Adelborg et al. 2019

Cardiovascular Outcomes and All-cause Mortality Following Measurement of Endogenous Testosterone Levels

PQHS500-OSIA Presentation
By Jason Huang
2020-04-09

Background

- Low testosterone causes
  - Body hair ↓
  - Muscle mass ↓
  - Sex drive ↓
  - Erectile dysfunction
  - Growth of breast tissue
- Testosterone therapy helps patients with **abnormally low testosterone levels**.
- Some recent studies suggest an increased **risk of cardiovascular disease**.
  - US FDA and Health Canada warn against testosterone therapy.
- Confounding by indication: the **indication of low testosterone itself** can be a confounder
  - Dictates the exposure
  - Might be associated with the outcome

Methods

Study Population

- Location: Central Denmark
- Duration: Jan 1, 2000 to Nov 1, 2015
- Men 18 years or older at the time of the measurement (first-ever testosterone measurement)
  - With normal or low testosterone levels based on age-specific reference
  - No prostate cancer, no exogenous testosterone and antiandrogen therapy 90 days prior to index date
- 4771 with low testosterone levels (treated) and 13467 with normal levels (control)

Methods, cont'd

Exposure and Outcome

- Exposure
  - **Low endogenous testosterone levels** based on age-specific reference v.s. **normal**
    - Data: the Clinical Laboratory Information System research Database
- Outcome
  - first-time **stroke**, myocardial infarction (**MI**), and venous thromboembolism (**VTE**)
    - Data: the Danish National Patient Registry (DNPR)
  - **All-cause death**
    - Data: the Danish Civil Registration System

Covariates

- Albumin level (inhibits testosterone function)
- Comorbidities and potential causes of low testosterone within 10 years before the index date
- Filled prescriptions

Determining Outcomes

For Each Outcome

- Whichever happens first
  - Outcome and/or death
  - emigration
  - 1 or 5 years of follow-up
  - Dec 31, 2015
- Censoring
  - **Testosterone treatment**
- Calculated **incidence rates(1 and 5 year cumulative)**

Adelborg et al. 2019

Propensity Score Analysis

PS with Inverse Probability of Treatment

- PS Estimated with **generalized boosted models**
- Transformed to **inverse probability of treatment weights (IPTW)** - ATT

Statistical Analysis

- Cox Regression for hazard ratios and CIs
  - before and after weighting
  - Proportional hazard assumption checked with log(-log) plots
- Standardized differences for all covariates
- Stratify by age

Results

- Table 1
  - Before and After IPTW
- 1-year and 5-year follow-up
  - Incidence Rate
  - Unadjusted hazard ratios
  - IPTW hazard ratios

Adelborg et al. 2019

Table One

	Unweighted cohorts		Propensity score weighted cohorts	
Variable	Low testosterone	Normal testosterone	Low testosterone	Normal testosterone
Number of men	4,771	13,467	4,771.0	4,470.7
Median age (25th–75th percentiles)	55.4 (38.3–69.2)	50.4 (33.8–63.4)	55.4 (38.3–69.2)	54.8 (37.9–68.5)
Albumin level				
Low	794 (17%)	734 (5.5%)	794.0 (17%)	652.1 (15%)
Normal	2,072 (43%)	5,772 (43%)	2,072.0 (43%)	1,985.2 (44%)
High	371 (7.8%)	1,031 (7.7%)	371.0 (7.8%)	336.4 (7.5%)
Missing	1,534 (32%)	5,930 (44%)	1,534.0 (32%)	1,496.9 (34%)

- After propensity score weighting, the standardized difference of all covariates **were less than 0.1** (Rubin's Rule 1).
- The **empirical cumulative distribution functions** of continuous variables was almost **identical** after weighting (Strong balance).

Adelborg et al. 2019

Table One, cont'd

Co-morbidities

- Myocardial infarction
- Congestive heart failure
- Peripheral vascular disease
- Cerebrovascular disease
- Dementia
- Chronic pulmonary disease
- Connective tissue disease
- Ulcer disease
- Mild liver disease
- Diabetes without end-organ damage
- Hemiplegia
- Moderate to severe renal disease
- Diabetes with end-organ damage
- Moderate to severe liver disease
- AIDS
- Hypogonadism
- Hypopituitarism
- Klinefelter's syndrome
- Down's syndrome
- Testicular torsion
- Varicocele
- Cryptorchidism
- Orchitis
- Chronic kidney disease
- Myxedema
- Obesity*
- Alcoholism
- Hypertension
- Any cancer (except prostate cancer)
- Illicit drug abuse

Comedications

- ACE/ARB
- Beta-blockers
- Statins
- Low-dose aspirin
- Clopidogrel
- Vitamin K antagonists
- Diuretics
- NSAID
- Opiods
- Antidepressants
- Antipsychotics
- Erectile dysfunction drugs

Adelborg et al. 2019

0-1 Year Follow-up

Outcome by testosterone level	No. at risk/No. of events	Incidence rate* (95% CI)	Unadjusted hazard ratio (95% CI)	Hazard ratio after IPTW weighting (95% CI)
Stroke				
Normal	13,034/65	0.55 (0.43; 0.70)	1.00 (ref)	1.00 (ref)
Low	4,469/45	1.17 (0.87; 1.57)	2.13 (1.46; 3.11)	1.33 (0.84;2.09)
Myocardial infarction				
Normal	13,026/50	0.42 (0.32; 0.55)	1.00 (ref)	1.00 (ref)
Low	4,460/29	0.75 (0.52; 1.08)	1.77 (1.12; 2.80)	1.47 (0.91; 2.38)
Venous thromboembolism				
Normal	13,266/42	0.35 (0.26; 0.47)	1.00 (ref)	1.00 (ref)
Low	4,656/33	0.82 (0.59; 1.16)	2.37 (1.50; 3.73)	1.10 (0.65; 1.85)
All-cause mortality				
Normal	13,467/263	2.14 (1.89; 2.41)	1.00 (ref)	1.00 (ref)
Low	4,771/421	10.25 (9.31; 11.27)	4.75 (4.07; 5.54)	2.08 (1.72; 2.52)

Adelborg et al. 2019

0-5 Year Follow-up

Outcome by testosterone level	No. at risk/No. of events	Incidence rate* (95% CI)	Unadjusted hazard ratio (95% CI)	Hazard ratio after IPTW weighting (95% CI)
Stroke				
Normal	13,034/199	0.47 (0.41;0.54)	1.00 (ref)	1.00 (ref)
Low	4,469/105	0.79 (0.65;0.95)	1.67 (1.32;2.12)	1.14 (0.87; 1.49)
Myocardial infarction				
Normal	13,026/154	0.36 (0.31; 0.42)	1.00 (ref)	1.00 (ref)
Low	4,460/65	0.49 (0.38; 0.62)	1.33 (1.00; 1.78)	0.95 (0.70; 1.30)
Venous thromboembolism				
Normal	13,266/113	0.26 (0.22; 0.31)	1.00 (ref)	1.00 (ref)
Low	4,656/65	0.47 (0.37; 0.60)	1.78 (1.31; 2.41)	1.10 (0.78; 1.55)
All-cause mortality				
Normal	13,467/919	2.08 (1.95; 2.22)	1.00 (ref)	1.00 (ref)
Low	4,771/848	5.95 (5.56; 6.36)	2.83 (2.58; 3.11)	1.48 (1.32; 1.64)

Key Results

Unadjusted (Low v.s. Normal)

- 1-year ❌ all four events
  - Higher **incidence rates**
  - Hazard ratio > 1
- 5-year ❌ all four events
  - Higher **incidence rates**
  - Hazard ratio > 1

Key Results

After IPTW Adjustment (Low v.s. Normal)

- 1-year and 5-year
  - Only all-cause mortality is significantly higher at α=0.05
  - Hazard ratios for stroke, MI, and VTE are not
- All-cause mortality attenuated
  - Between 1-year and 5-year
  - Before and after applying IPTW

Conclusion

For both 1-year and 5-year

- In unadjusted models
  - **A low testosterone level** was a strong predictor for **increased risk** of stroke, MI, VTE, and all-cause mortality
- After applying IPTW
  - **A low testosterone level** remained strong for **only all-cause mortality**

Discussion

Regarding the Results

- IPTW adjustments attenuated much of the effect on stroke, MI, and VTE.
- So, the association may mainly be driven by age and comorbidities included in the IPTW approach.
- Age-stratified results are broadly consistent.

Adelborg et al. 2019

Age-Stratified 0-1 Year Follow-up

Outcome by testosterone level	No. at risk/No. of events	Incidence rate* (95% CI)	Unadjusted hazard ratio (95% CI)	Hazard ratio after IPTW weighting (95% CI)
Myocardial infarction				
Normal	13,026/50	0.42 (0.32; 0.55)	1.00 (ref)	1.00 (ref)
Low	4,460/29	0.75 (0.52; 1.08)	1.77 (1.12; 2.80)	1.47 (0.91; 2.38)

Testosterone level	No. at risk / No. of events	Incidence rate* (95% CI)	Unadjusted hazard ratio (95% CI)	Hazard ratio after IPTW weighting (95% CI)
<55 years				
Normal	7,824 / 6	0.08 (0.04;0.18)	1.00 (ref)	1.00 (ref)
Low	2,335 / 2	0.09 (0.02;0.38)	1.13 (0.23;5.63)	0.89 (0.17;4.70)
55-64 years				
Normal	2,514 / 14	0.60 (0.36;1.02)	1.00 (ref)	1.00 (ref)
Low	793 / 5	0.72 (0.30;1.74)	1.17 (0.42;3.26)	0.96 (0.32;2.86)
65-74 years				
Normal	1,793 / 21	1.34 (0.87;2.05)	1.00 (ref)	1.00 (ref)
Low	670 / 6	1.07 (0.48;2.37)	0.80 (0.32;1.97)	0.75 (0.30;1.91)
+75 years				
Normal	895 / 9	1.17 (0.61;2.25)	1.00 (ref)	1.00 (ref)
Low	662 / 16	3.36 (2.06;5.48)	2.79 (1.23;6.35)	2.79 (1.17;6.65)

←Shouldn't it have balanced the age?

What about putting age in the Cox regression model?

Discussion

- Given how the results are strongly affected by age and comorbidity, unmeasured behavioral data might be strong confounders as well.
- Testosterone is a nonroutine test, and will not be performed unless there is a suspected reason.
- Previous studies were limited on surrogate end points as opposed to actual explicit outcomes.
- Denmark
  - Testosterone therapy has been limited in Denmark(~3% in the sample), making it possible to look at endogenous testosterone levels.
  - Data from a uniformly organized health system.
- Near-identical empirical cumulative distribution functions in place of Rubin's Rules?

Thank you!

OSIA presentation
Amy Attaway

4/9/20

Cardiovascular Outcomes and All-cause Mortality Following Measurement of Endogenous Testosterone Levels

Kasper Adelborg, MD, PhD[a,b,*], Thomas Bøjer Rasmussen, MSc[a],
Helene Nørrelund, MD, PhD, DMSc[a], J. Bradley Layton, PhD[c,d],
Henrik Toft Sørensen, MD, PhD, DMSc[a], and Christian Fynbo Christiansen, MD, PhD[a]

Methods

- Propensity score analysis first used generalized boosted models and the inverse probability of treatment weights
  - Weighting is based on the idea that the sample surveyed is not quite representative of the broader population.
  - For ex, some individuals assigned to a treatment group may be more likely to be assigned to the control group and so we want to give their outcomes as much weight as possible, whereas a much larger group of individuals who were placed in the expected treatment group need less weight, because we have more information on these individuals.

Methods (contd)

- Computed hazard ratios (HRs) with 95% confidence intervals (CIs) using Cox regression analysis before and after propensity score weighting, to compare individuals with low testosterone levels to those with normal levels
- After propensity score weighting, the standardized difference of all covariates were less than 0.1. (no plot?)

Results

Table 1
Characteristics of individuals with normal and low testosterone levels, Denmark, 2000 to 2015

	Unweighted cohorts		Propensity score weighted cohorts	
Variable	Low testosterone	Normal testosterone	Low testosterone	Normal testosterone
Number of men	4,771	13,467	4,771.0	4,470.7
Median age (25th–75th percentiles)	55.4 (38.3–69.2)	50.4 (33.8–63.4)	55.4 (38.3–69.2)	54.8 (37.9–68.5)

Results (contd)

Table 2
Risk of stroke, myocardial infarction, venous thromboembolism, and all-cause mortality in men with normal and low testosterone levels

Outcome by testosterone level	0–1 year of follow-up				0–5 years of follow-up			
	No. at risk/No. of events	Incidence rate* (95% CI)	Unadjusted hazard ratio (95% CI)	Hazard ratio after IPTW weighting (95% CI)	No. at risk/No. of events	Incidence rate* (95% CI)	Unadjusted hazard ratio (95% CI)	Hazard ratio after IPTW weighting (95% CI)
Stroke								
Normal	13,034/65	0.55 (0.43; 0.70)	1.00 (ref)	1.00 (ref)	13,034/199	0.47 (0.41;0.54)	1.00 (ref)	1.00 (ref)
Low	4,469/45	1.17 (0.87; 1.57)	2.13 (1.46; 3.11)	1.33 (0.84;2.09)	4,469/105	0.79 (0.65;0.95)	1.67 (1.32;2.12)	1.14 (0.87; 1.49)
Myocardial infarction								
Normal	13,026/50	0.42 (0.32; 0.55)	1.00 (ref)	1.00 (ref)	13,026/154	0.36 (0.31; 0.42)	1.00 (ref)	1.00 (ref)
Low	4,460/29	0.75 (0.52; 1.08)	1.77 (1.12; 2.80)	1.47 (0.91; 2.38)	4,460/65	0.49 (0.38; 0.62)	1.33 (1.00; 1.78)	0.95 (0.70; 1.30)
Venous thromboembolism								
Normal	13,266/42	0.35 (0.26; 0.47)	1.00 (ref)	1.00 (ref)	13,266/113	0.26 (0.22; 0.31)	1.00 (ref)	1.00 (ref)
Low	4,656/33	0.82 (0.59; 1.16)	2.37 (1.50; 3.73)	1.10 (0.65; 1.85)	4,656/65	0.47 (0.37; 0.60)	1.78 (1.31; 2.41)	1.10 (0.78; 1.55)
All-cause mortality								
Normal	13,467/263	2.14 (1.89; 2.41)	1.00 (ref)	1.00 (ref)	13,467/919	2.08 (1.95; 2.22)	1.00 (ref)	1.00 (ref)
Low	4,771/421	10.25 (9.31; 11.27)	4.75 (4.07; 5.54)	2.08 (1.72; 2.52)	4,771/848	5.95 (5.56; 6.36)	2.83 (2.58; 3.11)	1.48 (1.32; 1.64)

CI, confidence interval; IPTW, inverse probability of treatment weighting.
*Per 100 person-years. For nonfatal outcomes, individuals with previous events were excluded, for example, when stroke was the outcome, individuals with previous stroke were excluded events.

Results (contd)

- Results for stratifying by age

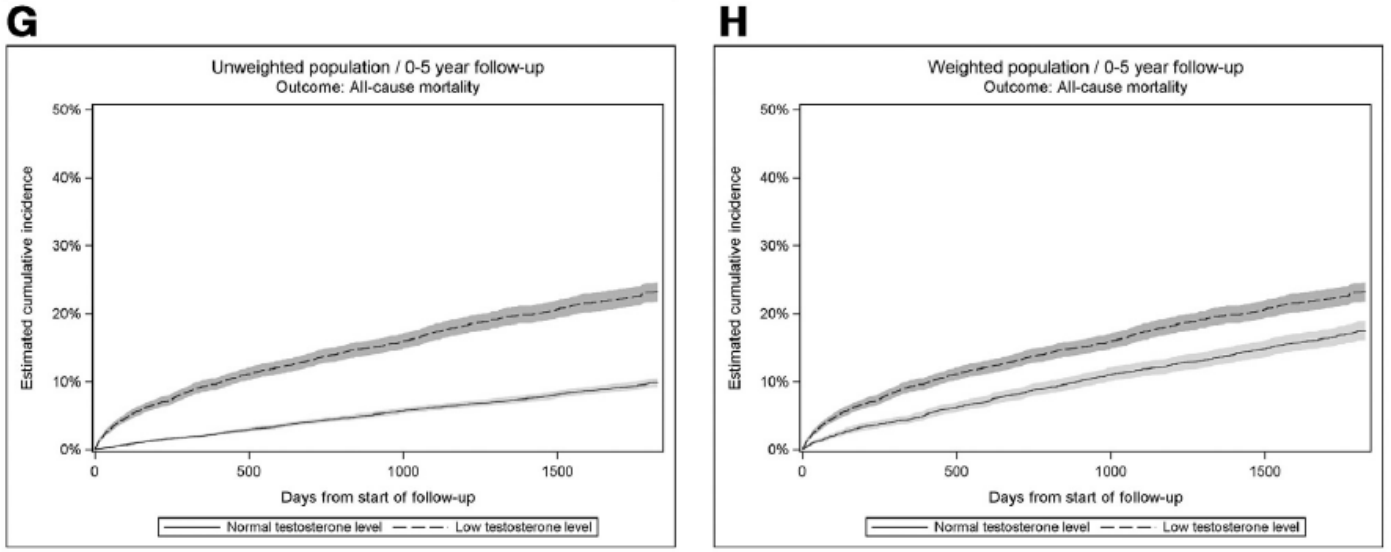
Supplementary Table 6. Risk of all-cause mortality in men with normal and low testosterone levels by age groups.

	0-1 year of follow-up				0-5 years of follow-up			
Testosterone level	**No. at risk / No. of events**	**Incidence rate* (95% CI)**	**Unadjusted hazard ratio (95% CI)**	**Hazard ratio after IPTW weighting (95% CI)**	**No. at risk / No. of events**	**Incidence rate* (95% CI)**	**Unadjusted hazard ratio (95% CI)**	**Hazard ratio after IPTW weighting (95% CI)**
<55 years								
Normal	7,871 / 38	0.52 (0.38;0.72)	1.00 (ref)	1.00 (ref)	7,871 / 132	0.49 (0.41;0.58)	1.00 (ref)	1.00 (ref)
Low	2,360 / 34	1.59 (1.13;2.22)	3.04 (1.91;4.82)	2.18 (1.24;3.81)	2,360 / 83	1.04 (0.84;1.29)	2.14 (1.63;2.82)	1.55 (1.15;2.10)
55-64 years								
Normal	2,619 / 52	2.15 (1.64;2.82)	1.00 (ref)	1.00 (ref)	2,619 / 198	2.26 (1.96;2.60)	1.00 (ref)	1.00 (ref)
Low	841 / 59	8.00 (6.20;10.33)	3.71 (2.55;5.38)	2.20 (1.47;3.29)	841 / 135	5.21 (4.40;6.17)	2.30 (1.85;2.86)	1.60 (1.27;2.02)
65-74 years								
Normal	1,944 / 71	4.13 (3.28;5.22)	1.00 (ref)	1.00 (ref)	1,944 / 264	4.76 (4.22;5.37)	1.00 (ref)	1.00 (ref)
Low	773 / 93	14.30 (11.67;17.52)	3.44 (2.53;4.69)	2.01 (1.40;2.89)	773 / 212	10.19 (8.91;11.66)	2.14 (1.78;2.57)	1.61 (1.31;1.97)
+75 years								
Normal	1,033 / 102	11.47 (9.44;13.92)	1.00 (ref)	1.00 (ref)	1,033 / 325	12.30 (11.04;13.72)	1.00 (ref)	1.00 (ref)
Low	797 / 235	40.59 (35.72;46.13)	3.41 (2.71;4.30)	2.12 (1.61;2.79)	797 / 418	26.01 (23.63;28.63)	2.06 (1.78;2.38)	1.51 (1.28;1.77)

*Per 100 person-years
Abbreviations: CI: confidence interval; IPTW: inverse probability of treatment weighting

Results

- After accounting for measured confounders using IPTW, the cumulative incidence of stroke, MI, and VTE were comparable for persons with low and normal testosterone, whereas the cumulative incidence of all-cause mortality remained higher for persons with low testosterone levels



OSIA – Spending And Quality After Three Years Of Medicare's Voluntary Bundled Payment For Joint Replacement Surgery

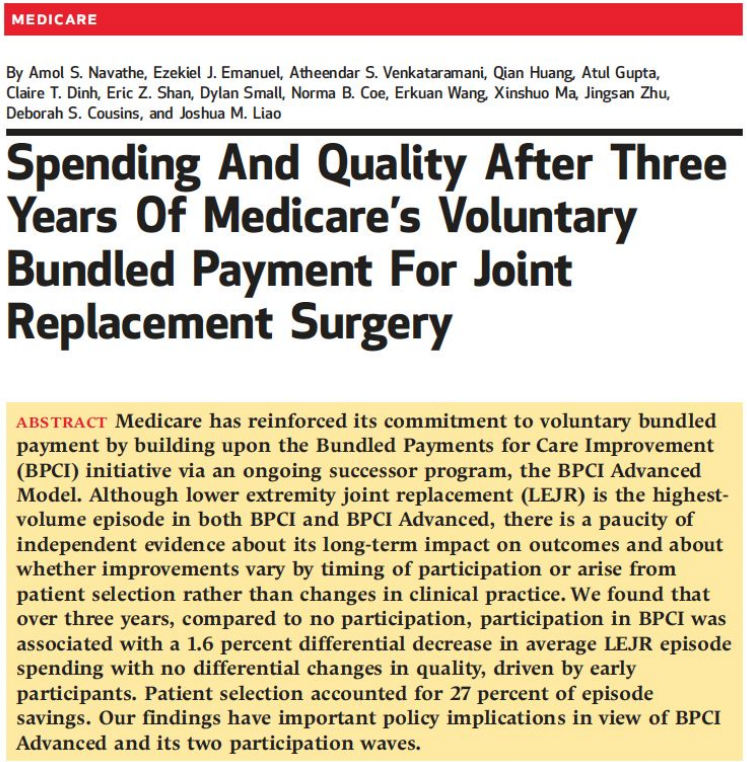
Navathe *et al.* (2020)

Joseph Hnath

9 April 2020

Background

- Bundled Payments
  - Innovation to reduce costs & improve outcomes
  - Give hospitals / provider freedom to coordinate and choose appropriate care
  - Good Idea and initial evidence, but what about widespread?
- LEJR
  - 2014: Over 400,000 procedures and $7 billion for only hospitalizations
  - Costs & quality (obviously) vary widely
    - Some facilities have 3x complications
    - Costs from $16,500 to $33,000
- Comprehensive Care for Joint Replacement Model



MEDICARE

By Amol S. Navathe, Ezekiel J. Emanuel, Atheendar S. Venkataramani, Qian Huang, Atul Gupta, Claire T. Dinh, Eric Z. Shan, Dylan Small, Norma B. Coe, Erkuan Wang, Xinshuo Ma, Jingsan Zhu, Deborah S. Cousins, and Joshua M. Liao

Spending And Quality After Three Years Of Medicare's Voluntary Bundled Payment For Joint Replacement Surgery

ABSTRACT Medicare has reinforced its commitment to voluntary bundled payment by building upon the Bundled Payments for Care Improvement (BPCI) initiative via an ongoing successor program, the BPCI Advanced Model. Although lower extremity joint replacement (LEJR) is the highest-volume episode in both BPCI and BPCI Advanced, there is a paucity of independent evidence about its long-term impact on outcomes and about whether improvements vary by timing of participation or arise from patient selection rather than changes in clinical practice. We found that over three years, compared to no participation, participation in BPCI was associated with a 1.6 percent differential decrease in average LEJR episode spending with no differential changes in quality, driven by early participants. Patient selection accounted for 27 percent of episode savings. Our findings have important policy implications in view of BPCI Advanced and its two participation waves.

Methods – Overall

- Used data from
  - Jan 2011 - September 2013 (pre-BPCI)
  - October 2013 - June 2015 (early-BPCI)
  - July 2015 - December 2016 (late-BPCI)
    - Through September 2016 to allow for 90 day post-discharge to December 2016
- Major outcomes
  - Costs
    - Primary: average Medicare payment per LEJR episode
    - Secondary: subgroups of costs (hospital, readmit, etc.)
  - Quality
    - 90 day post discharge risk-standardized mortality, unplanned readmission, and emergency department visit rates, and LEJR-specific complication rates.
- 1:1 match of non-BPCI and BPCI hospitals with Difference-in-differences model
  - Excluded non-BPCI in BPCI markets because referral patterns could change
  - Instrumental variable also looked referral patterns for predicted hospitalization

STATISTICAL ANALYSES We compared patient, hospital, and market characteristics between BPCI and non-BPCI hospitals and markets. Chi-square tests were used to compare categorical variables, while *t*-tests and Wilcoxon rank-sum tests were used to compare continuous variables.

We conducted unadjusted analyses that compared changes in outcomes across study periods. In adjusted analyses we used a difference-in-differences method to estimate differential changes in spending and quality outcomes for BPCI hospitals compared to a 1:1 propensity-matched set of non-BPCI hospitals in the pre-BPCI and BPCI periods.[20,21] We used baseline hospital and market characteristics and matched BPCI and non-BPCI hospitals based on propensity score values with identical first two digits. Because hospitals that did not participate in BPCI but were located in BPCI markets could have been affected by the program (for example, as a result of changes in market share or referral patterns after it began), only non-BPCI hospitals in non-BPCI markets were used in propensity score matching (for details about our propensity score matching approach, see the appendix methods 1 section and appendix exhibit 1, both in the online appendix).[22]

Methods - Matching

EXHIBIT 1

Baseline characteristics of hospitals that participated in the Bundled Payments for Care Improvement (BPCI) initiative in 2011 and those that did not, after propensity score matching

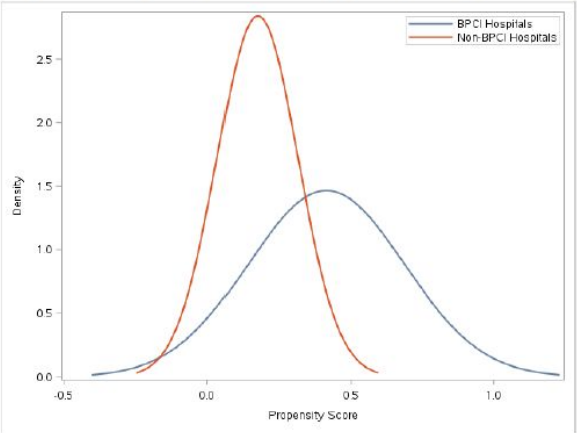
Characteristic	Non-BPCI hospitals (*n* = 244)	BPCI hospitals (*n* = 244)	Standardized difference
HOSPITAL ADMISSIONS			
Mean annual admissions for top 10 BPCI episodes (%)[a]	22.7	22.6	–0.03
Median no. of annual admissions for LEJR	170	179	0.01
Mean share of discharges to the highest-volume SNF (%)	28.9	29.5	0.04
Median share of discharges to the highest-volume IRF (%)	90.0	80.0	–0.10
Median 90-day readmission rate	11.0	11.0	0.12
Median 90-day LEJR episode spending ($)	23,936	24,355	0.09
HOSPITAL ORGANIZATIONS			
Median no. of beds	235	261	0.06
Ownership status (%)			
For profit	17.2	16.4	–0.02
Not for profit	77.1	78.7	0.04
Government owned	5.7	4.9	–0.04
Member of a system (%)	79.1	78.7	–0.01
Teaching status (%)[b]			
Major teaching	13.5	12.7	–0.02
Minor teaching	31.6	34.4	0.06
Nonteaching	54.9	52.9	–0.04
Median ratio of interns and residents to beds	0.0	0.0	0.08
Median DSH payments ($)[c]	2,162,258	2,472,058	0.02
Urban location (%)	99.2	99.2	0.00
Mean Medicare days (% of total patient days)	52.9	52.0	–0.07
Median market share (%)	9.2	9.2	0.03
MARKETS			
Median no. of beneficiaries	1,068,113	1,321,591	0.14
Median income ($)	52,303	53,089	0.05
Median SNF beds	6,124	6,308	–0.01
Median IRF beds	122	134	0.05
Medicare Advantage penetration[d]	24.7	25.4	0.05
Hospital market concentration (HHI)	2,216	2,181	–0.02

Appendix Exhibit 1. Comparison of baseline characteristics of BPCI hospitals and non-BPCI hospitals before and after propensity matching, 2011

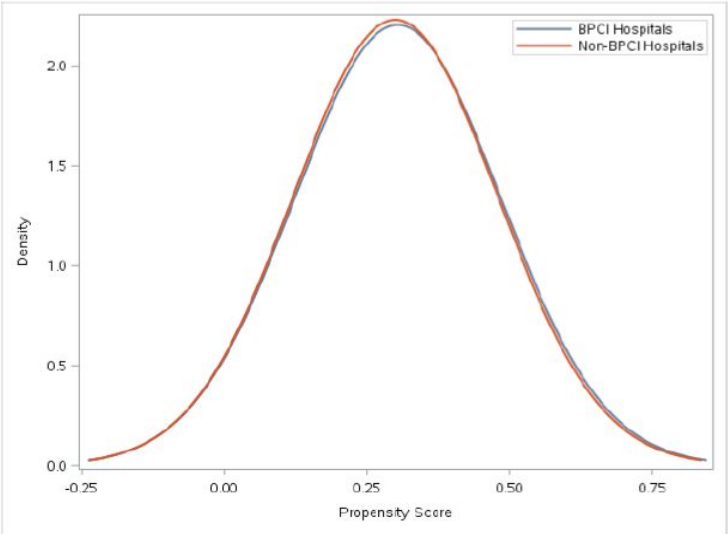
Patient Characteristics						
	Before Propensity Matching			**After Propensity Matching**		
	Non-BPCI Hospitals	**BPCI Hospitals**	**Standardized Difference**	**Non-BPCI Hospitals**	**BPCI Hospitals**	**Standardized Difference**
Annual admissions for top 10 BPCI episodes, mean % (SD)[a]	24.8 (6.7)	22.1 (5.0)	-0.46	22.7 (5.2)	22.6 (5.0)	-0.03
Annual admissions for LEJR, median (IQR)[b]	118 (55 to 229)	175 (94 to 296)	0.39	170 (90 to 314)	179 (96 to 293)	0.01
Proportion of discharges to highest-volume SNF, mean % (SD)	35.7 (19.5)	28.2 (17.0)	-0.41	28.9 (14.8)	29.5 (17.6)	0.04
Proportion of discharges to highest-volume IRF, median % (IQR)[b]	50.0 (0.0 to 100)	80.0 (0.0 to 99.6)	0.22	90.0 (0.0 to 100.0)	80.0 (0.0 to 100.0)	-0.10
90-day readmission rate, median (IQR)[b]	10.5 (5.7 to 16.7)	11.3 (9.1 to 15.3)	0.25	7.1 (16.3)	11.0 (9.0 to 15.0)	0.12
90-day LEJR episode spending, median $ (IQR)[b]	22,996 (20,134 to 26,813)	24,606 (22,343 to 27,523)	0.39	23,936 (21,092 to 27,930)	24,355 (21,977 to 27,309)	0.09
Hospital Characteristics						
	Before Propensity Matching			**After Propensity Matching**		
	Non-BPCI Hospitals	**BPCI Hospitals**	**Standardized Difference**	**Non-BPCI Hospitals**	**BPCI Hospitals**	**Standardized Difference**
Number of beds, median (IQR)[b]	176 (94 to 313)	268 (164 to 402)	0.53	235 (148 to 401)	261 (154 to 391)	0.06
Ownership status, %						
For-profit	19.0	18.2		17.2	16.4	
Not-for-profit	66.7	78.0	0.36	77.1	78.7	0.05
Government	14.4	3.8		5.7	4.9	
Member of a system, %	63.8	80.2	0.37	79.1	78.7	-0.01
Teaching status, %[c]						
Major teaching	7.9	14.7		13.5	12.7	
Minor teaching	30.1	36.1	0.3	31.6	34.4	0.05
Non-teaching	62.0	49.2		54.9	52.9	

Methods - Matching

Density plot: propensity score distribution of all BPCI and all Non-BPCI Hospitals in Non-BPCI Markets, before matching



Density plot: propensity score distribution of BPCI (n=244) and Non-BPCI (n=244) Hospitals, after matching

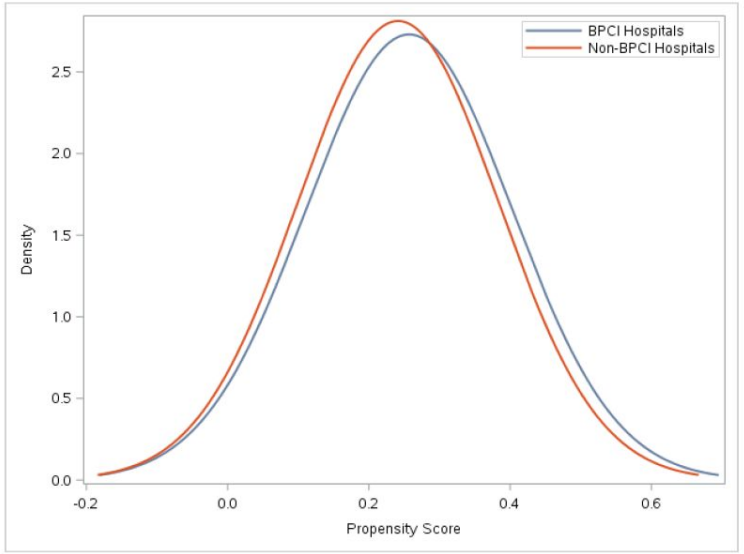


These density plot figures illustrate the area of common support and the distribution of propensity scores before and after matching using a minimum of a 2 digit propensity score match in a 1:1 match.

Methods – Sensitivity Analysis

SENSITIVITY ANALYSES When we extended the approach from prior CMS evaluations, we found that BPCI participation was associated with a 2.2 percent differential decrease in mean episode spending (appendix exhibit 12),[22] a value that was 27 percent lower than estimates from our main analyses (1.6 percent differential decrease in mean episode spending). Analyses that excluded data for January–September 2013, used less stringent propensity matching, and removed hospitals that participated in the Comprehensive Care for Joint Replacement model yielded generally similar results (appendix exhibits 13–15).[22]

Density plot: less stringent propensity score distribution of BPCI (n=225) and Non-BPCI (n=628) Hospitals, after matching



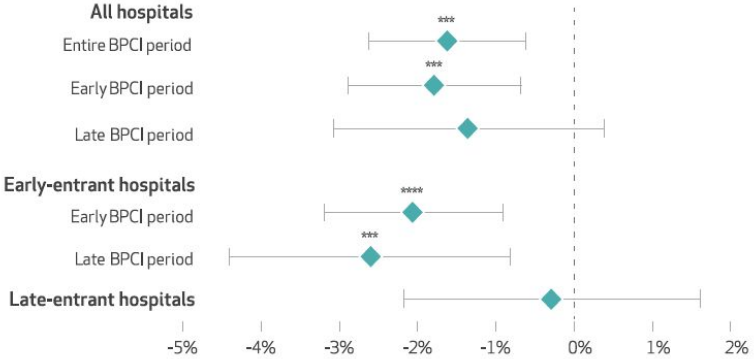
These density plot figures illustrate the area of common support and the distribution of propensity scores before and after matching using a less stringent optimal propensity match method by varying the match ratio from 1:1 to 1:3 and using a caliper=0.05. This resulted in 225 BPCI hospitals matched to 628 Non-BPCI Hospitals.

Results

- For BPCI patients, mean episode spending was $23,552 in the pre-BPCI period and $22,129 in the BPCI period (a reduction of $1,423; p < 0:001)
- Mean episode spending among non-BPCI patients was $22,834 in the pre-BPCI period and $22,073 in the BPCI period (a reduction of $761; p < 0:001)
- Significant cost decreases
    - We see expected early-entrant hospitals with better improvements than late entrants, especially comparing early & late periods

EXHIBIT 3

Percent changes in episode spending associated with hospital participation in the Bundled Payments for Care Improvement (BPCI) initiative, by program period and timing of participation



SOURCE Authors' analysis of Medicare claims data for 2011–16. **NOTES** This exhibit shows the results from a series of difference-in-differences models that used an instrumental variable to evaluate the association between participation in BPCI and differential changes in mean episode spending. Separate models were used to evaluate associations between BPCI participation and changes in mean episode spending for the overall cohort and study period (that is, all hospitals and the entire BPCI period, October 2013–December 2016), all hospitals by program period (early versus late BPCI period, defined in the notes to exhibit 2), early-entrant hospitals (those that began participating in the early BPCI period) by program period, and late-entrant hospitals (those that began participating in the late BPCI period) in the late BPCI period. All spending estimates were standardized and transformed into 2016 dollars. Negative estimates indicate savings. The error bars indicate 95% confidence intervals. ***$p < 0.01$ ****$p < 0.001$

Results

- No statistically significant difference in quality outcomes
- Important that costs are not being reduced with the tradeoff of worse/less care
- Implications about removing inefficiencies/redundant care

EXHIBIT 4

Changes in quality outcomes associated with hospital participation in the Bundled Payments for Care Improvement (BPCI) initiative, by program period and timing of participation

		Program period		Timing of participation		
				Early-entrant hospitals		Late-entrant hospitals
Outcome	**Overall**	**Early BPCI**	**Late BPCI**	**Early BPCI**	**Late BPCI**	**Late BPCI**
Mortality rate	–0.15%	–0.03%	–0.34%	0.01%	–0.27%	–0.37%
Unplanned readmission rate	0.15	0.21	0.03	0.24	–0.10	0.18
ED visit rate	–0.19	–0.22	–0.08	–0.25	0.15	0.54
LEJR-specific complication rate[a]	0.12	0.21	–0.03	0.24	0.003	–0.09

SOURCE Authors' analysis of Medicare claims data for 2011–16. **NOTES** The exhibit shows results from difference-in-differences models that evaluated the association between participation in BPCI and differential changes in quality outcomes, with changes displayed for the overall cohort and intervention period (entire BPCI period) as well as for program period (early and late BPCI, defined in the notes to exhibit 2) and timing of participation (early- and late-entrant hospitals, defined in the notes to exhibit 3). Negative estimates indicate reductions in rates (that is, quality improvements). Emergency department (ED) visits are those without a hospitalization. LEJR is lower extremity joint replacement. Appendix exhibit 11 contains a fuller version of this exhibit (see note 22 in text). [a]Defined by Hospital Compare.

Strengths & Limitations

- Strengths
  - Robust study combining many statistical techniques (propensity score, DiD, IV)
  - Sensitivity analysis help prevent bias of unobserved patient characteristics

- Weaknesses
  - Selection bias of which hospitals will join BPCI program (opt-in)
    - Positive outliers, overestimate effect
    - Different from mandatory programs like Comprehensive Care for Joint Replacement model
  - DiD is good for clear cut changes over time, but BPCI could have some gradual evolutions and systems and providers improve
    - Used early vs. late BPCI, but difference could be more nuanced

Questions?

Spending and Quality After Three Years of Medicare's Voluntary Bundled Payment for Joint Surgery Replacement

Navathe et al. *Health Affairs* (2020)

Observational Studies in Action

PQHS 500

2020-04-08

Second Reader: Wyatt Bensken

SCHOOL OF MEDICINE
CASE WESTERN RESERVE UNIVERSITY

Email: wyatt.bensken@case.edu

Twitter: @Wyatt_Bensken

Overview of Conclusions & Implications

Participation in the Bundled Payments for Care Improvement initiative was associated with a 1.6 percent decrease in average lower extremity joint replacement episode spending.

1. Early reductions in episode spending accounted for the savings observed over 3 years under voluntary LEJR bundled payment
2. Early 1.8% savings came from early entrants with sustained savings
3. Savings may not be generalizable across all participants
4. Patient selection exists in LEJR bundled payment, it does not fully account for associated savings
5. Several measures of quality did not appear to change under LEJR bundled payment, even among hospitals that participated in BPCI for close to two years

Strengths & Weaknesses to Highlight

- "While our use of an instrumental variable accounted for unobserved selection in ways that prior studies did not, more work is needed to ensure that all sources of selection are accounted for in policy evaluations."
- "Fifth, results might not be generalizable to medical condition episodes or episodes initiated by physician group practices."
- **Strengths: Many complex analytical approaches**
  - Is this truly a strength? Could these questions have been answered in a different, less complex way?

Future Directions + Next Steps

- “These findings indicate that despite generating financial savings, early participants were unable to redesign practice or coordinate care in ways that reduced mortality, complication, readmission, or emergency department visit rates. On the other hand, this analysis suggests that financial savings did not appear to come at the expense of quality.”
- “Future work should evaluate other measures of quality, such as patient-reported outcomes, that might not appear in administrative data.”

- So where do we go from here?