Reconciling Conflicting Evidence Across Target Trial Emulations: What Standards Should We Use?

Sebastien Haneuse [1], Luke Benz [1]
Valerie Smith [2, 3], David Arterburn [4], Matthew Maciekewski [2, 3]

[1]Harvard University [2]Duke University [3]Veterans Affairs [4]Kaiser Permanente

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Society for Epidemiologic Research

Introduction

Substantial observational evidence exists (e.g. **Fisher et al.** [2]) in support of bariatric surgery being associated with a reduction in risk for CVD outcomes in patients with diabetes (T2DM).

Madenci et al. [1]:

- prior work is flawed because it does not align with any corresponding target trial
- concern about the 'pre-operative' period
- emulate two target trials using VA data
- claim evidence of no CVD benefit for bariatric surgery in T2DM patients

We use data from a prior Kaiser Permanente (KP), originally reported on by **Fisher et al.** [2], to emulate a trial that mimics the methods employed by **Madenci et al.** [1].

Target Trials for Bariatric Surgery

EHR data on adults with T2DM at one of three KP systems between 01/2005 to 12/2011, with follow-up through 09/2015.

Trial 1A
* Eligibility closest to **Fisher et al. [2]**

Trial 1B
* Additional criteria (non-current smoker, 1-year BMI change < +2 kg/m^2)
* Acknowledge possible lifestyle changes in pre-operative period

Trial 1C
* Additional exclusions for various comorbidities
* Eligibility closest to **Madenci et al. [1]**

As in **Madenci et al.** [1], we fit a marginal structural discrete time hazard model, using propensity scored-based IPW for confounding

- PS model #1 linear terms
- PS model #2 splines

Results

Figure 1. Marginal, adjusted cumulative incidence comparing bariatric surgery to no surgery.

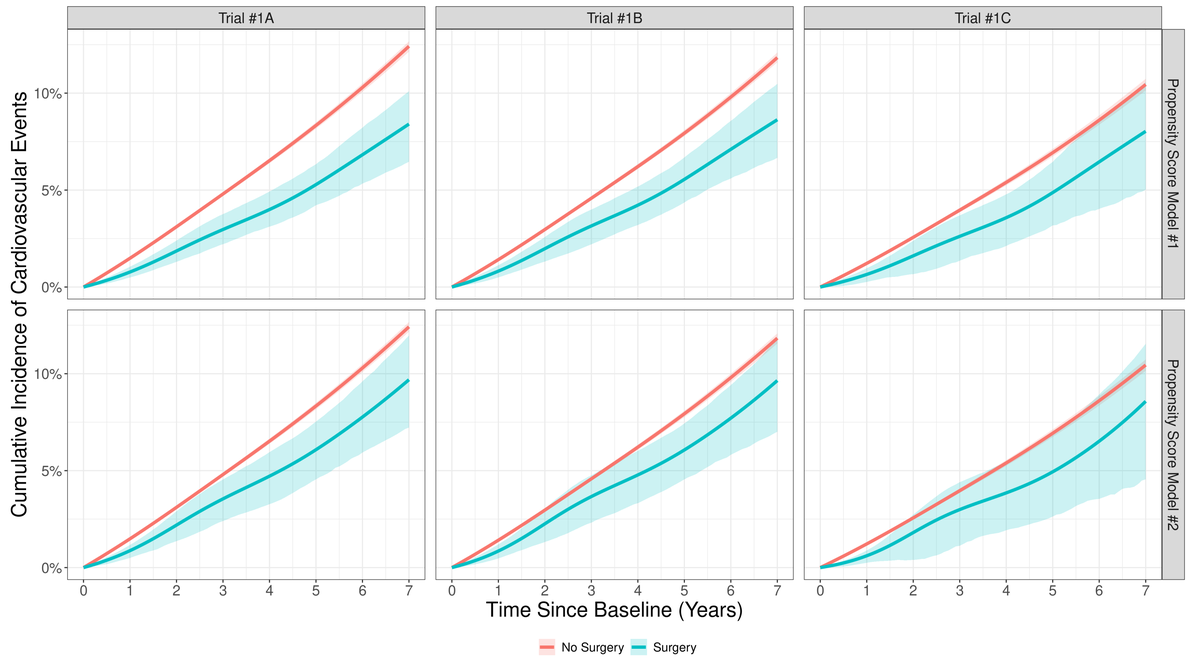

Table 1. Adjusted 5- and 7-year risk ratio and differences for CVD events comparing bariatric surgery to no surgery across combinations of eligibility criteria and PS model.

	5-Year		7-Year	
	RR	**RD**	**RD**	**RD**
Trial 1A				
PS model #1	0.63 (0.51, 0.76)	-3.1% (-4.1 , -2.0)	0.68 (0.52, 0.81)	-4.0% (-6.0 , -2.3)
PS model #2	0.73 (0.54, 0.90)	-2.3% (-3.8 , -0.8)	0.78 (0.58, 0.97)	-2.7% (-5.2 , -0.4)
Trial 1B				
PS model #1	0.70 (0.55, 0.84)	-2.4% (-3.5 , -1.2)	0.73 (0.56, 0.89)	-3.2% (-5.2 , -1.4)
PS model #2	0.76 (0.57, 0.94)	-1.9% (-3.4, -0.5)	0.82 (0.59, 1.00)	-2.2% (-4.9 , 0.0)
Trial 1C				
PS model #1	0.70 (0.45, 0.94)	-2.1% (-3.8 , -0.4)	0.77 (0.47, 1.01)	-2.4% (-5.6 , 0.1)
PS model #2	0.71 (0.38, 0.99)	-2.0% (-4.3 , -0.1)	0.82 (0.42, 1.11)	-1.9% (-6.0 , 1.2)

Summary of Results

Figure 1 and Table 1 provide evidence that bariatric surgery is associated with reductions in long-term CVD risk in patients with T2DM. This conclusions is consistent with **Fisher et al.** [2], who used the same data but employed a matched cohort study design, and the vast majority of the clinical literature. However, these conclusions discordant from those of **Madenci et al.** [1].

Reconciling Discordant Results

(1) Differences in # bariatric cases

- This work & **Fisher et al.** [2] $\sim$ 5,000
- **Madenci et al.** [1] only had 435

Q: Implications for estimation/inference?

(2) Differences in patient populations

	KP	VA
Non-surgical controls		
Mean age (years)	51.4	58.0
Bariatric surgery cases		
Mean age (years)	47.4	53.0
% Female	77.4	26.0
Mean baseline BMI (kg/m^2)	44.0	40.1

Q: A comparison of apples and oranges?

(3) Conservative clinical trial thinking

Focus is typically on establishing efficacy through tight control of inclusion criteria

Well meaning exclusions removed patients who actually underwent surgery in the KP data (!)

- 5.1% w/ BMI $\uparrow \geq$ 2 kg/m^2 in prior year
- Trial #1C excluded 46% of observed cases
- 97% b/c of depression or bipolar disorder

Q: Compromised generalizability?

Takeaways

- In the absence of a gold standard, care is needed when holding up any given approach, or set of results, as definitive
- Evidence triangulation may provide a framework for constructive progress

References

[1] A. Madenci et al. Estimating the effect ... *Epidemiology*, 35(5):721–729, 2024.

[2] D. Fisher et al. Association between ... *JAMA*, 320(15):1570–1582, 2018.