500 Class 07

https://thomaselove.github.io/500-2023/

2023-03-02

Today's Agenda

- The Right Heart Catheterization Example
- Some Extensions to Matching (originally in Slides 5)
- Discussion of Kubo et al. (2020)
- Discussion of Lab 3 - how did it go?

Section 1

Right Heart Catheterization and the SUPPORT Study

The SUPPORT Study

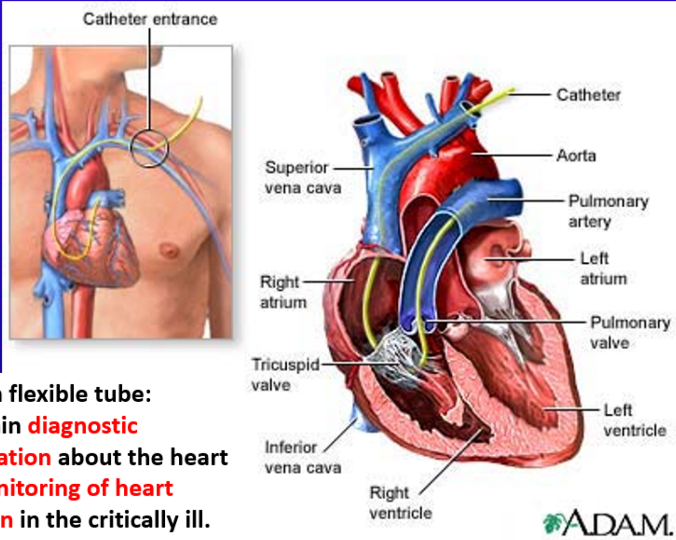
This example is based on the Right Heart Catheterization data set available at Vanderbilt University.

The key reference is Connors AF et al. (1996) The effectiveness of RHC in the initial care of critically ill patients. JAMA 276: 889-897.

Connors et al. used a logistic regression model to develop a propensity score then: [a] matched RHC to non-RHC patients and [b] adjusted for propensity score in models for outcomes, followed by a sensitivity analysis.

The key conclusions were that RHC patients had decreased survival time, and any unmeasured confounder would have to be somewhat strong to explain away the results.

Right Heart / Swan-Ganz / Pulmonary Artery Catheterization



Pass a thin flexible tube:

1. **to obtain diagnostic information about the heart**
2. **for monitoring of heart function in the critically ill.**

http://www.nlm.nih.gov/medlineplus/ency/imagepages/18087.htm

Does the RHC do more harm than good?

Prior (small) observational studies comparing RHC to non-RHC patients:

- RR of death higher in RHC elderly patients than non-RHC elderly
- RR of death higher in RHC patients with acute MI than non-RHC patients with MI
- Patients with higher than expected RHC use had higher mortality

Big Problem: Selection Bias. Physicians (mostly) decide who gets RHC and who doesn't.

Why not a RCT?

- RHC directly measures cardiac function
- Some MDs believe RHC is necessary to guide therapy for some critically ill patients
- Procedure is very popular - existing studies haven't created equipoise

Characteristics used to predict PS(RHC usage)

- Age, Sex, Race
- Education, Income, Insurance
- Primary and Secondary Disease category
- Admission diagnosis category (12 levels)
- ADL and DASI 2 weeks before admission
- DNR status on day 1
- Cancer (none, local, metastasized)
- 2 month survival model
- Weight, temperature, BP, heart rate, respiratory rate
- Comorbid illness (13 categories)
- Body chemistry (pH, WBC, PaCO_2_, etc.)

Panel (7 specialists in clinical care) specified important variables related to the decision to use or not use a RHC.

RHC vs. Non-RHC patients

RHC patients were more likely to

- Be male, have private insurance, enter the study with ARF, MOSF or CHF

RHC patients were less likely to

- Be over 80 years old, have cancer, have a DNR order in the first 24 hours of hospitalization

RHC patients had significantly

- Fewer comorbid conditions,
- More abnormal results of vital signs, WBC count, albumin, creatinine, etc.
- Lower model probability of 2-month survival

What's in the RHC Example?

- exposure/treatment is the installation of a Swan-Ganz (right heart) catheter on day 1
- 3 outcomes
  - binary: in-study mortality
  - quantitative: hospital length of stay, in days
  - time-to-event: time to death (with censoring)
- 50 covariates, including socio-demographics, presentation and diagnoses at admission, comorbid illness and transfer status, summary measures of presentation and lab results on day 1

What Analyses are Presented?

- Unadjusted analyses
- Estimating the PS and checking balance before adjustment
- Six different Matching Approaches and resulting outcomes
  - 1:1 greedy matching without replacement
  - 1:2 greedy matching without replacement
  - 1:1 matching without replacement using genetic search
  - 1:1 greedy matching with replacement
  - 1:1 caliper matching without replacement
  - 1:2 greedy matching with replacement
- ATT weighting using TWANG
- Double Robust (weighting + regression) analysis
- Sensitivity Analyses after Matching

Section 2

Some Extensions to Propensity Matching

Is Regression Adjustment Unnecessary?

- Matching and stratification are old and trusted methods of adjustment for observational studies, but the difficulty of implementing them led earlier practitioners to prefer regression.
- Modern extensions to matching methods let us perform optimal matches, full matches and optimal full matches, and to control imbalance (or at least reduce bias reduction) in ways that have become attainable only in recent years.

Good references include Rosenbaum (2010) and Hansen (2004) for example.

General Approaches to Optimal or Near-Optimal Constrained Matching

1. Calculate propensity scores
2. Establish a **distance matrix**

This is just a table with one row for each treated subject and one column for each potential control.

- The “distances” can be squared differences in propensity scores between the subjects, Mahalanobis distances, or something else.
- To use calipers, we set to ∞ all cells in the table corresponding to a propensity difference which exceeds the caliper.

A Small Distance Matrix

Consider four treated subjects (T1, T2, T3 and T4) and six control subjects (C1, C2, C3, C4, C5 and C6.)

- We have a difference score (perhaps the absolute difference in propensity for treatment) for each comparison. Some of these are infinite.
- We also have each subject categorized as (Y)oung or (O)ld, and we haven't decided yet how important this is for our matching.

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

OK, so Who Gets Matched?

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

- Now, who gets matched?

OK, so Who Gets Matched?

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

- Now, who gets matched?
- Treated subject T1 matches to C1

OK, so Who Gets Matched?

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

- Now, who gets matched?
- Treated subject T1 matches to C1
- T2 matches to C3

OK, so Who Gets Matched?

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

- Now, who gets matched?
- Treated subject T1 matches to C1
- T2 matches to C3
- T3 matches to C4 (or maybe C6 - is age important?)

OK, so Who Gets Matched?

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

- Now, who gets matched?
- Treated subject T1 matches to C1
- T2 matches to C3
- T3 matches to C4 (or maybe C6 - is age important?)
- T4 matches to C6 (or C2, or C4, hmmm....)

Almost Exact Matching

- Suppose a few of the covariates are of enormous importance - want to match exactly on them wherever possible.

We could add a penalty (but perhaps not an infinite penalty) to the distance matrix when the specified covariates fail to match, and that is the main approach that we use.

- Adding 2 to the Mahalanobis distance for mismatches roughly doubles the importance of that covariate as compared to the others, for example.

There's a lot of active work in this area developing various algorithms that permit finer control.

“Fine Balance” in Matching

- Constrain optimal matching that forces a nominal variable to be balanced, without restricting who is matched to whom.

This is especially useful if...

- you have a nominal variable with many levels
- you have a rare binary variable that is difficult to control using a distance
- you are focused on the interaction of several nominal variables

It is also possible to get specific imbalance patterns.

Fine Balance: Initial Distance Matrix

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24

Suppose we want to get optimal balance on the propensity score while matching perfectly on the age category (Y/O).

- We have 4 treated subjects (1 young, 3 old)
- We have 6 potential controls (2 young, 4 old)
- So we need to remove 1 young and 1 old in matching

Fine Balance: Augmented Distance Matrix

Subject	C1 (Y)	C2 (O)	C3 (O)	C4 (Y)	C5 (O)	C6 (O)
T1 (Y)	.23	.47	.39	∞	.51	.35
T2 (O)	.45	∞	.28	.31	.42	∞
T3 (O)	∞	.35	∞	.27	.44	.28
T4 (O)	.31	.26	.51	.29	∞	.24
Extra 1	0	∞	∞	0	∞	∞
Extra 2	∞	0	0	∞	0	0

Add 2 rows to the matrix, then run the match

- *Extra 1* pulls away one young control
- *Extra 2* pulls away one old control

The binary age category will be perfectly balanced across the matched sample, but the partners within each individual pair are not required to be in the same age category.

Fine Balance General Procedure

To get the minimum distance match with fine balance (on a nominal covariate, say GROUP)...

1. Cross tabulate GROUP with treatment indicator
2. Determine # of controls to remove from each category of GROUP to achieve perfect balance
3. Add one row for each control that must be removed, with 0 distance to its own category and infinite distance to all others
4. Find an optimal match for this square matrix
5. Discard extra rows and their matched controls

Section 3

Full Matching

Full Matching in Observational Studies

- In the past, it has been tough to implement full matching in observational studies, even though it is appealing in principle.
- Alignment of comparable treated and control subjects is as good as any alternate method, and potentially much better.
- Hansen (2004) modifies full matching with modifications to minimize variance as well as bias

In this example,

- Optimal full matching removes as much as 99% of the bias along a PS on which treated and control means are separated by 1.1 SD's.
- Reduces to insignificance biases along 27 covariates, while making use of more, not less, of the data than regression based analyses.

Hansen (2004) SAT Coaching Study

- Survey of a random sample of 1995-1996 SAT test takers about their preparation
- 12% of respondents had completed extracurricular test preparation courses
- Matching looked unattractive to the original researchers due to significant reduction in sample size, but they only considered 1:1 matching.
- Do 1:k matching options look better?

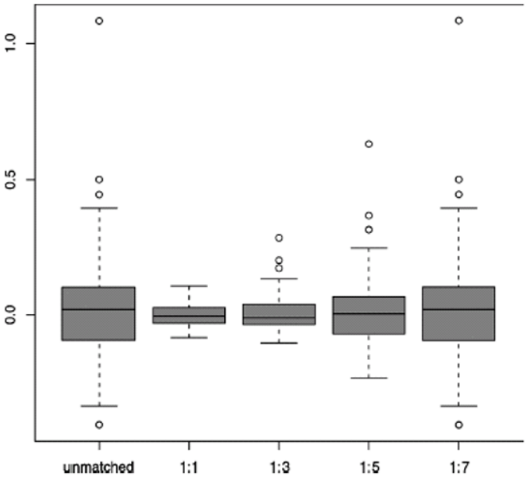


Figure 1. Covariate Imbalances in 1:k Matching. Each boxplot represents standardized biases in the 99 categories of the 27 categorical covariates along with standardized bias in the propensity score (which in each plot is the uppermost outlier). Strictly speaking, the matching represented at far right is not a 1:7 matching but a blend of six 1:6 and 494 1:7 matched sets.

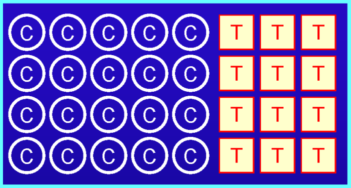
Covariate Imbalances in 1:k Matching

- In all of these cases, we're using less data
- Still some imbalance

Hansen 2004

Optimal Full Matching

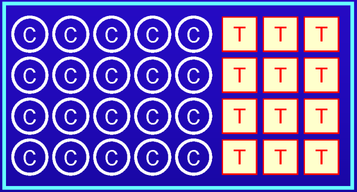
ORIGINAL SAMPLE



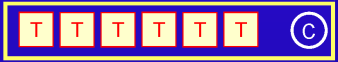
- OFM minimizes propensity score distances (discrepancies) while using all treated and all control subjects (i.e. discarding no units).

Optimal Full Matching

ORIGINAL SAMPLE

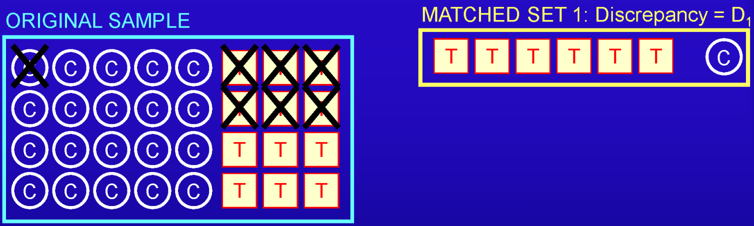


MATCHED SET 1: Discrepancy = D_1



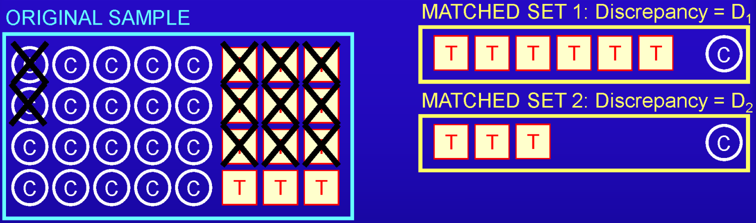
- OFM minimizes propensity score distances (discrepancies) while using all treated and all control subjects (i.e. discarding no units).
- Here, infinite distances force matches on Race×Sex

Optimal Full Matching



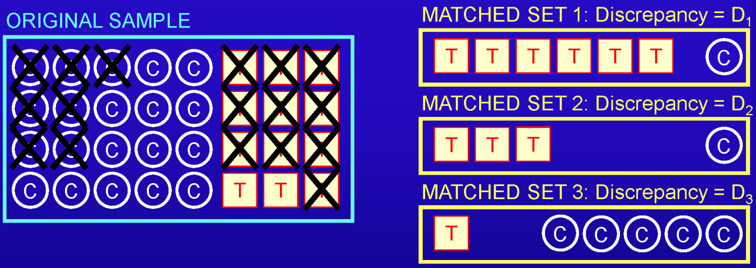
- OFM minimizes propensity score distances (discrepancies) while using all treated and all control subjects (i.e. discarding no units).
- Here, infinite distances force matches on Race×Sex

Optimal Full Matching



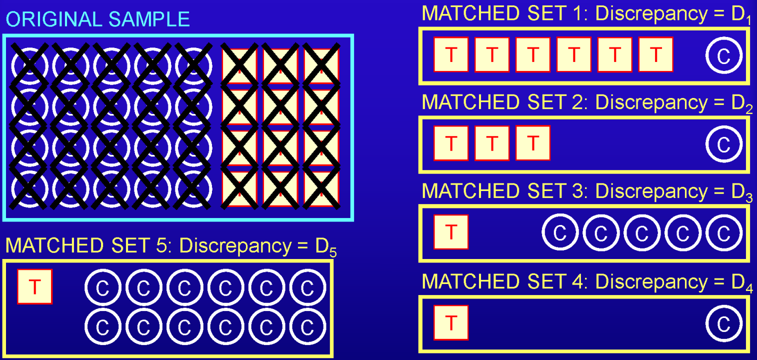
- OFM minimizes propensity score distances (discrepancies) while using all treated and all control subjects (i.e. discarding no units).
- Here, infinite distances force matches on Race×Sex

Optimal Full Matching



- OFM minimizes propensity score distances (discrepancies) while using all treated and all control subjects (i.e. discarding no units).
- Here, infinite distances force matches on Race×Sex

Optimal Full Matching



- OFM minimizes propensity score distances (discrepancies) while using all treated and all control subjects (i.e. discarding no units).
- Here, infinite distances force matches on Race×Sex

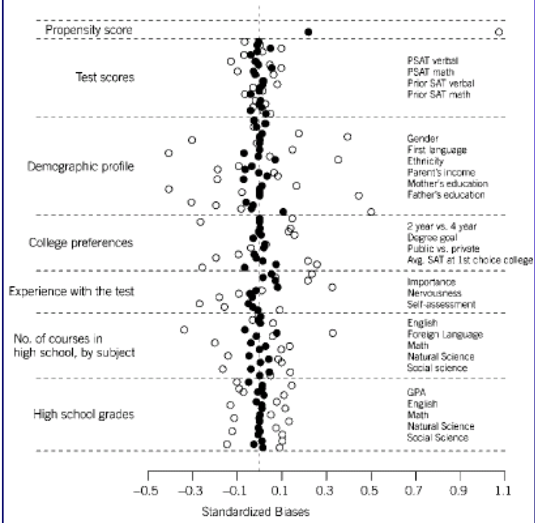


Figure 3. Standardized Biases Without Stratification or Matching, Open Circles, and Under the Optimal [.5, 2] Full Match, Shaded Circles.

Standardized Bias Plot

- Open circles are for standardized biases before matching
- Shaded circles describe results after full match

SAT Coaching Study Results

- Raw differences of treated and control group means were 41 points on Math and 9 on Verbal
- Full matching leads to aggregate contrasts of 26 points on Math and 1 point on the verbal.
  - Standard errors for these estimates are around 5 points.
- Surprised that Verbal effect is so small?
  - Control is not "no prep at all"
  - Estimated effect of treatment on the controls is 3 for Math and -8 on Verbal.
- Method doesn't require homogeneity of coaching effects.
- Whether and to what degree coaching is beneficial appears to vary greatly across students.

Section 4

Walking Through A Paper (Kubo et al. 2020): What To Look For

Kubo *et al. BMC Sports Science, Medicine and Rehabilitation* (2020) 12:72
https://doi.org/10.1186/s13102-020-00223-7

BMC Sports Science, Medicine and Rehabilitation

RESEARCH ARTICLE **Open Access**

Effects of preoperative low-intensity training with slow movement on early quadriceps weakness after total knee arthroplasty in patients with knee osteoarthritis: a retrospective propensity score-matched study

Yusuke Kubo*, Shuhei Sugiyama, Rie Takachu, Takeshi Sugiura, Masahiro Sawada, Kaori Kobori and Makoto Kobori

Kubo et al. (2020) Paper

Effects of preoperative low-intensity training with slow movement on early quadriceps weakness after total knee arthroplasty in patients with knee osteoarthritis: a retrospective propensity score-matched study

- **Population**
- **Outcome**
- **Treatment**
- **Covariates**

Key Words and Abbreviations

Key Words (from Abstract): Exercise preconditioning, Ischemic preconditioning, Ischemia-reperfusion injury, Knee swelling, Low-intensity training, Prehabilitation, Quadriceps weakness, Slow movement, Thigh swelling, Total knee arthroplasty

Abbreviations

TKA: Total knee arthroplasty; QW: Quadriceps weakness; IR: Ischemiareperfusion; IPC: Ischemic preconditioning; EPC: Exercise preconditioning; LST: Low-intensity resistance exercise with slow movement and tonic force generation; QST: Quadriceps strength test; TUG: Timed up and go test; SCT: Stair climb test; JKOM: Japanese Knee Osteoarthritis Measure; VAS: Visual analog scale; SMD: Standardized mean difference

Background

Background: Severe and early quadriceps weakness (QW) after total knee arthroplasty (TKA), which is caused by acute inflammation resulting from surgical trauma and tourniquet-induced ischemia-reperfusion (IR) injury, can be especially problematic. We focused on tourniquet-induced IR injury, because it has been shown to be preventable through ischemic and exercise preconditioning. Low-intensity resistance exercise with slow movement and tonic force generation (LST) share some similarities with ischemic and exercise preconditioning. The present study primarily aimed to clarify the efficacy of preoperative LST program as prehabilitation for early QW among patients with TKA using propensity score matching analysis.

Methods

Methods: This single-center retrospective observational study used data from patients with knee osteoarthritis ($n = 277$) who were scheduled to undergo unilateral TKA between August 2015 and January 2017. Those with missing outcome data due to their inability to perform tests were excluded. The LST group included participants who performed LST and aerobic exercise (LST session) more than seven times for three months prior to surgery. The control group included participants who performed less than eight LST sessions, a general and light exercise or had no exercise for three months prior to surgery. Knee circumference, thigh volume, knee pain during quadriceps strength test (QST) and timed up and go test (TUG), quadriceps strength, and TUG were measured before and 4 days after surgery. Knee swelling, thigh swelling, Δknee pain, QW, and ΔTUG were determined by comparing pre- and postoperative measurements.

Check 1

Can we describe the ...

- Population
- Outcome
- Treatment
- Covariates

Statistical Analysis Section (Start)

Statistical analysis was conducted using the IBM SPSS version 26 statistical software package (IBM Corp., Armonk, N.Y., USA). Participants were divided into the LST group and control group. The LST group included participants who performed category 1 sessions (LST and aerobic exercise) more than seven times for three months prior to surgery. The control group included participants who performed less than eight category 1 sessions, category 2 sessions (a general and light exercise) or had no prehabilitation (no exercise) for three months prior to surgery. Propensity score matching was used to balance group characteristics that could affect the LST program's instructions and formulae. Propensity scores were estimated using a logistic regression model where treatment status was regressed on age, gender, body mass index, and preoperative measurements, including quadriceps strength of the affected leg, knee pain during the QST and TUG, the TUG, the SCT, and JKOM scores. Propensity scores were subsequently used to match participants on a one-to-one basis using the nearest-neighbor method without replacement and a caliper width of 0.2 standard deviations of the logit of the propensity score.

Figure 1 from Kubo et al. (2020)

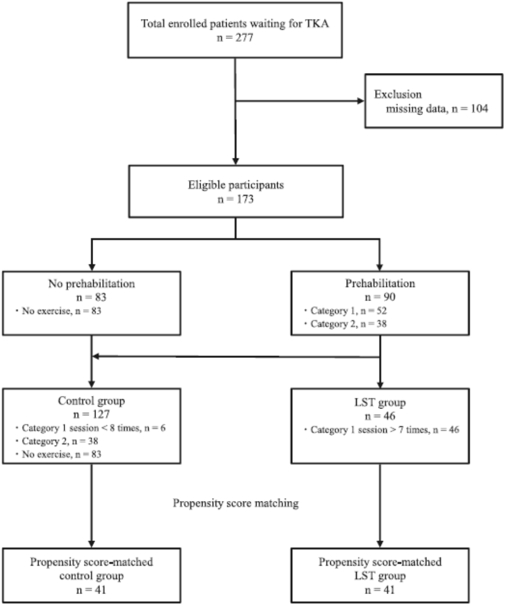


Fig. 1 Study flowchart. The LST group included participants who performed category 1 sessions (LST and aerobic exercise) more than seven

Table 1 from Kubo et al. (2020)

Table 1 Preoperative characteristics of participants and tourniquet time in the control and LST groups

	All participants			**Matched participants**		
	Control (*n* = 127)	**LST (*n* = 46)**	**SMD**	**Control (*n* = 41)**	**LST (*n* = 41)**	**SMD**
Age (years), median (IQR)	74 (68, 79)	71 (66, 75)	0.06	71 (66, 75)	71 (67, 75)	0.00
Male, n (%)	40 (31)	4 (9)	0.59	5 (12)	4 (10)	0.08
BMI (kg/m^2), median (IQR)	26 (23, 28)	25 (23, 28)	0.02	26 (24, 29)	25 (23, 28)	0.00
Current medical history, n (%)						
Heart disease	15 (12)	3 (7)	0.18	2 (5)	3 (7)	0.10
Diabetes	24 (19)	8 (17)	0.04	8 (20)	8 (20)	0.00
Hyperlipidemia	51 (40)	20 (43)	0.07	18 (44)	18 (44)	0.00
Rheumatoid arthritis	1 (1)	2 (4)	0.23	0 (0)	2 (5)	0.32
KL grade 3 of surgical knee, n (%)	14 (12)	5 (11)	0.00	4 (10)	5 (12)	0.08
Contralateral knee, n (%)						
OA and TKA, n (%)	102 (80)	38 (83)	0.06	32 (78)	33 (80)	0.06
Quadriceps strength, median (IQR)	1.3 (1.1, 1.7)	1.4 (1.2, 1.7)	0.05	1.3 (1.1, 1.7)	1.4 (1.2, 1.7)	0.01
T-handle cane usage, n (%)	7 (6)	1 (2)	0.17	0 (0)	1 (2)	0.22
SCT (s), median (IQR)	24 (16, 34)	20 (14, 25)	0.03	21 (17, 31)	22 (16, 27)	0.01
JKOM (points), median (IQR)	37 (25, 49)	32 (25, 45)	0.01	36 (32, 46)	34 (27, 48)	0.00
Tourniquet time (min), median (IQR)	58 (54, 66)	56 (52, 67)	0.01	56, (53, 63)	56 (52, 65)	0.01

Preoperative characteristics and tourniquet time between the groups were compared using standardized mean differences. *Abbreviations*: *LST* low-intensity resistance exercise with slow movement and tonic force generation, *SMD* standardized mean difference, *IQR* interquartile range, *BMI* body mass index, *KL* Kellgren and Lawrence, *OA* osteoarthritis, *TKA* total knee arthroplasty, *SCT* stair climb test, *JKOM* Japanese Knee Osteoarthritis Measure

Study Limitations

There are several limitations that need to be considered. First, the study included a small number of each group participants. Second, this was a single-center retrospective study; accounting for all unmeasured or unknown confounders affecting the outcomes was impossible, even after propensity score matching. Third, some variables remained imbalanced after propensity score matching. However, it is important to note that most imbalanced variables were worse in the LST group than that in the control group, suggesting that preoperative LST program may have improved early QW even in cases with relatively low physical function. Finally, given that QW was assessed only on postoperative day 4, it remains uncertain whether early QW suppression can optimize long-term postoperative recovery. In future, a large-scale multicenter randomized controlled trial with long-term follow up is needed to address these limitations.

Results and Conclusions

Results: Propensity score matching generated 41 matched pairs who had nearly balanced characteristics. The LST group had a significantly lower knee and thigh swelling, QW, and ΔTUG compared to the control group (all, $p < 0.05$). No significant differences in Δknee pain during the QST and TUG were observed between both groups (both, $p > 0.05$).

Abstract Conclusions: The present study demonstrated the beneficial effects of preoperative LST program on knee swelling, thigh swelling, QW, and walking disability immediately after TKA.

Conclusions Section (in the body of the paper)

The present study showed that preoperative LST program exerted beneficial effects on knee and thigh swelling, QW, and walking disability immediately after TKA. Future research addressing the limitations of this study is nonetheless needed to confirm the validity of our findings.

How Did Lab 3 go?

Answer Sketch should be posted to our Shared Drive by class time