Randomized Experiment



Randomization

- ▶ Randomized studies are considered by many the **gold** standard for inferring causality
- ▶ Not always feasible for reasons ranging from cost to ethics
- ▶ After randomization, if large enough, groups A and B are representative of the population of origin
- ▶ Differences in outcomes are then likely associated with group membership (treatment), not confounders!

Data Available In an Observational Study



Randomized versus Observational Studies

- ▶ Randomized experiments and observational studies are not worlds apart!
- ▶ On the contrary, they are part of the same continuum of study designs (Imbens and Rubin, 2015)
- ▶ What separates, but also unites them, is the group assignment **mechanism**
- ▶ In a randomized study, that mechanism could simply be a coin toss, hence no participant information involved
- ▶ In an observational study, participant characteristics play a **key** role in the group assignment!
- ▶ Other factors may (and likely will) be involved: environment characteristics, physician's experience and preferences, etc.

Further Insight Into Observational Studies

- ▶ It is self-evident that in any study, we only have access to **observable** information!
- ▶ Some relevant information may not be collected due to many reasons: cost, lack of awareness, non-feasibility
- ▶ **Randomized experiments** tend to balance both **observed** and **unobserved** variables
- ▶ In contrast, in **observational studies** it is not reasonable to expect that **unobservables** can be well controlled!
- ▶ Key to balancing covariates in observational studies is the **propensity score**

The Propensity Score

- ▶ Assume that vector \mathbf{Z}_i contains observed covariates collected for individual i **prior** to assigning them to group $G = A$ or B
- ▶ In biomedical studies \mathbf{Z}_i might include age, sex, body mass index, laboratory tests, medication and family history
- ▶ Importantly, \mathbf{Z}_i may or may not include **all** relevant pieces of information that contribute to group A or B assignment!
- ▶ The probability $P(G = A|\mathbf{Z})$ is called the **propensity score** (PS) to be assigned to group A
- ▶ To estimate the PS, one option is to use a logistic regression model

The Propensity Score as a Balancing Score

- ▶ A fundamental difference: \mathbf{Z} is a vector, but the PS is simply a **number**! And not only that!
- ▶ **Given** the PS, covariates \mathbf{Z} have the **same distribution** in groups A and B !
- ▶ This is **comparable**, but **not identical** to being randomized to treatment A or B

How Propensity Score Matching Works: An Intuitive Description

- ▶ Obviously, each patient is assigned to one and only one group: A or B
- ▶ Under different circumstances (another MD, repeat lab tests?) one might have been assigned to the other group, but was not!
- ▶ Hence, **given** their covariates \mathbf{Z} , we model the probability of being assigned to group A
- ▶ This probability $P(\text{Group} = A|\mathbf{Z})$ is the **propensity score** (PS)

How PS-Matching Works – Continued

- ▶ Pool groups A and B together
- ▶ Each individual in the pooled group will have their own *estimated* propensity score \widehat{PS}
- ▶ Then, find individuals with **similar** PS values in groups A and B and match them
- ▶ Here is a small-scale hypothetical, yet illustrative, example

PS-Matching: Hypothetical Example

- ▶ Estimated PS values in groups A and B
- ▶ Find **similar** PS values in the two groups and match them!

Group A		Group B	
Patient	Estimated Propensity Score	Patient	Estimated Propensity Score
1	0.83	3	0.12
2	0.79	4	0.66
6	0.23	5	0.29
7	0.54	8	0.78
9	0.11	10	0.37
11	0.27		
12	0.65		

PS-Matching: Key Considerations

- ▶ PS-matching is essentially a form of **data processing** and is no substitute for data analysis
- ▶ No analyses of outcomes should be performed until the PS-matching part of data processing has been finalized
- ▶ It is **not principled** to seek the PS-matched groups in which differences in outcomes are most significant!
- ▶ Balancing assessment is performed for each covariate, typically using the standardized mean difference (SMD)
- ▶ For a variable X measured in groups A and B , the SMD is

$$\frac{\text{mean}(X_A) - \text{mean}(X_B)}{\sqrt{0.5[\text{Var}(X_A) + \text{Var}(X_B)]}}$$

Statins in Cardiac Surgery Example (Vaduganathan et al., Annals of Thoracic Surgery, 2012)

Midterm Benefits of Preoperative Statin Therapy in Patients Undergoing Isolated Valve Surgery

Muthiah Vaduganathan, MD, MPH, Neil J. Stone, MD, Adin-Cristian Andrei, PhD, Richard Lee, MD, MBA, Preeti Kansal, MD, Robert A. Silverberg, MD, Robert O. Bonow, MD, and Patrick M. McCarthy, MD

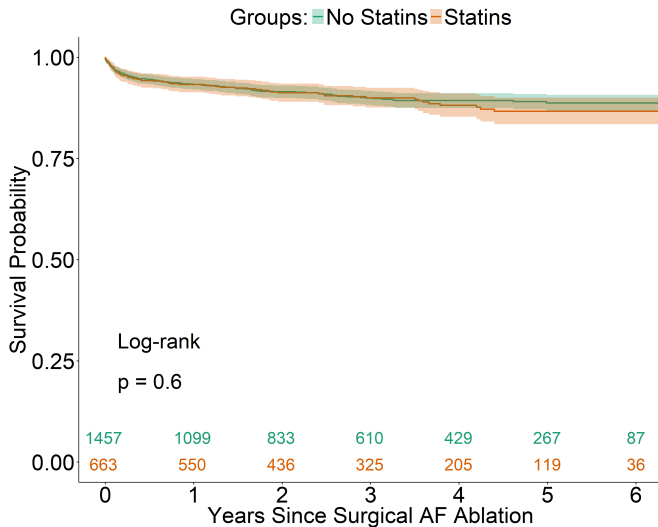
- ▶ Evaluate association of **preoperative** statins with overall survival after cardiac surgery
- ▶ Decision to prescribe statins was part of a treatment strategy specifically tailored to each individual's set of comorbidities

Background

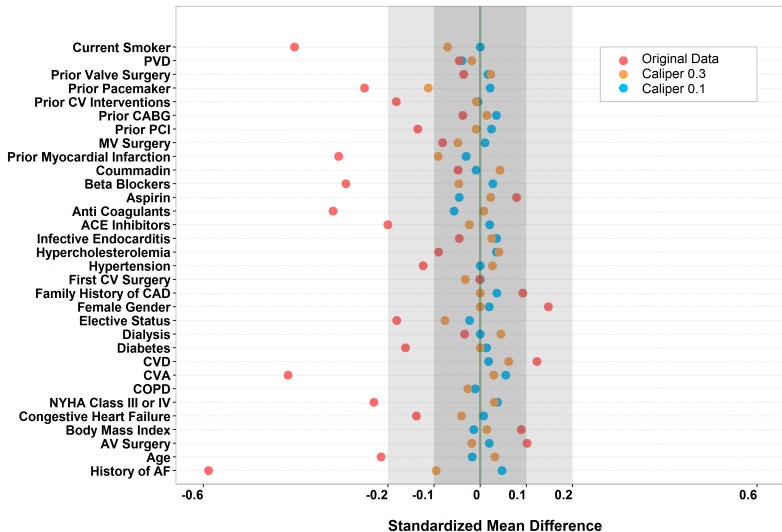
- ▶ Statins are first line drugs in lipid lowering strategies
- ▶ The benefits of statins in patients undergoing isolated valvular heart surgery are still being researched
- ▶ 2120 consecutive patients underwent isolated cardiac valvular surgery at NM between April 2004 and April 2010
- ▶ 663 (31%) patients were administered statins before surgery and 1457 (69%) were not

Table 1 Preoperative and Intraoperative Characteristics by Preoperative Statins Status				
Variable	N	No Statins (N=1457)	Statins (N=663)	P-value
Age (years)	2120	57.6 ± 16.1	65.9 ± 12.2	<.001
BMI (kg/m ²)	2120	27.1 ± 5.9	28.4 ± 5.8	<.001
Female	2120	641 (44%)	258 (39%)	0.028
Current Smoker	2120	52 (4%)	14 (2%)	0.07
Family History of CAD	2120	169 (12%)	108 (16%)	0.003
Diabetes	2120	133 (9%)	112 (17%)	<.001
Dyslipidemia	2120	356 (24%)	570 (86%)	<.001
Hypertension	2120	708 (49%)	456 (69%)	<.001
Infectious Endocarditis	2120	135 (9%)	40 (6%)	0.012
Chronic Lung Disease	2120	172 (12%)	116 (17%)	<.001
First CV Surgery	2120	1189 (82%)	516 (78%)	0.042
Peripheral Vascular Disease	2120	71 (5%)	51 (8%)	0.010
Cerebrovascular Disease	2120	147 (10%)	87 (13%)	0.039
Previous CV Intervention	2120	409 (28%)	248 (37%)	<.001
Previous CABG	2120	69 (5%)	91 (14%)	<.001
Previous Valve	2120	182 (12%)	66 (10%)	0.09
Previous PCI	2120	55 (4%)	73 (11%)	<.001
Previous Pacemaker	2120	78 (5%)	43 (6%)	0.30
Previous MI	2120	41 (3%)	65 (10%)	<.001
Congestive Heart Failure	2120	404 (28%)	209 (32%)	0.07
NYHA Class III/IV	2111	458 (32%)	252 (38%)	0.003
Preop Beta Blockers	2120	646 (44%)	356 (54%)	<.001
Preop ACE Inhibitors	2120	320 (22%)	221 (33%)	<.001
Preop Aspirin	2119	190 (13%)	192 (29%)	<.001
AV Surgery Alone	2120	661 (45%)	362 (55%)	<.001
MV Surgery Alone	2120	645 (44%)	245 (37%)	0.002

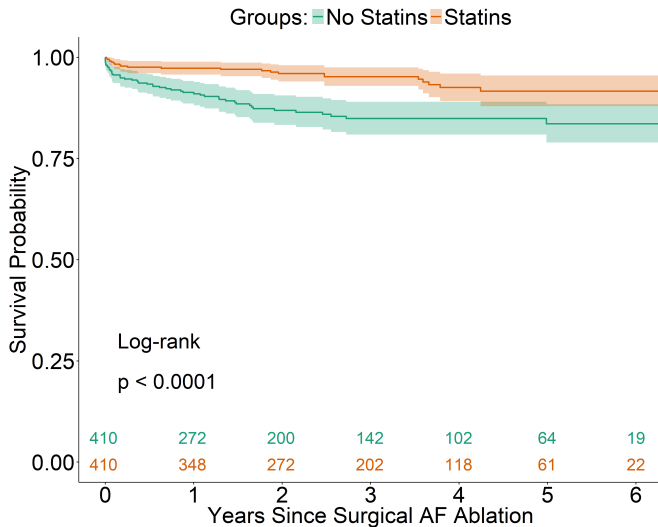
Overall Survival in the Original Groups



Standardized Mean Differences Before and After PS-Matching



Overall Survival in the Propensity Score Matched Groups



Gender Differences in Bicuspid Aortic Valve Patients (Andrei et al., American Journal of Cardiology, 2015)

Comparison of Outcomes and Presentation in Men-Versus-Women With Bicuspid Aortic Valves Undergoing Aortic Valve Replacement



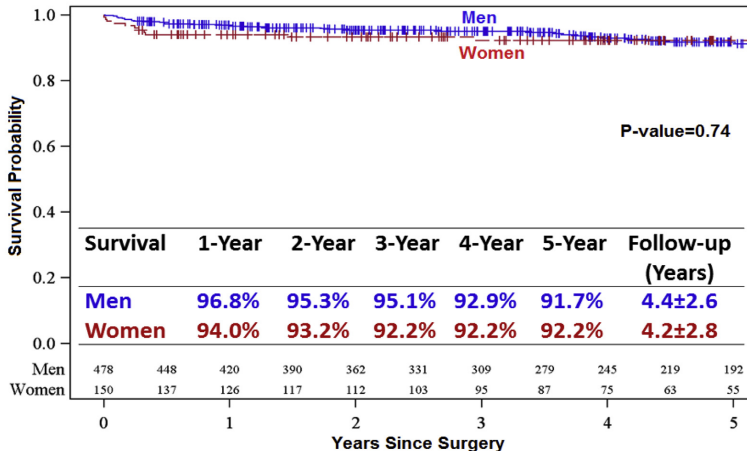
Adin-Cristian Andrei, PhD^{a,*}, Ajay Yadlapati, MD^b, S. Chris Malaisrie, MD^a, Jyothy J. Puthumana, MD^b, Zhi Li, MS^a, Vera H. Rigolin, MD^b, Marla Mendelson, MD^b, Colleen Clennon, BSN^a, Jane Kruse, BSN^a, Paul W.M. Fedak, MD^{a,c}, James D. Thomas, MD^b, Jennifer A. Higgins, MD^a, Daniel Rinewalt, MD^a, Robert O. Bonow, MD^b, and Patrick M. McCarthy, MD^a

- ▶ Gender disparities in outcomes documented in AV replacement surgery, but not in bicuspid AV (BAV) patients
- ▶ Retrospective analysis of 628 consecutive BAV patients who underwent AVR surgery from 04/2004 to 12/2013
- ▶ Observational study, but different: **natural experiment!**

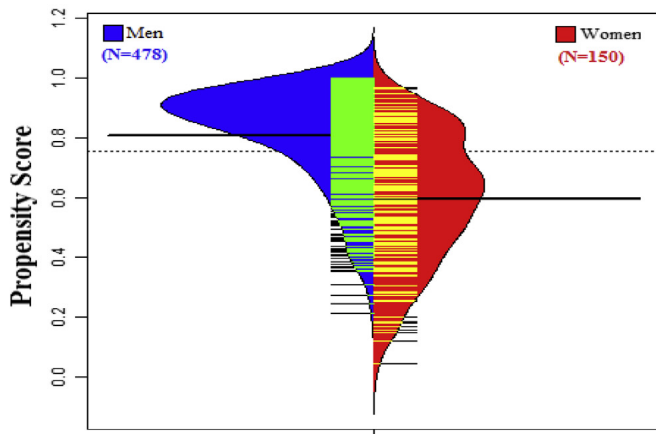
Gender Differences in Bicuspid Aortic Valve Patients

	Men (N=478)	Women (N=150)	P-value
Preoperative			
Age (years)	56.3 ± 13.6	60.7 ± 13.8	<.001
Body Mass Index (kg/m ²)	28.2 ± 4.4	27.6 ± 7.2	0.22
Repeat Sternotomy	57 (12%)	16 (11%)	0.67
Prior Valve Surgery	40 (8%)	18 (12%)	0.18
NYHA Functional Class III or IV Symptoms	93 (20%)	37 (25%)	0.18
Left Ventricular Ejection Fraction (Q1, Q3)	60.0 (53.0, 65.0)	61.0 (60.0, 65.0)	<.001
Severe Aortic Stenosis †(N=620)	307 (65%)	117 (79%)	0.001
Aortic Insufficiency ≥ 2+	228 (48%)	44 (29%)	<.001
Aortic Valve Area (cm ²)	1.4 ± 1.2	0.9 ± 0.7	<.001
Aortic Valve Mean Gradient (mmHg) (N=551)	34.2 ± 21.3	43.4 ± 22.6	<.001
Aortic Valve Peak Gradient (mmHg) (N=419)	53.6 ± 33.6	68.9 ± 35.3	<.001
Intraoperative			
Perfusion Time (minutes)	114.0 (81.0, 169.0)	83.5 (64.0, 126.0)	<.001
Cross Clamp Time (minutes)	95.0 (71.0, 134.0)	72.5 (55.0, 104.0)	<.001
Ascending Aortic Replacement (N=602)	82 (18%)	15 (11%)	0.040
Bioprosthetic Aortic Valve Implant	423 (88%)	140 (94%)	0.390
Aortic Valve Implant Size ≤ 21mm (%)	10 (2%)	63 (44%)	<.001
Coronary Artery Bypass Grafting	92 (19%)	18 (12%)	0.042
Postoperative			
Postoperative Blood Products	165 (35%)	72 (48%)	0.003
Any Complication*	171 (36%)	52 (35%)	0.80
Post-Operative Length of Stay (Days)	5.0 (4.0, 6.0)	5.0 (5.0, 7.0)	0.011
Total ICU Hours	27.4 (23.2, 46.7)	31.2 (24.0, 69.5)	0.014
Readmission within 30 Days	59 (12%)	13 (9%)	0.24
30-Day Mortality	2 (0%)	4 (3%)	0.014

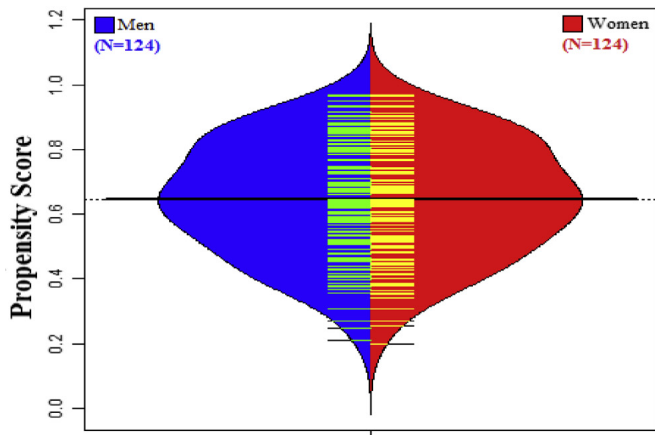
Gender Differences in BAV: Overall Survival in Original Groups



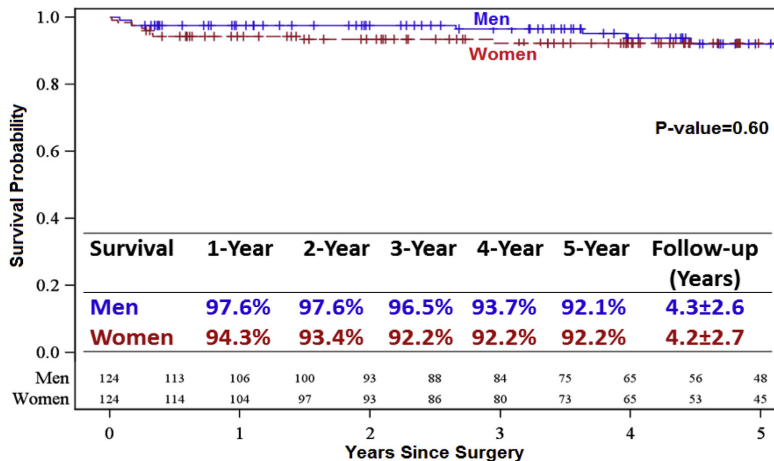
Gender Differences in BAV: Beanplots of PS in the Original Groups




Gender Differences in BAV: Beanplots of PS in the PS-Matched Groups



Gender Differences in BAV: Overall Survival in PS-Matched Groups



Late Reinterventions After AF Ablation (Andrei et al., Heart Rhythm, 2015)

Overcoming reporting challenges: How to display, summarize, and model late reintervention outcomes, follow-up, and vital status information after surgery for atrial fibrillation 

Adin-Cristian Andrei, PhD, MS,^{*} Patrick M. McCarthy, MD,^{*} James D. Thomas, MD,^{*} Travis O. Abicht, MD,^{*} S. Chris Malaisrie, MD,^{*} Zhi Li, MS,^{*} Jane Kruse, BSN,^{*} Albert L. Waldo, MD, PhD (Hon), FHRS,[†] Hugh Calkins, MD, FHRS,[‡] James L. Cox, MD[§]

- ▶ Atrial fibrillation (AF) is the most common heart rhythm dysfunction in the US (> 2.5m individuals)
- ▶ AF treatment options include cardioversion (**CV**), catheter ablation (**CA**) or surgical ablation (**SA**)
- ▶ Appropriate risk/benefit ratio metrics are necessary when considering a sinus rhythm (**SR**) restoration intervention

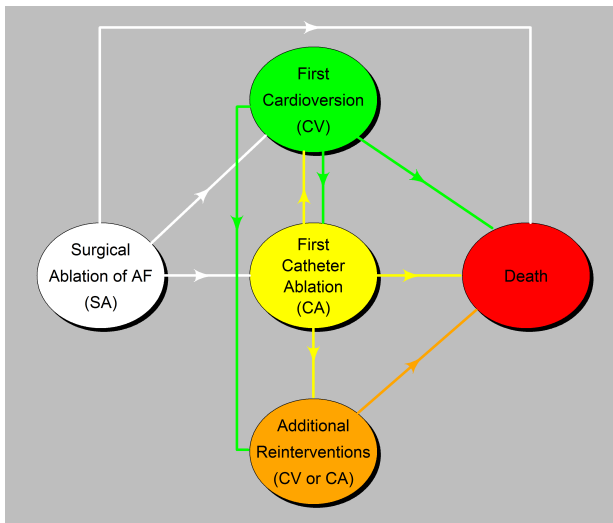
Background

- ▶ 2012 HRS/EHRA/ECAS Consensus Document on Catheter and Surgical AF Ablation, defines **failure** as
 - ▶ any symptomatic or asymptomatic episode of AF, AFL or AT
 - ▶ at least 30 seconds duration
 - ▶ after the 3-month blanking period off-antiarrhythmic drug therapy
- ▶ Widely used as an endpoint of AF ablation trials, yet very strict and may underestimate clinical benefit of AF ablation
- ▶ **SR success** typically defined as the freedom-from-AF at pre-specified timepoints (yearly marks or last follow-up)

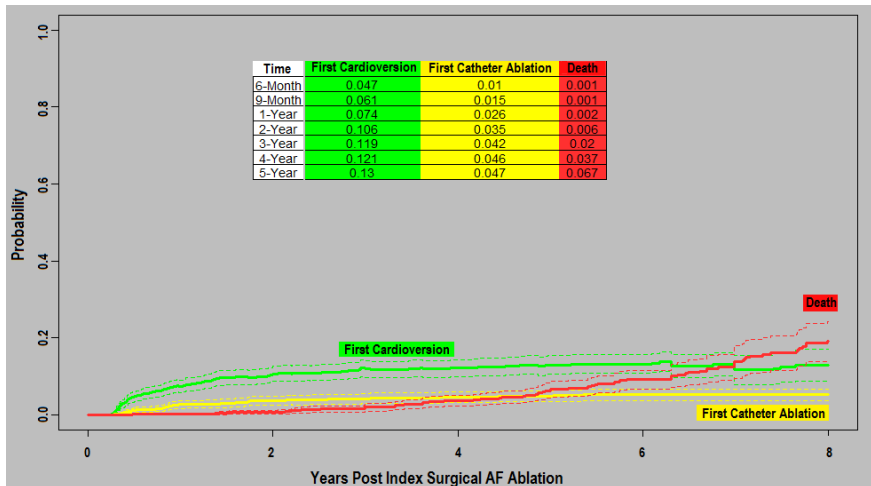
Background

- ▶ Overall temporal dynamics of success only partially conveyed
- ▶ Likelihood of AF detection is directly proportional to the duration and frequency of arrhythmia monitoring
- ▶ Not straightforward to summarize success when patients experience intermittent AF episodes before SR restoration
- ▶ Instead, we focus on freedom-from-late-reinterventions, such as **CV** or **CA** (**hard** endpoints)

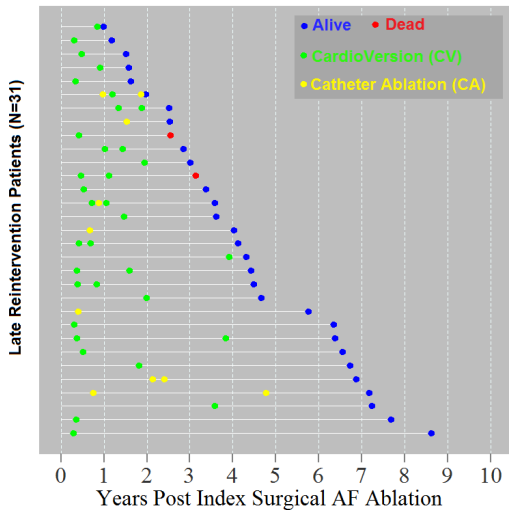
└ Multi-State Models: Diagram of Possible States For a Patient



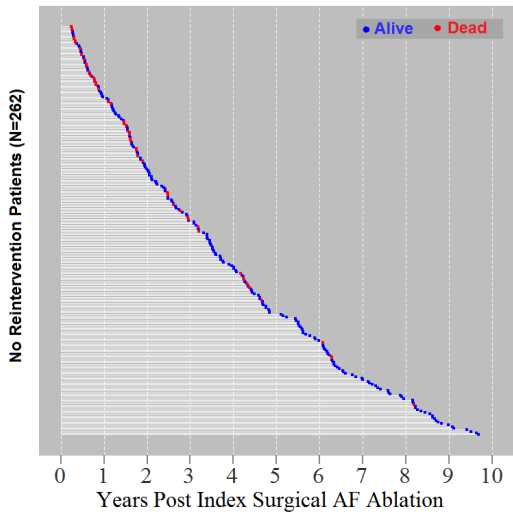
└ Multi-State Models: A Summary of State Occupancy Probabilities



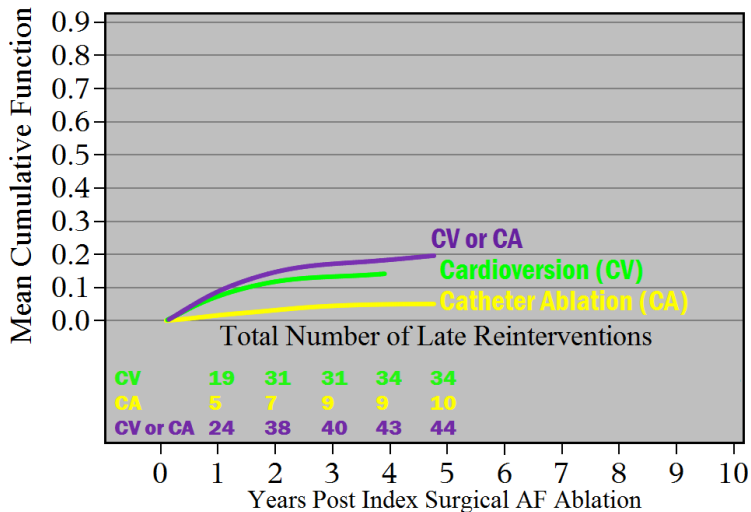
Summary of Mortality And Outcomes Reported Over Time



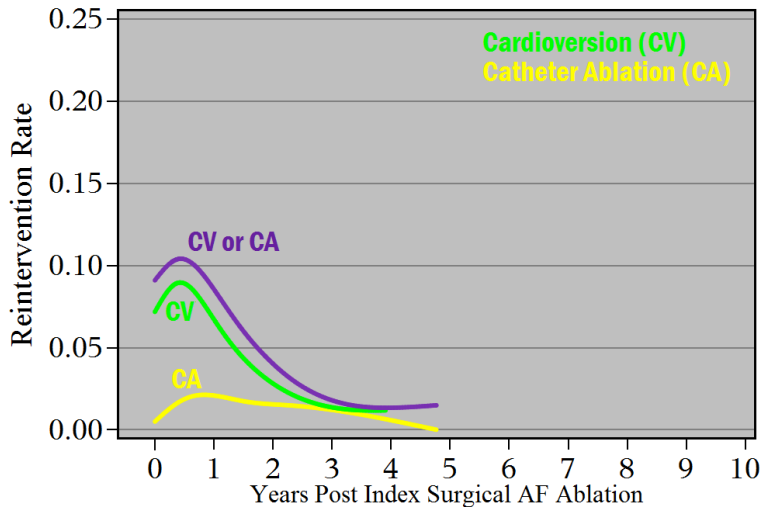
SMART Plot: Patients Age 75 Or Older



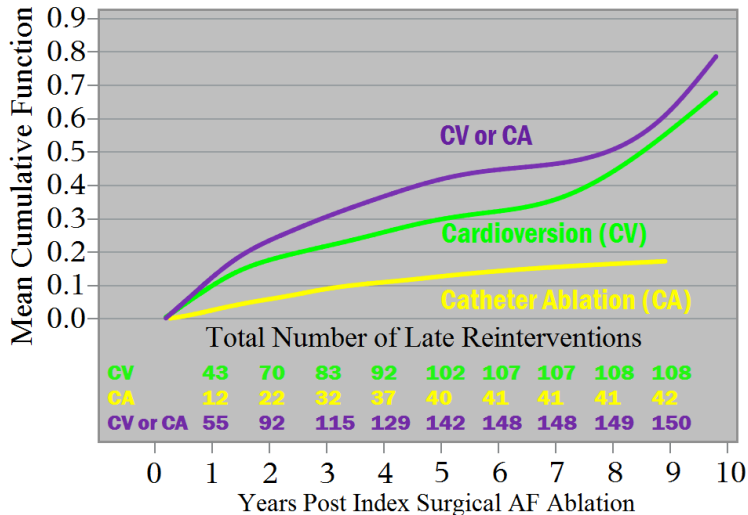
↳ Mean Cumulative Number of Reinterventions Over Time: Patients Age 75 Or Older



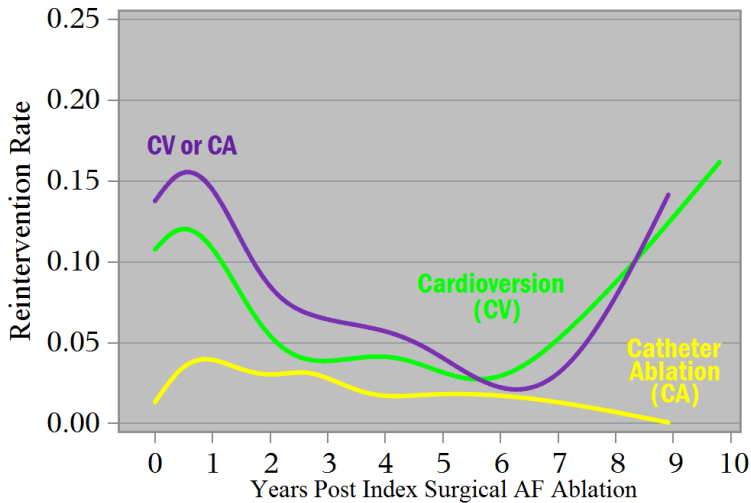
└ Reintervention Rates Over Time: Patients Age 75 Or Older



Mean Cumulative Number of Reinterventions Over Time: Patients Age 60 to 75



└ Reintervention Rates Over Time: Patients Age 60 to 75



Examples Discussed in Relationship to POMs

- ▶ **Study design:** Observational – PS-matching as a way to balance covariates in two groups (akin to randomization)
- ▶ **Measure outcome:** AF is a **soft** endpoint, but CV or CA are **hard** endpoints
- ▶ **At-risk period:** multi-state models are elegant ways to achieve this
- ▶ **Risk adjustment:** models for recurrent events are natural ways to risk-adjust and go beyond Cox regression
- ▶ **Data structure:** accounting for the fact that multiple (recurrent) events per patient are recorded
- ▶ **Reporting period and at-risk period:** clearly delineated
- ▶ **Reliability and validity testing:** should perhaps constitute the new standards [Andrei (JT CVS 2016)]

Opportunities and Conclusions

- ▶ Study design remains highly relevant and always supersedes analysis in the order of importance (Rubin 2008)
- ▶ Underutilized statistical tools that can help define or refine POM reporting: SMART plots, recurrent events, multi-state models
- ▶ And also lots of overutilized methods that might not be entirely adequate...
- ▶ To identify these tools, disease area knowledge and team participation are important for a statistician

Thank you!

Questions?